BIOORGANIC MOLECULES IN THE COSMOS AND THE ORIGIN OF DARWINIAN MOLECULAR SYSTEMS

By

ALONSO RICARDO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2004

This work is dedicated to my family.

ACKNOWLEDGMENTS

I will like to thank Dr. Steven Benner for the research opportunities and guidance that he has offered me. I also thank Dr. Matthew Carrigan for his help, friendship and for being an excellent lab partner. I am grateful to the people in the mass spectrometry laboratory at the University of Florida and to Dr. Maurice Swanson for providing laboratory space and collaboration . I also thank Fabianne for her unconditional love and support. Finally, I am forever grateful to my family for their constant support, care, and consideration that helped me to made this work possible.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

BIOORGANIC MOLECULES IN THE COSMOS AND
THE ORIGIN OF DARWINIAN MOLECULAR SYSTEMS

By

Alonso Ricardo

May, 2004

Chair: Steven Benner
Major Department: Chemistry

Two critical unsolved issues in the origins of life field are the prebiotic formation of the molecular building blocks of life and from these, the appearance of a self-replicating molecule that undergoes Darwinian evolution. In the present dissertation, both issues were addressed by an experimental approach from which the following findings are reported.

A plausible prebiotic route for the synthesis of sugar pentoses starting from materials known in the interstellar matrix was achieved. Ribose, one of the pentoses and constituents of ribonucleic acid, was generated from a reaction mixture containing boron minerals. The role of boron in this process was found to be dual. Boron coordinates to glyceraldehyde blocking the enolization process and binds to the pentose sugar preventing decomposition. The formation of ribose appears to be the natural consequence of the intrinsic chemical reactivity of compounds available from the interstellar medium under alkaline, calciferous conditions. As these conditions are not excluded from the

early Earth, it is also not possible to exclude the availability of pentoses at the time when life originated.

In vitro selections performed in the presence of Mg^{++} generated DNA sequences capable of cleaving an internal ribonucleoside linkage. Several of these, surprisingly, displayed intermolecular catalysis and catalysis independent of Mg^{++}, features that the selection protocol was not explicitly designed to select. A detailed physical organic analysis was applied to one of these DNAzymes, termed **614**. The DNAzyme **614** is more active in *trans* than in *cis*, and more active at temperatures below the selection temperature than at the selection temperature. Many of these properties are unreported in similar systems, and these results expand the phenomenology known for this class of DNA-based catalysts. A brief survey of other catalysts arising from this selection found other Mg^{++}-independent DNAzymes, and provided a preliminary view of the ruggedness of the landscape relating function to structure in sequence space.

Finally, in the last chapter of this dissertation a method was designed, that allows the evaluation and detection of potential organic molecules on Mars. The method was tested with synthetic salts of mellitic acid, that are likely to be formed under Martian conditions. The presence of these molecules in the martian soil, was evaluated by direct comparison with the recently published Infrared spectra of the Mars surface

CHAPTER 1
ORIGINS

Background and Significance

Prebiotic Chemistry

Experimental prebiotic chemistry as a modern approach to study the origins of life was born just over 50 years ago, in the work of Stanley Miller (1953). Miller demonstrated that applying electric discharges to a mixture of reduced gases in the presence of water generates a brown solution containing amino acids. Neither sugars nor nucleobases were produced under these conditions. Miller's experiment, which at the time was considered to reproduce conditions on early Earth atmosphere's, was a modern example of the abiotic origin of biological molecules.

Since then, the approach in prebiotic experiments remains the same. A mixture containing molecules believed to be present in a prebiotic environment is exposed to a source of energy. If any amount of the desired product is detected, it is then claimed that, if there was plausibly a historical moment during which the Earth contained the starting materials, it is then possible to assume that the mentioned reaction took place, allowing the compound to accumulate over long periods of time.

This logic seems intuitively valid. The methodology, however, can be flawed at different stages:

- The retrosynthetic analysis used to determine the most likely starting materials, may ignore the geochemical constraints necessary to make the reaction relevant.

Starting materials can be chosen mainly based on reactivity rather than prebiotic relevance.

- Water is included/excluded in the model by convenience; no logical reasons other than the instability of reactants and products in the chosen solvent are addressed.
- In some cases concentrations of reactants are controlled at stoichiometric ratios to bias the reaction towards one desired product. This is a common practice in organic chemistry. When applied to prebiotic chemistry, however, this practice fuels arguments that origin required intelligent design. It is worth mentioning that limiting reagents are not prohibited as long as their abundance can be explained through geochemical constraints.

The statements expressed above do not necessarily apply to prebiotic chemistry elsewhere in the universe. For example, an inventory of interstellar compounds detected by radioastronomical methods and the inventory of organic compounds found in meteorites suggest that the non-terrean chemical repertoire is rich in molecules that have a short half life under terrean conditions, but nonetheless exist in the cosmos. These molecules constitute valuable starting materials for terrean chemistry.

Biologically relevant compounds may eventually arise when interstellar material delivered to the Earth (by meteor or comet impact) interacts with terrean molecules, volcanic emissions, water, or surface minerals. Therefore, for a prebiotic experiment to be meaningful, chemistry and geology must be linked; only through a concise analysis of the chemical evolution (organic and inorganic) of our planet will we be able to explore the possible outcomes of an experiment within natural constrains.

In the end, this methodology will not conclude how in fact life began, or what chemistry indeed happened, but at the very least, will offer evidence of a plausible mechanism (able to suffer scientific scrutiny) for the formation of key molecules present in modern life.

Sugars in the Prebiotic Environment

Sugars, and pentoses in particular, are one of the building blocks of nucleic acids. Other sugars, including hexoses, are key throughout metabolism and structural biochemistry. Sugars therefore are logical targets for prebiotic chemistry experiments.

Early attempts to synthesize sugars from simpler molecules were not done explicitly to reproduce prebiotic events. In 1861 Butlerow reported the formation of a brown, sweet tasting compound resulting from the reaction of an aqueous formaldehyde solution in the presence of calcium hydroxide ($Ca(OH)_2$). The product had a molecular formula (calculated by elemental analysis) corresponding to $C_7H_{14}O_6$, and was named "methylenitan." The term "formose" to describe the same product composition was introduced by Loew in 1886 while reacting gaseous formaldehyde and calcium hydroxide. Further characterization of the reaction products by derivatization showed that instead of being a single product, the formose sugar was a heterogeneous mixture of monosaccharides. This suggested a complex reaction mechanism with many possible outcomes.

The first mechanistic studies on the formose reaction focused on the isolation of intermediates. From this work, it was found that glycolaldehyde, glyceraldehyde and dihydroxyketone were early addition products of formaldehyde condensation (Henry, 1895; Neuberg, 1902). These compounds can then react with either additional

formaldehyde, or cross-react to produce a mixture containing tetroses, pentoses, hexoses, and branched sugars, in both aldo- and keto- forms.

A kinetic analysis of the formose reaction shows an initial period during which products are not made (the induction period). This is followed by a period in which all the compounds are formed relatively quickly.

Breslow suggested a mechanism to explain the formose reaction in 1959 (Figure 1.1). According to Breslow, if any direct joining of formaldehyde molecules exists, it must be very slow. In this way, he explained the induction period. The novelty in Breslow's work was to introduce the concept of an autocatalytic reaction during which the first addition product, glycolaldehyde, is regenerated. Because the autocatalytic process does not require condensations between two electrophiles (as in the condensation of formaldehyde), the autocatalytic formation of glycolaldehyde proceeds at a fast rate in alkaline conditions.

Several authors tested this mechanism by starting the reaction from the intermediates suggested by Breslow (Pfeil & Ruckert, 1960; Ruckel, Pfeil & Scharf, 1965). These authors obtained similarly complex products. Further, the addition of glycolaldehyde at the beginning of the reaction was found to reduce significantly the induction period.

With the invention of chromatography, formose sugar composition was better characterized and found to consist of 10% C_4, 30% C_5, 55% C_6, 5 % > C_6 (detectable by GC-MS analysis) sugars when formaldehyde is ca. 99% consumed (Weiss *et a1*. 1970; Decker *et al*. 1982).

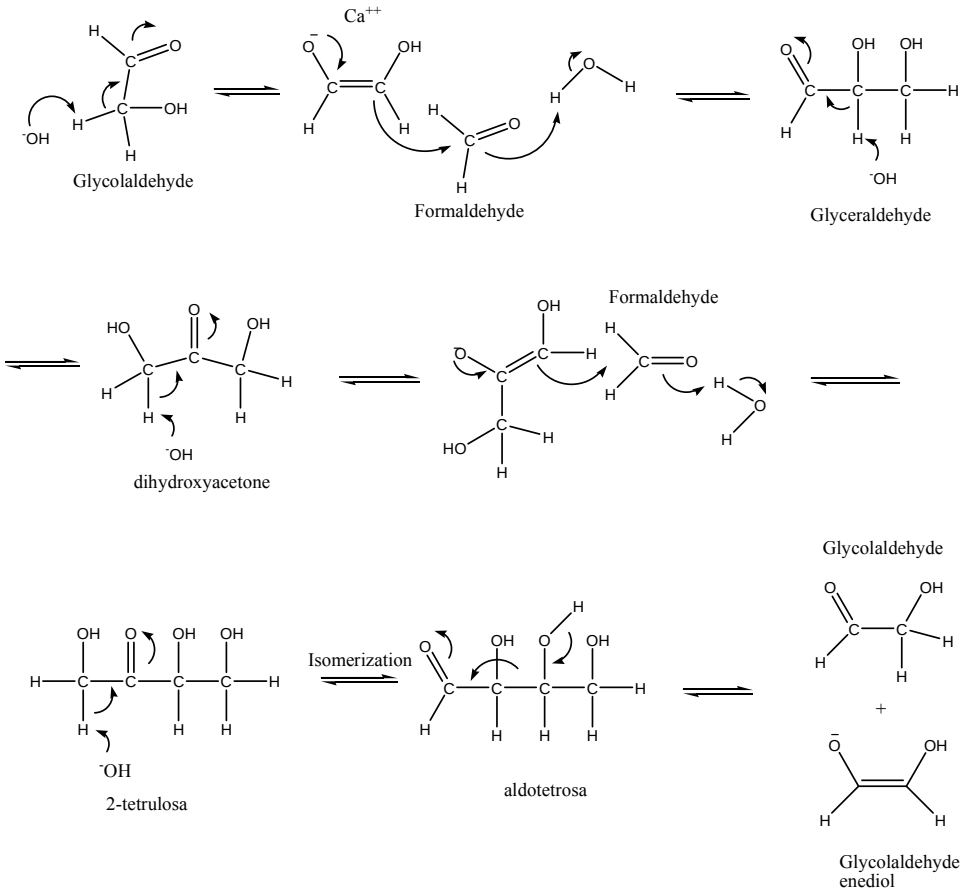


Figure 1.1. Proposed autocatalytic glycolaldehyde regeneration in the formose reaction.

Parallel to the discovery of the formose reaction, studies on the chemical composition of nucleic acids were also yielding some important results. In a series of papers between 1891-1894 Albrecht Kossel deduced the structure of the heterocyclic bases that constitute nucleic acids. Levene and Jacobs (1909) completed the structure of the nucleoside of the RNA (known at the time as "nucleic acid of plant origin") by finding that the carbohydrate portion was made of the sugar D-ribose. The elusive structure of the carbohydrate in DNA (named "nucleic acid of animal origin") remained unsolved until in 1929, when Levene and London concluded that deoxyribose was indeed the sugar component.

With the advent of prebiotic chemistry in the 1960's, it was not long before the formose synthesis was reexamined from a prebiotic context as a way of obtaining the sugars necessary for biology. With the realization that ribose is the backbone of RNA, and having detected this sugar as a product in the formose reaction (although ribose overall yield and long term stability were low in the experiment), prebiotic chemists were satisfied, with the idea of having found that ribose was prebiotically available.

RNA World and the Problem of Ribose Accumulation

In 1962 Alexander Rich hypothesized that RNA might have played both catalytic and genetic roles in early forms of life. This idea was furthered by comments from Woese (1967), Orgel (1968) and Crick (1968), who remarked that transfer RNA (tRNA) appeared to be an RNA molecule attempting to fold like an enzyme. The discovery of catalytic RNA by Cech, Altman, Usher, and others supported this notion (Kruger *et al*. 1982; Guerrier-Takada *et al*. 1983; Usher & McHale, 1976). Gilbert proposed in 1986 that organisms in an "RNA world" may have been a precursor to the contemporary protein-DNA-RNA organisms that dominate life on Earth today. In the RNA world, life used RNA as the sole genetically encoded biological catalysts. However, the plausibility of the RNA world hypothesis is obviously conditioned to the success of generating RNA from a plausible prebiotic soup.

The difficulties of generating RNA without the assistance of a pre-existing living system have been noted by many authors and can be divided into two categories: (a) those directly related to the activity of water, and (b) those that are not.

In the first category, RNA is thermodynamically unstable in water with respect to hydrolysis. Further, many RNA nucleobases are hydrolytically unstable. Cytidine, for

example, deaminates with a half life of ca. 10^2 years to give uridine (Frick, Mac Neela & Wofelden, 1987). Adenosine deaminates to give inosine at a slower rate, while guanosine deaminates to give xanthosine. These reaction modes, favored at high pH, are matched by depurination and depyrimidinylation reactions at low pH, which are also favored thermodynamically in water.

A series of criticisms of the "RNA origins of life" hypothesis are not directly related to water, however, but rather concern ribose itself. These focus on difficulties of creating the necessary amounts of ribose to support a RNA world in the early Earth, and the instability of ribose under prebiotic conditions where it might be generated. This problem was addressed by Robert Shapiro in 1988, who, focusing on the formose reaction, concluded that the synthesis and accumulation of ribose in any significant amount under prebiotic conditions were very unlikely events.

Shapiro's comments are correct; the overall yield of ribose in formose is less than 1% after an arbitrary length of time, and less if the incubation is allowed to continue indefinitely. Ribose itself contains both an electrophilic center (carbon-1) and a nucleophilic center (carbons 1 and 2 of the enediolate) (Figure 2.2). This makes ribose unstable under basic conditions with respect to further reactions with formaldehyde, glycolaldehyde, or itself, or other nucleophiles and electrophiles that are emerging under formose conditions.

Not surprisingly, in the presence of $Ca(OH)_2$, ribose is converted to higher condensation products, branched chain sugars, and (ultimately) a brown, largely intractable polymer of undefined composition. "Browning" of the mixture is pronounced within an hour at room temperature and within minutes at 60 ºC. Thus, the extent of

accumulation of ribose as a product of formaldehyde and glycolaldehyde at steady state (formation minus destruction) is quite low. Given this, Stanley Miller and coworkers, commented that the rate of this decomposition reaction is so high that it suggested that "the backbone of the first genetic material could not have contained ribose or other sugars because of their instability" (Larralde, Robertson & Miller, 1995).

The Role of Minerals on Ribose Formation and Accumulation

While the scientific literature contains many reports concerning the use of minerals in prebiotic experiments, few of these publications include a geochemical explanation of how and why the mineral was available on the early Earth.

Ponnamperuma reported the use of the clay mineral kaolin (aluminium hydroxide silicate ($Al_2(OH)_4.Si_2O_5$) a weathering product of feldspar), as a catalyst in the formose reaction at low concentrations of formaldehyde (Gabel & Ponnamperuma, 1967). Kaolin was shown to facilitate the condensation reaction of formaldehyde at a pH lower than that of calcium hydroxide solutions. Because of this, the sugar products obtained were stable over longer periods of time. Still, the problem of selective formation of ribose was not solved; the overall yield was later calculated by Miller to be approximately 3.8% (Miller, 1984), again after arbitrary time under arbitrary conditions.

Recently, Zubay (1998) reported the use of a combination of lead and magnesium salts in the presence of formaldehyde to generate aldopentoses that constituted 30% of the total product. Here, the lower basicity and solubility of lead and magnesium hydroxide was exploited to moderate the formose decomposition processes. The discussion on the geochemistry of lead was only limited to a list of lead containing minerals.

CHAPTER 2
PREBIOTIC SYNTHESIS OF SUGARS

Introduction

We do not know what organic molecules were present on the early Earth. We may, however, look at compounds in the interstellar medium (ISM), within meteorites, in comets, and in other solar system bodies, to provide a clue.

No direct information indicates what fraction of interstellar and cometary compounds would be delivered to early Earth in unaltered form. It is likely that some would be transformed in icy bodies, especially by high energy particles and photons (Bernstein *et al*. 2002). It is also known that Earth-based chemistry would influence the composition of material. Hydrogen cyanide (HCN), is generated by example, upon comet impact, by an unknown mechanism. At this point in our development of knowledge of the chemistry of the solar system, it is pragmatic to assume a set of compounds such as those shown in Figure 2.1 as our starting point.

How might the molecules in Figure 2.1 be transformed on Earth to give ribose in a stable form ? It is clear that the prebiotic soup would be exposed to rocks and minerals that have some solubility in water at atmospheric pressure (these are found on modern Earth as evaporates). In this light, we re-examined the formose reaction, recognizing that both formaldehyde and glycolaldehyde are found in the ISM.

In its native form, glycolaldehyde can act as an electrophile (the carbon of the C=O group) and as a nucleophile (the alpha carbon, once the 2-position proton is abstracted). As the enolate, however, glycolaldehyde can act only as a nucleophile. To facilitate the

enolization, a cationic species that coordinates the two oxygens of the enediol is needed. The O-C=C bond angle is 120 º. This places the two oxygens at some distance, implying the need for a large metal ion to bridge the long O-O distance. The large Ca^{++} serves this role. Indeed Ca^{++} was the cationic species originally used by Butlerow in 1861, and is the most common catalyst used at present for the formose reaction.

Formaldehyde can act only as an electrophile. The calcium-stabilized enolate of glycolaldehyde can react only a a nucleophile. Therefore, the reaction of formaldehyde and the calcium-stabilized enolate of glycolaldehyde is constrained to give glyceraldehyde as a pair of enantiomers. Glyceraldehyde has a 1,2 diol unit.

Glyceraldehyde, however, can act intrinsically both as an electrophile (the carbon of the C=O group) and as a nucleophile (the alpha carbon, once the 2-position proton is abstracted). The ability of glyceraldehyde to act as both a nucleophile and an electrophile means that it can cross-react to form compounds that resemble tar. For example, reaction of glyceraldehyde as a nucleophile with formaldehyde gives a branched sugar lacking the 1,2-diol moiety. Tar formation is, of course, the standard outcome of the formose reaction, which has been criticized for its prebiological relevance, as noted above.

When glyceraldehyde acts as an electrophile with the calcium-stabilized enolate of glycolaldehyde acting as a nucleophile, however, a pentose is the only product. Four enantiomeric pairs of diastereomeric pentoses exist: ribose, arabinose, xylose and lyxose. As it is drawn in the open chain form, it appears as if ribose can also act both as an electrophile and as a nucleophile (Figure 2.2). This would also permit it to form undesired further products.

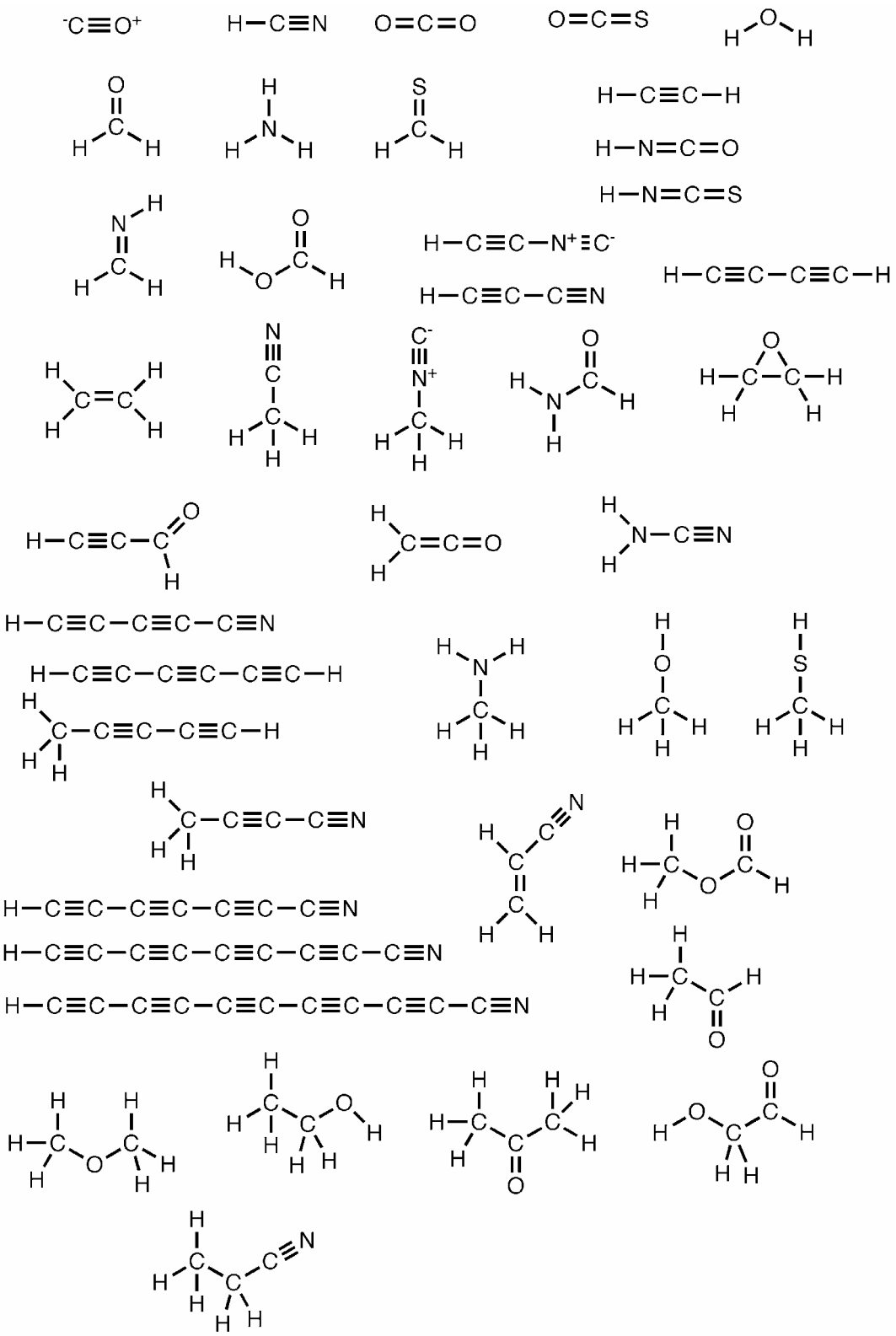


Figure 2.1. Some organic compounds detected in the ISM.

It is clear, however, that these pentoses cannot do so in their ring closed form. Thus, ribose closes to give either a six-membered ring (a pyranose, both in the alpha and beta anomeric forms, about 75% of the total) or a five membered ring (furanose, both in the alpha and beta anomeric forms, about 25% of the total).

We then asked: what mineral components might stabilize the glyceraldehyde against undesired reaction, while directing it towards the reaction that creates ribose?

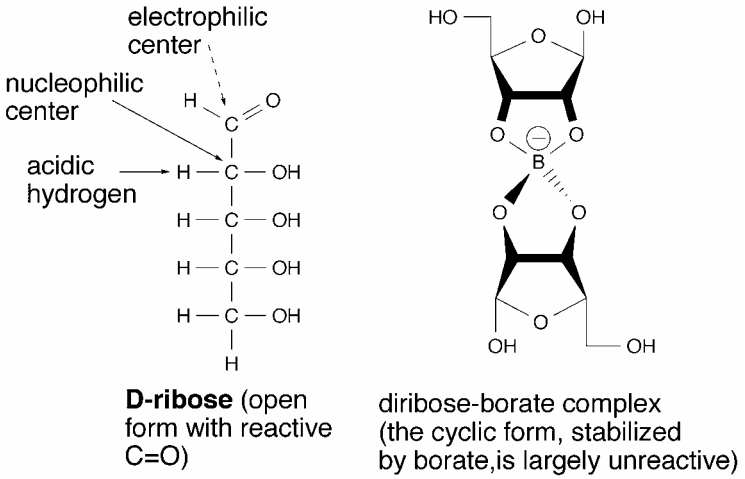


Figure 2.2. D-ribose structure. The open form of ribose contains electrophilic and nucleophilic centers.

Here, we do not seek a large complexing species that can bridge the distant oxygens on an enediol, but rather a small complexing species that can bridge the short oxygen-oxygen distance on a 1,2 diol. Here the O-C-C bond angle is only 109º. The most obvious complexing species for this purpose is small borate. Borate is well known to form a complex with diols, with a micromolar dissociation constant (Böeseken, 1949).

As a borate complex, glyceraldehyde can act as an electrophile. The C=O group is not affected. The borate complex of glyceraldehyde is not expected to enolize easily, however, under alkaline conditions. Abstraction of the 2-proton by base is discouraged by the negative charge already on boron.

Interestingly, borate is expected to stabilize the cyclic form of ribose as well (Figure 2.2). The cyclic form presents two hydroxyl groups in a cis configuration. It is well known that borate complexes with 1,2-diol and 2,3-diol are specially stable. These observations generated the hypothesis that is tested in this chapter: Perhaps borate, if it were present under formose conditions, would manage the reactivity of glyceraldehyde, and stabilize ribose, the desired product.

Boron is known in carbonaceous chondrites, where it is almost certainly present as borate (Zhai & Shaw, 1994). Boron is relatively scarce, relative to carbon and other light elements, however, due to the inefficiency of its synthesis in nuclear reactions.

Borate is, however, excluded from many silicate minerals. For this reason, it appears in the residual melts as lava cools. Here, it is found in tourmalines, minerals that are found in many forms, including colorful forms used as gemstones. Tourmaline weathers from rocks as they are exposed on the surface to generate borate salts, which are generally modestly soluble in water. For example, colemanite is soluble in water to the extent of 0.82 g/L. As a consequence, colemanite and other borate-containing minerals are found in deserts and other dry environments, often under alkaline conditions. Here, they are known as *evaporites*, as they are crystallized from water as it evaporates. These evaporites form under conditions that are close to the conditions that generate pentose.

Boron Chemistry and Complexation Mechanism in Polyols

A clear understanding of the chemistry of boron is necessary for the evaluation of our hypothesis. Boron has five electrons, which can be assigned to an electronic configuration $1s^2\ 2s^2\ 2p_x^1$. In the hybrid orbital state, one electron from the 2s orbital is promoted to the 2p orbital ($2s^1\ 2p_x^1\ 2p_y^1$) to make a an sp^2 hybrid orbital in which each of the three electrons is located in an orbit and able to accept one electron from another

element to form a covalent complex with the boron atom (i.e.: $B(OH)_3$ boric acid). The additional 2p electron orbit is able two hold a pair of electrons from another element, (which completes the octet around boron) this property explains the Lewis acid character of trigonal boron complexes.

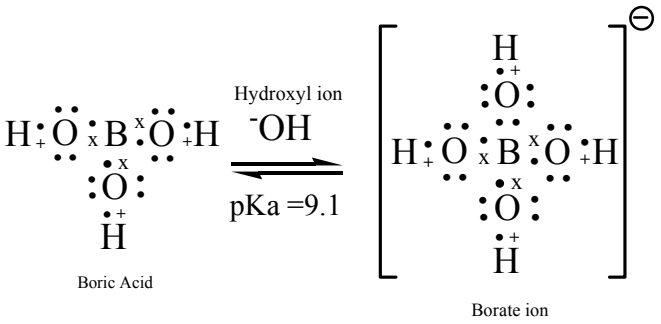


Figure 2.3. Lewis structure representation of boric acid and borate.

Boric acid has a trigonal planar structure in which the B-O bond length is 1.37 Å. In the tetrahedral borate ion, the B-O bond length is 1.48 Å, which makes the hydroxyl group a better proton acceptor (more basic) and therefore a better leaving group when compared to the hydroxyl group in trigonal boron (Pizer & Tihal, 1992).

The reactive form of boron (trigonal vs tetrahedral) towards complexation is dependent on the pH of the solution under study. Is generally assumed that in reactions carried out at the pK_a of boric acid (and above), the reactive species is borate, while the trigonal boron is responsible for the reactivity at lower pH values.

Van Duin *et al*. (1984) suggest that aqueous boric acid exists as an adduct with a water molecule, to give a species whose geometry is tetrahedral, but still neutral. This suggestion makes sense when explaining the reactivity of boric acid towards esterification reactions, in this way a loosely bound water molecule is easily substituted by a hydroxyl group).

The complexation of boric acids and borate with dicarboxylic acids, α-hydroxycarboxylic acids, diols and polyols, has been studied in detail for many years (Mazurek & Perlin, 1963; Pizer & Kustin, 1968; Davis & Mott, 1979; Van duin *et al*. 1984; Van duin *et al*. 1985; Verchere & Hlaibi, 1986; Pizer & Tihal, 1996; Ito *et al*. 2003).

In the case of acidic ligands such as dicarboxylic acids, α-hydroxycarboxylic acids and 1,2 dihydroxybenzenes, the mechanism of boronic acid complexation (at pH values lower than the pK_a of the boronic acid under study) is believed to occur through the nucleophilic attack of a hydroxyl group of the ligand (**1)** to the trigonal boron (**2)** generating an associative transition state in which a proton from the entering ligand is transferred to a leaving hydroxide originally coordinate to boron (Figure 2.4). (Kustin & Pizer, 1968; Pizer & Tihal, 1992, Pizer & Tihal, 1996)

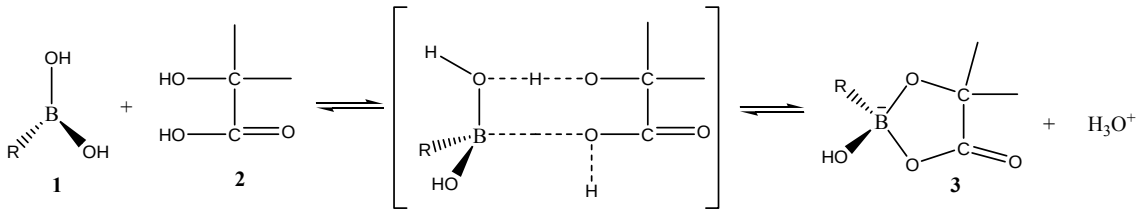


Figure 2.4. Mechanism of boronic acid complexation by acidic ligands. In boric acid (R= OH), the resulting borate monoester (**3)**, has the potential to form a bis-substituted complex by reacting with an additional ligand molecule.

In non-acidic ligands, such as diols and polyols, the predominant complexation reaction occurs with the tetrahedral borate (Figure 2.5). The mechanism of complexation to borate its not yet well understood. But it has been suggested (Pizer & Tihal, 1992) that complexation occurs in two steps, is probably associative in character, and also involves proton transfer. An associative mechanism does not necessarily imply an increase in the coordination number on boron. But even if this is the case, hypervalent pentacoordinated

boron complexes are known in the literature (Yamashita *et al*. 2000; Lee & Martin, 1984). Therefore, a pentacoordinated transition state in borate complexation remains as a possibility.

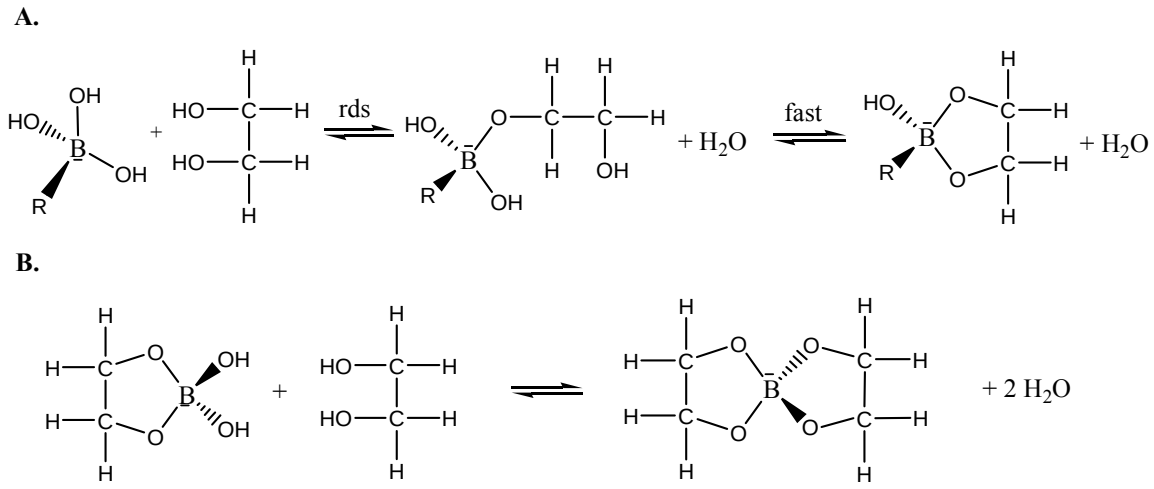


Figure 2.5. Borate complexation by non-acidic ligands. A) Attack of the hydroxyl moiety of the diol to borate is rate limiting. The cyclization reaction to make the five member borate ester is fast. B) In the case of esters of boric acid at high pH, bis-substituted complexes are form by reaction with an additional diol molecule.

pH dependence on the stability of boric acid esters and borate esters.

^{11}B NMR experiments allowed comparison of esters of boric and borate formed by diols, α-hydroxycarboxylic acids and dicarboxylic acids (Van Duin *et al*, 1984). In general it was found that boric acid esters of α-hydroxycarboxylic acids proved to be more stable than those of diols or dicarboxylic acids. Also, it was concluded that the optimal pH stability of boric acid and borate esters, can be predicted by using the "charge rule":

"Esters of boric acid and borate in aqueous medium show the highest stability at that pH where the sum of the charges of the free esterifying species is equal to the charge of the ester."

This rule is better illustrated as:

- Esters of boric acid are most stable at low pH where dissociation of $B(OH)_3$ and ionization of the bidentate ligand (L) hardly occurs (for carboxylic acids and alpha-hydroxyacids).

$$B(OH)_3 + L \rightleftharpoons B(OH)L + 2H_2O$$

- Dissociation of boric acid favors formation of borate esters of 1,2-diols. A pH ≥ pK_a of boric acid is necessary (in other words, diols only form borate complexes at an alkaline pH).

$$B^-(OH)_4 + nL \rightleftharpoons B^-L_n + 2n\ H_2O$$

- In the case of α-hydroxycarboxylic acids and dicarboxylic acids, pH dependent optima are involved. The borate diester of an α-hydroxycarboxylic acid shows maximal stability at pH = pK_a (L) :

$$B(OH)_3 + L + L^- \rightleftharpoons B^-L_2 + 3\ H_2O$$

- The borate monoester, however, occurs preferentially at pK_a (L) < pH < pK_a boric acid. A similar situation applies to dicarboxylic acids at pH = $(pK_{a1} + pK_{a2})/2$

$$B(OH)_3 + L^- \rightleftharpoons B^-(OH)_2L + H_2O$$

The previous rules agree with the expected behavior for a reaction with an associative transition state in which proton transfer is involved postulated by Pizer & Tihal in 1992.

Differential coordination of borate to polyols.

The preferentiality of borate binding to different diols and polyols was studied by measuring the association constants of different borate complexes (Van Duin *et al*, 1985). Experiments were done by ^{11}B NMR at a high pH allowing only borate ions to be the

reactive species. The results of the relative stabilities of complexes obtained by these experiments can be summarized as:

- Tridentate borate > bidentate borate > monodentate borate complexes
- Increasing the number of hydroxyl groups increases the stability of borate esters.
- Diols in a threo-α,β diol conformation > terminal diol > erythro-α,β diol (explained by the steric interactions of the R' groups (Figure 2.6), in the case of terminal diols a loss of enthropy is responsible for the difference)

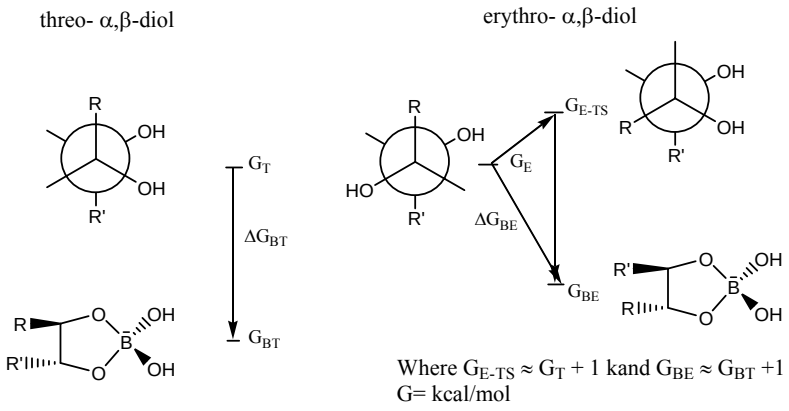


Figure 2.6. Free energy diagram of threo and erythro diols and their borate complexes.

- Coulumbic repulsion decreases stability (introduction of carboxylate residues in the diols, creates repulsion with the negatively charged borate)
- syn-α,γ diols > anti-α,γ-diols (Figure 2.7).

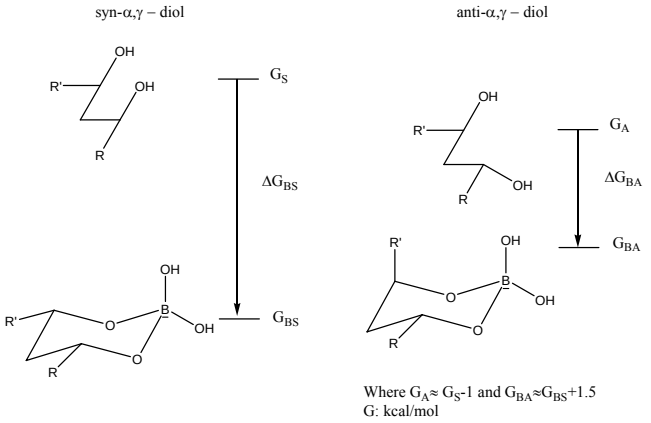


Figure 2.7. Free energy diagram of syn-α,γ and anti-α,γ-diols and their borate complexes.

Interaction of boron with carbohydrates : aldoses and ketoses

The interaction of boron with carbohydrate molecules has been studied in the past with special interest due to the potential applications in stereochemistry and structure determination. Also in the case of carbohydrate mixtures, chromatographic separations may be improved by having boron preferentially coordinated to specific molecules within a family of structurally related sugars.

Several methods have been published to measure the association constants (K_{asso}) between boron and carbohydrates: Potentiometry, pH titration, (^{11}B, ^{13}C, ^{1}H) NMR, calorimetry. However, reported values in the literature differ dramatically depending of the technique used, reasons for these differences other than experimental error include:

- Uncertainty of the anomeric and conformational composition of carbohydrates-borate complexes at equilibrium.
- Stoichiometric ratio of borate(boric)-carbohydrate species is assumed to be homogeneous at a certain pH value and either consistent of borate monoester or borate diester complexes (work by Van Bekkum's group has shown that borate monoester and diesters coexist at certain pH values)

Reproducible values of association constants have been obtained by using a combination of techniques. Mazurek & Perlin (1963) used thermometric measurements of vapor pressure equilibria and ^{1}H-NMR to determine the different complexes of borate with D-glucose, D-threose and *cis*-3,4-dihydroxytetrahydrofuran (Table 2.1). The most important conclusion of their work is the finding that the complexation of D-glucose to borate proceeds via a pyranose to furanose interconversion. This finding is important, since the calculation of association constants by other groups, assume that the ratio of

sugar conformers in the presence of borate is the same as in the aqueous sugar without borate. Mazurek also suggested that pyranose *cis*-diols do not make strong borate complexes, and that a sugar-borate ratio 2:1 yields spirane complexes of the sugars under study.

Verchere & Hlaibi (1987) presented the first comprehensive analysis that considers the effect of borate in the conformational equilibrium of carbohydrates and later these authors include the contribution of this effect in the calculation of the association constants. By using a combination of potentiometric titration and ^{11}B NMR spectroscopy, Verchere confirmed that 1:1 and 2:1 carbohydrate-borate complexes were obtained for every sugar and then calculated the stability constant for each of those complexes (Table 1.1) (Chapelle & Verchere, 1988; *ibid*, 1989). Structural information of the different borate complexes was also deduced from NMR studies (Figure 2.8), allowing a direct interpretation of the stability constants for the monoester β_1 and the diester β_2.

General conclusions from data in Table 2.1 and Figure 2.8 included:

- While sugars in aqueous solution mainly adopt the 6 member ring pyranose conformation, in the presence of borate, this configuration is forced into the five member borate-furanose form.

- At limited amounts of borate, formation of spirane complexes is preferred over the monochelate complex.

- The trend in stability of spirane comples in the monosaccharides shown is :

ketohexoses > aldopentoses > aldohexoses

Table 2.1. Borate complexes of aldopentoses, aldohexoses and ketohexoses. n.d: not determined. Only the strong furanose-borate esters are shown.

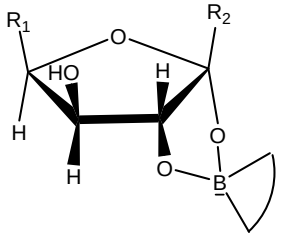
Sugar (furanose form)	Donor sites B^-L_2 (borate ester)	Stability constant($log\ \beta_1$) $B^-(OH)_2L$	Stability constant ($log\ \beta_2$) B^-L_2	% Complexed as B^-L_2
1,4-anhydroerythritol	2,3	n.d	n.d	n.d
β-D-threose	1,2	n.d	n.d	n.d
α-D-ribose (70%)	1,2	2.26	4.28	95 %
β-D-ribose (30%)	2,3			
β-D-arabinose	1,2	2.14	2.99	30 %
α-D-xylose	1,2	1.95	3.74	80 %
α-D-lyxose (60%)	2,3	2.15	3.39	50 %
β-D-lyxose (40%)	1,2			
α-D-allose	1,2	n.d	3.9- 4.4	85 %
β-D-allose	2,3			
α-D-glucose	1,2	1.80	3.05	55 %
α-D-mannose	2,3	2.01	2.74	20 %
α-L-galactose	1,2	1.99	2.56	20 %
α-D-psicose	2,3	n.d	≥ 6	> 99 %
β-D-fructose	2,3	2.82	4.97	n.d
α-D-sorbose	2,3	< 3.5	5.75	n.d
α-D-tagatose	3,4			
β-D-tagatose	2,3	n.d	n.d	n.d

- The anomeric hydroxyl posses a high reactivity towards the formation of borate complexes. In those cases where the anomeric hydroxyl is not involved in complexing, is found in a *trans* position to the vicinal borate-diol ring system.

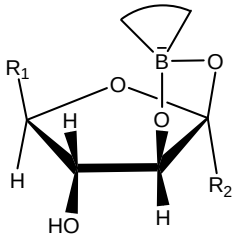
- Within the aldopentoses, the trend in stability of spirane borate complexes follows:

ribose > xylose > lyxose > arabinose

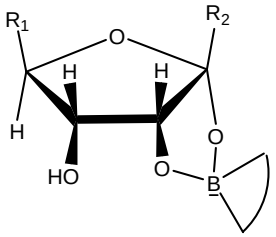
This difference in stability can be explained by considering steric effect between the R_1 group: CH_2OH and the borate ring. In 1,2-*cis* coordinated ribose and xylose



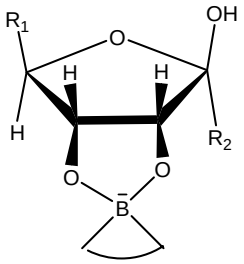
1,2-α-D-xylofuranose $R_1 = CH_2OH$; $R_2 = H$
1,2-α-D-glucofuranose $R_1 = CHOH\text{-}CH_2OH$; $R_2 = H$
2,3-α-D-sorbofuranose $R_1 = R_2 = CH_2OH$



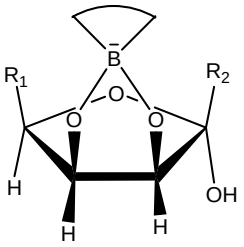
1,2-β-D-arabinofuranose $R_1 = CH_2OH$; $R_2 = H$
1,2-α-L-galactofuranose $R_1 = CHOH\text{-}CH_2OH$; $R_2 = H$
2,3-β-D-fructofuranose $R_1 = R_2 = CH_2OH$



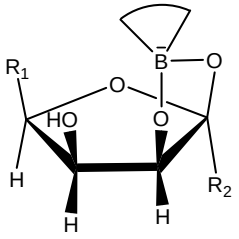
1,2-α-D-ribofuranose $R_1 = CH_2OH$; $R_2 = H$
1,2-α-D-allofuranose $R_1 = CHOH\text{-}CH_2OH$; $R_2 = H$
2,3-α-D-psicofuranose $R_1 = R_2 = CH_2OH$



2,3-β-D-ribofuranose $R_1 = CH_2OH$; $R_2 = H$
2,3-β-D-allofuranose $R_1 = CHOH\text{-}CH_2OH$; $R_2 = H$



2,3-α-D-lyxofuranose $R_1 = CH_2OH$; $R_2 = H$
2,3-α-D-mannofuranose $R_1 = CHOH\text{-}CH_2OH$; $R_2 = H$
3,4-α-D-tagatofuranose $R_1 = R_2 = CH_2OH$



1,2-β-D-Lyxofuranose $R_1 = CH_2OH$; $R_2 = H$
1,2-β-D-mannofuranose $R_1 = CHOH\text{-}CH_2OH$; $R_2 = H$
2,3-β-D-tagatofuranose $R_1 = R_2 = CH_2OH$

Figure 2.8. Structures of the B-L_2 spirane complex formed by monosacharides and borate. The stability constants for these compounds are tabulated in table 1.1.

the borate ring is *trans* to the R_1 group preventing steric interactions. With xylose, the 3'-OH and the R_1 group are *cis* destabilizing the furanose ring.

- In 1,2 *cis* coordinated lyxose and arabinose, the R_1 group is in *cis* position to the borate ring, creating steric interference. The arabinose complex has the 3'-OH in a *trans* position to both R_1 and borate ring, in 1,2-coordinated lyxose the 3 groups are *cis* decreasing the stability of the complex. However lyxose has the ability of

coordinating through hydroxyls 2,3 to borate, increasing the overall concentration of the complex.

Stabilization of pentoses towards decomposition in the presence of borate

Borate complexation with carbohydrates should change the reactivity of the sugars towards enolization. As mentioned before in the case of the pentoses, borate is expected to lock the pentose in a closed form, rendering the complexed sugar largely unreactive. A literature review shows that indeed this is the case. Borate protects monosacharides towards alkaline degradation and the extent of protection is proportional to the stability of the complex. (Bruijn, Kieboom & Van Bekkum, 1986). However the stabilization of aldopentoses had not been reported.

Mendecino (1960) reports the isomerization D-xylose at alkaline pH in the presence of borate to yield D-xylulose (2-ketopentose). Once formed, the ketopentose was found to be stabilized by borate towards degradation at high values of pH and temperature.

In the present chapter we will explore the role of borate and its minerals in the prebiotic synthesis and stabilization of the aldo-pentoses and in particular ribose.

Materials and Methods

Chemicals

All reagents used for synthesis or as standards were purchased from Sigma-Aldrich Co., and were of their highest quality, if not mentioned otherwise. Glycolaldehyde was obtained as the dimer from ICN Biomedicals. Sodium deuteroxide was purchased from Cambridge Isotope Laboratories. Pyridine was purchased from Fluka in anhydrous

quality. Calcium deuteroxide was prepared *in situ* from metallic calcium and deuterium oxide.

^{13}C fully isotopically labeled D-ribose was purchased from Omicron Biochemicals Inc. Silicon chips for desorption ionization on silicon (DIOS) analysis were obtained from Silicon Sense Inc.

Enzymes

Aerobacter aerogenes Type I lyophilized cells as a crude source of ribitol dehydrogenase were obtained from Sigma-Aldrich.

Analytic Instrumentation

Ultraviolet analysis (UV)

UV analysis was performed in a Cary Varian spectrophotometer interfaced to a MS windows based computer.

Gas chromatography (GC)

Gas Chromatography Analysis was perform in a Perkin Elmer 1500 GC equipped with a flame ionization detector and a DB-5 capillary column.

Mass spectrometry (MS)

Gas Chromatography- Mass spectrometry analysis was performed in a Finnigan LCQ – Ion Trap GC-MS equipped with a DB-5 column. Injector temperature : 300 ºC, the temperature program used for the analysis was: 60 ºC- 250 ºC at 3.5 ºC/min, 250 ºC – 300 ºC at 20 ºC/min, final temperature was held for 20 minutes. Analysis were done in collaboration with Dr. Lidia Nikole Matveeva, HPLC-ESI analysis were performed by Dr. Jodie Johnson at the University of Florida.

Desorption Ionization On Silicon (DIOS-MS) analysis was performed on a Bruker Daltonics (Billerica, MA) Reflex II MALDI-TOF mass spectrometer in the negative reflectron mode in collaboration with Qian Li.

NMR Spectroscopy

Mercury 300 MHz. ^{1}H NMR spectra, referenced to the respective solvent (D_2O).

Synthetic Preparations

Synthesis of colemanite

Calcium hydroxide (0.74 g ; 10 mmole) and boric acid (1.84 g; 30 mmole) were added to 30 mL of Milli-Q water. The resulting suspension was stirred for an hour, and then the pH was adjusted to a final value of 12 by dropwise addition of sodium hydroxide (10 M) under stirring. Clear off-white crystals of colemanite ($Ca_2B_6O_{11}.5H_2O$) precipitated from the slurry after one hour.

Synthesis of deuterated Colemanite

Calcium hydroxide (74 mg ; 1 mmole) and boric acid (184 mg; 3 mmole) were added to a Fisher brand conical tube (10 mL, polypropylene) containing deuterium oxide (10 mL). The resulting solution was immersed in liquid nitrogen and lyophilized. The resulting white solid of deuterated colemanite was stored in a desiccator.

Synthesis of pentoses in the presence of colemanite

Glycolaldehyde dimer (6 mg, 0.05 mmol) and D,L-glyceraldehyde (15 mg, 0.16 mmol were dissolved together in an aqueous alkaline colemanite slurry (15 mL, pH: 12). The mixture was heated at 45 ºC for 60 minutes. The reaction was quenched by addition of Dowex W-50 resin H^+ form until a pH of 5.0 was obtained. The resulting clear solution was quickly filtered through a Nalgene 0.2 μm filter, into a Fisher brand centrifuge tube (50 mL) and immediately immersed into liquid nitrogen until complete

solidification of the solution. Lyophilization yielded a white solid powder which was dissolved in anhydrous methanol (5 mL) and subjected to rotary evaporation under high vacuum (repeated 3 times) yielding a white solid (18 mg, 88 % mass recovery of the total carbon input).

Synthesis of pentoses in the presence of calcium hydroxide

Glycolaldehyde dimer (6 mg, 0.05 mmol) and D,L-glyceraldehyde (15 mg, 0.16 mmol were mixed and dissolved into an aqueous alkaline calcium hydroxide slurry (0.74 g, 15 mL, pH: 12). The mixture was heated at 45 ºC for 60 minutes (the solution turned clear brown after 20 minutes). The reaction was quenched by addition of Dowex W-50 resin H^+ form until a pH of 5.0 was obtained. The resulting clear brown solution was quickly filtered through a Nalgene 0.2 μm filter, into a Fisher brand centrifuge tube (50 mL) and immediately immersed into liquid nitrogen until complete solidification of the solution. Lyophilization yielded a dark brown syrup with an strong odor reminiscent of caramel. The syrup was dissolved in anhydrous methanol (5 mL) and subjected to rotary evaporation under high vacuum (repeated 3 times) the syrup remained unchanged after this treatment and it was not possible to obtain a solid even after high vacuum exposure (19 mg, 91% mass recovery of total carbon input).

Derivatization of pentoses for gas chromatography analysis

Derivatization of pure samples of pentose was usually achieved by adding anhydrous pyridine (400 μL) to the corresponding solid pentose (5 mg, 33 nmol) , followed by addition of N,O-bis(trimethylsilyl)trifluoroacetamide (100 μL) under Ar atmosphere, the derivatized samples were used as standards for GC analysis.

Pentoses resulting from synthetic procedures, were derivatized by addition of anhydrous pyridine (800 μL) and of N,O-bis(trimethylsilyl)trifluoroacetamide (200 μL) under Ar atmosphere.

Sugars Degradation Experiments

Sugar decomposition in the presence of calcium deuteroxide

Calcium deuteroxide (16.5 mg) and sodium benzoate (6.1 mg) were added to a 2.5 mL Eppendorf tube containing 1.5 mL of deuterium oxide. This solution was vortexed and then equilibrated at room temperature for 30 min (pD: 12). After equilibration, the tube was immersed into a 4 ºC ice-water bath. An aqueous solution of the corresponding pentose (D-arabinose, D-lyxose, D-ribose, D-xylose and D-ribulose; 500 μL, 1.0 M) was transferred by pipette to the ice-cold solution and vortexed for a period of 1 min. An aliquot (900 μL) of the resulting solution was then removed and used as the sample for the ^{1}H NMR degradation experiment.

Pentose degradation was measured by calculating the ratio of the integral for the sharpest signal corresponding to the anomeric proton (hydrogen in C-1) over the value of the internal standard integral (sodium benzoate). Half-life for the decomposition of each pentose was calculated by plotting the mentioned ratio versus time.

Sugar decomposition in the presence of calcium deuteroxide and synthetic colemanite

Calcium deuteroxide (16.5 mg), deuterated colemanite powder (83 mg), and sodium benzoate (6.1 mg) were added to a 2.5 mL eppendorf tube containing 1.5 mL of deuterium oxide. This solution was vortexed and then equilibrated at room temperature for 30 min (pD adjusted to 12 by sodium deuteroxide addition when necessary). After equilibration, the tube was immersed into a 4 ºC ice-water bath. An aqueous solution of

the corresponding pentose (D-arabinose, D-lyxose, D-ribose, D-xylose and D-ribulose, 500 μL, 1.0 M) was transferred by pipette to the ice-cold solution and vortexed for a period of 1 min. An aliquot (900 μL) of the resulting solution was then removed and used as the sample for the ^{1}H NMR degradation experiment.

Pentose degradation was measured by calculating the ratio of the integral for the sharpest signal corresponding to the anomeric proton (hydrogen on C-1) over the value of the internal standard integral (sodium benzoate). Half life for the decomposition of each pentose was calculated by plotting the mentioned ratio versus time.

Enzymatic Assays

Ribitol dehydrogenase assay

An aliquot (1 mL) of the reaction mixture between glyceraldehyde + glycolaldehyde (with/without boron) was removed after 60 min and treated with a sodium borohydride solution (100 μL, 0.06g/mL in cold water) followed by incubation at room temperature for 30 min. This solution was then acidified to a final pH of 5.0 by addition of aqueous HCl (200 μL, 3 M) and passed through a short column containing Dowex-50 resin H^{+} form. An aliquot (650 μL) of the eluent was lyophilized and the resulting solid was dissolved in anhydrous methanol (650 μL) and subjected to rotary evaporation under high vacuum (repeated 2 times).

The white solid obtained after methanol evaporation was dissolved in potassium phosphate buffer (200 μL, 1.0 M, pH: 10) and then aqueous sodium borate (10 μL, 19 mM) was added. The enzymatic reaction was initiated by addition of nicotinamide adenine dinucleotide (NAD^{+}, 15 μL, 100 mM in water), and of *Aerobacter aerogenes* Type I lyophilized cells suspension (50 μL 0.02 g/mL in 1 M potassium phosphate buffer) and incubated for 60 min at 30 ºC. After incubation, the reaction was placed on an

ice bath for 5 minutes, and then centrifugated to remove cell debris. The presence of 2-pentulose in the resulting solution was confirmed by the cysteine-carbazole test.

Cysteine-carbazole test

An aliquot (100 μL) of solution from the ribitol dehydrogenase assay, was transferred to a quartz cuvette containing a cysteine solution (20 μL, 1.5% in water). sulfuric acid (600 μL, 15.2 M) and carbazole (20 μL, 0.12% in ethanol) were then added and mixed thoroughly. The reaction was kept at room temperature for a period of 60 min. Presence of ribulose in the mixture was confirmed by the appearance of a strong violet color (UV/vis : λ_{max} = 535 nm).

DIOS Analysis

Preparation of PSi surfaces

PSi surfaces were manufactured with HOME (H_2O_2-Metal) etching method: a silicon chip of 2×2 cm was cut from an n-doped crystalline silicon wafer and sputter coated with a thin layer of Au at an argon pressure of 40 Torr and a current of 10 mA. For the production of an array PSi chip, an aluminum mask with 4×4 holes (0.8 mm in diameter) was used in Au coating. The Au coated silicon chip was then immersed in an etching solution (49% HF: 30% H_2O_2: Ethanol = 1: 1: 1, v: v: v) for 10-20 s (Figure 2.9).

Samples of sugars (0.5 μL) were directly pipetted from the mother solution into the PSi surfaces.

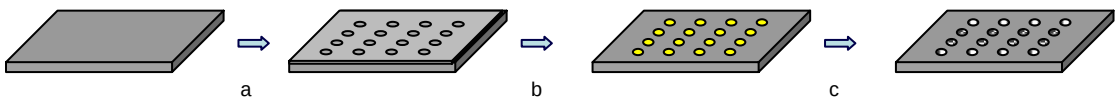


Figure 2.9. Psi surface preparation: (a) An aluminum mask was placed on top of a crystalline silicon chip, (b) Only exposed areas were coated with Au, (c) PSi spots were produced on Au-coated regions

Competition Experiments

1, 4-Anhydroerythritol (AET) vs Pentoses

1,4-Anhydroerythritol (100 μL, 0.1 M, in water) and the corresponding pentose (100 μL, 0.1 M) in water were mixed in a Eppendorf tube (1.5 mL) containing water (700 μL) and vortexed thoroughly. An aqueous sodium borate solution (100 μL, 0.025 M) was then added, and the resulting mixture was vortexed and equilibrated for 2 h at room temperature before DIOS analysis.

^{13}C-Ribose vs Pentoses

^{13}C-Ribose (10 μL, 0.1 M, in water) and the corresponding pentose (10 μL, 0.1 M in water) were mixed into a Eppendorf tube (500 μL) containing water (70 μL) and vortexed thoroughly. An aqueous sodium borate solution (70 μL, 0.025 M) was then added and the resulting mixture was vortexed and equilibrated for 2 h at room temperature before DIOS analysis.

Results

Synthetic Preparations

Synthesis of pentoses in the presence of colemanite

The presence of pentoses in the reaction products after 60 minutes of incubation under formose conditions was directly confirmed by HPLC/ESI (-)-MS analysis. The dipentose-borate complex was detected as an m/z: 307 molecular ion (Figure 2.10). Analysis of a standard of di-ribose-borate complex is shown for comparison in Figure 2.11.

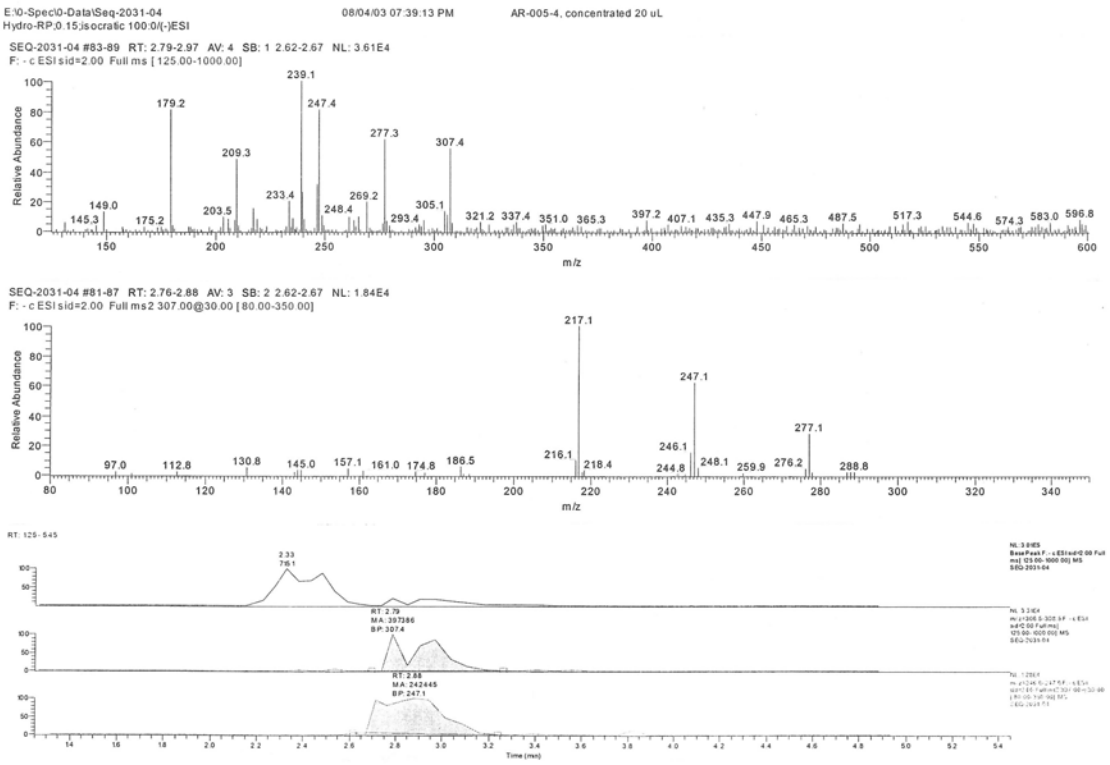


Figure 2.10. HPLC-MS analysis of reaction mixture containing colemanite. Analysis of the products from the reaction between glyceraldehyde an glycolaldehyde in the presence of synthetic colemanite ($Ca_2B_6O_{11}.5H_2O$) at pH:12, 45 °C. (top) ESI-MS indicates the presence of the pentose-borate dimeric complex ion MW: 307. (middle) MS/MS fragmentation pattern of the 307 ion/peak. Fragmentation is consistent with that observed from the standard. (bottom) HPLC-MS trace of reaction mixture showing only various isomeric pentoses.

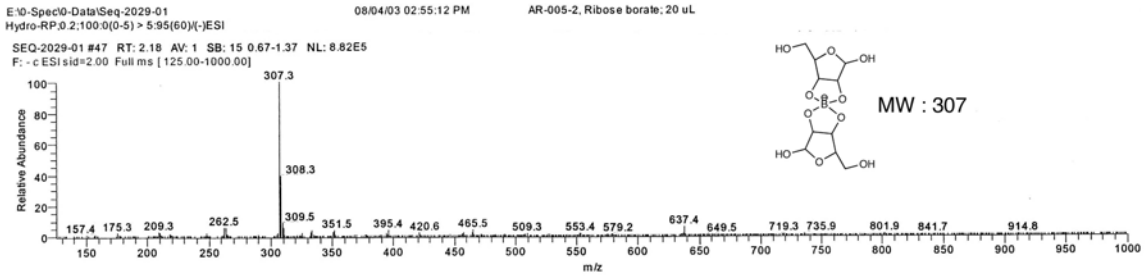


Figure 2.11. Detection of ribose-borate comples by ESI (-) ion mode.(electrospray ionization negative ion mode).

Gas chromatography analysis of the trimethylsilyl derivatives of the reaction products is shown in Figure 2.12. Peaks were assigned by coinjection of authentic standards, using retention time values (Table 2.2).

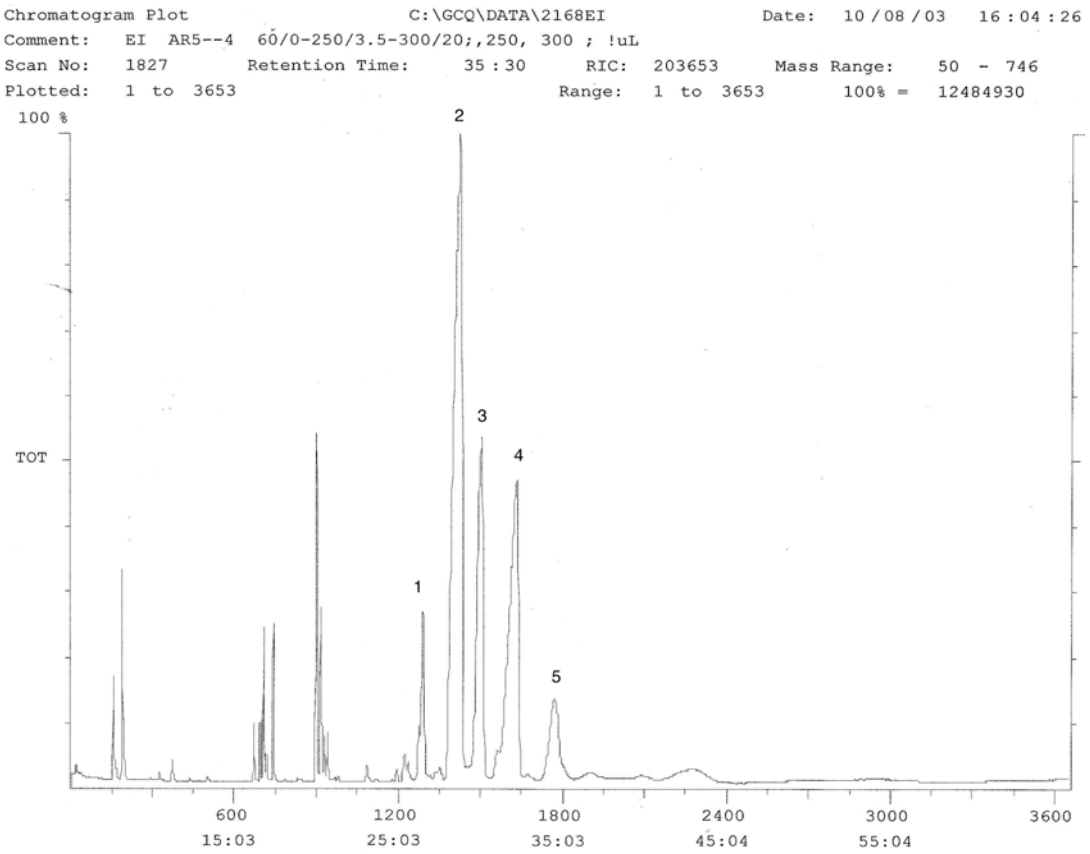


Figure 2.12. GC trace of the reaction mixture containing colemanite. Analysis of the silylated products (after 60 min) from the reaction between glyceraldehyde and glycolaldehyde in the presence of synthetic colemanite($Ca_2B_6O_{11}.5H_2O$) at pH:12, 45 °C. Peak assignation: (1) Lyxose/Arabinose, (2) Lyxose/Arabinose, (3) Ribose, (4)Xylose, (5) Xylose. Anomeric forms of the sugars are not assigned. Notice the absence of peaks corresponding to compounds eluting after 5.

Table 2.2. Retention times of trimethylsilyl derivative of pentoses by GC analysis.

Pentose	Retention Time (min)	Intensity ratios
Arabinose	28.34; 29.49	10:1
Lyxose	28.17; 29.81	1:1
Ribose	30.04	1
Xylose	28.79; 31.97; 34.22	1:2:5

Synthesis of pentoses in the presence of calcium hydroxide

A large portion of insoluble material for this reaction remained insoluble in pyridine after trimethylsilylation at 60 min. Gas chromatography analysis showed an

heterogeneous composition with a low signal intensity (data not shown). Analysis of the reaction products after just 20 minutes is shown in Figure 2.13. Peaks were assigned by coinjection of authentic standards and retention time values.

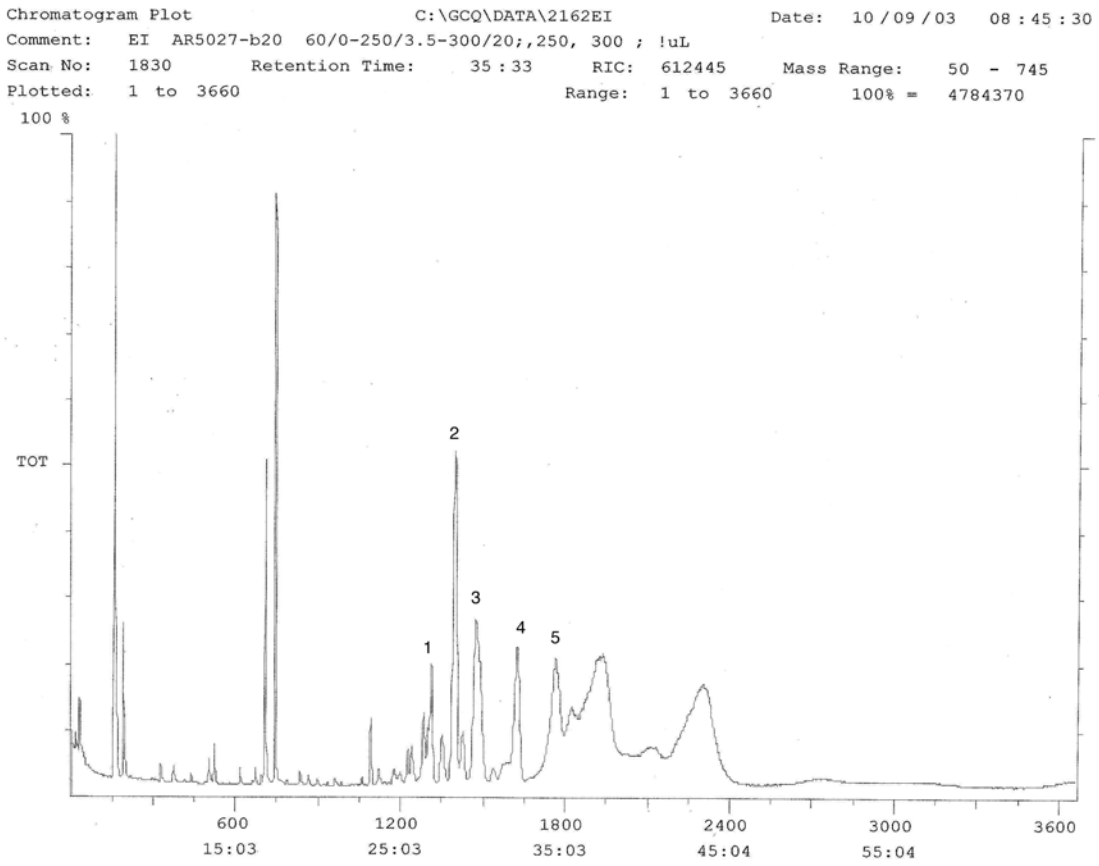


Figure 2.13. GC trace of the reaction mixture containing $Ca(OH)_2$. Analysis of the silylated products (after 20 min) from the reaction between glyceraldehyde and glycolaldehyde in the presence of $Ca(OH)_2$ at pH:12, 45 °C. Peak assignation: (1) Lyxose/Arabinose, (2) Lyxose/Arabinose, (3) Ribose, (4) Xylose, (5) Xylose + other. Anomeric forms of the sugars are not assigned. Compounds to the right, eluting after peak, 5 include tar.

Sugar Degradation Experiments

Sugar decomposition in the presence of Calcium Deuteroxide

Alkaline degradation of sugars in the presence of calcium deuteroxide occurred at a high rate. After a few minutes of reaction, the sugar solutions turned clear brown, and after 1 h some precipitate formed in the NMR tube. The calculated rates of

decomposition (Table 2.3) were obtained by measuring the value of the integral of the signal associated with the anomeric proton and comparing this value with that of the internal standard (Figure 2.14).

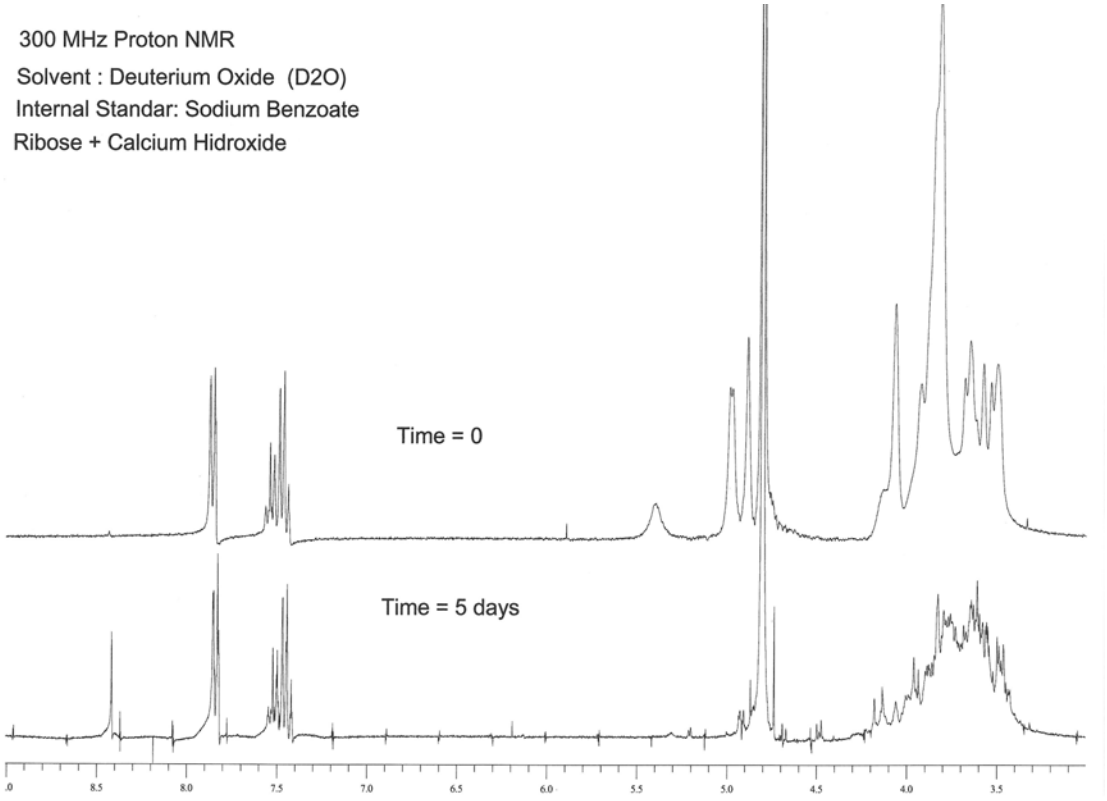


Figure 2.14. Incubation of ribose in $Ca(OH)_2$ solution. Decomposition was monitored by ^1H NMR at 25 °C. Note loss of signals at δ (ppm): 4.85, 4.95, 5.4, corresponding to the signal of H-1 from various anomeric forms of ribose.

Sugar decomposition in the presence of Colemanite

Addition of colemanite had a dramatic effect in the rate of alkaline degradation. All of the sugars were stabilized towards decomposition (when compared to the samples without colemanite). Sugar solutions remained clear after several hours and in the case of ribose it remained unchanged for several days (Figure 2.15). The calculated rates of decomposition are shown in Table 2.3.

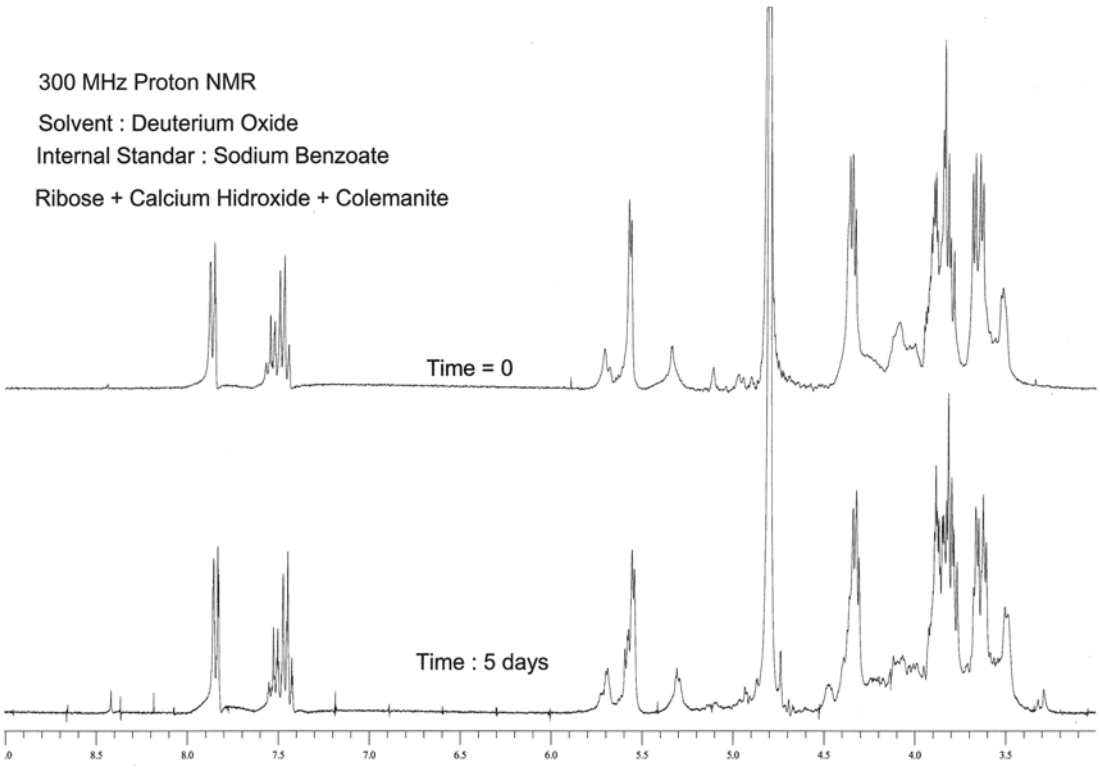


Figure 2.15. Incubation of ribose in the presence of $Ca(OH)_2$ + colemanite. Degradation was monitored by ^1H NMR at 25 °C. Note that signals corresponding to H-1 of various anomeric forms of ribose at δ (ppm): 5.3, 5.5, 5.7, remain unchanged even under incubation for 5 days at high pH.

Table 2.3. Half life of pentoses under alkaline conditions determined by ^1H NMR estimated by loss of selected NMR signals (see appendix 1).

Pentose	Half life (min): $Ca(OD)_2$	Half life (min): $Ca(OD)_2$ + Colemanite
D-ribose	291	2700
D-arabinose	144	259; 716; 331
L-arabinose	124	380; 274
D-lyxose	69; 177	1382
L-xylose	66	572; 258
L-Lyxose	184	1201
ribulose	14	2028

Enzymatic Assays

The presence of ribulose after ribitol dehydrogenase treatment was confirmed by the cysteine-carbazole test. A dark violet color was observed when samples of the reaction between glyceraldehyde and glycolaldehyde were analyzed by this assay.

DIOS analysis: Competition Experiments.

1, 4-Anhydroerythritol (AET) vs Pentoses

The different complexes detected by DIOS are depicted in Figure 2.16. A typical spectra obtained by DIOS analysis is shown in Figure 2.17.

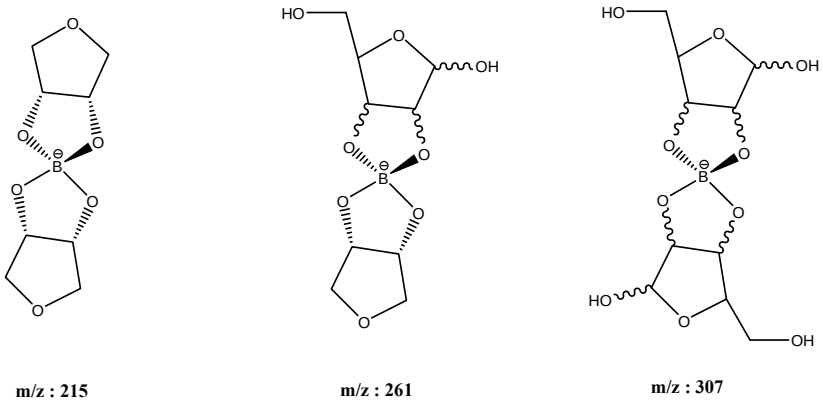


Figure 2.16. Anhydroethrythritol (AET)-pentose borate ions detected by DIOS. Pentose borate complex is depicted through C2 and C3 (cis diol). However, binding through C1-C2 (cis diol) is also known to occur when possible.

Each pentose was analyzed independently, in competition with 1,4-anhydroerythritol (AET). The relativies intensities reported for each detected ion (Figure 2.18) were calculated using equation (1) :

$$\% \text{ Relative intensity} = 100 \times \frac{\text{Intensity of individual ion}}{\sum \text{Intensity of all detected ions}} \quad \text{: equation (1)}$$

Error bars represent the standard deviation calculated from a set of 10 different measurements (for complete set of spectra see appendix 2).

^{13}C-Ribose vs Pentoses

Fully isotopically labeled ribose (13C-ribose, 98.8% labeled) was used for competitions experiments because the difference in mass allows resolution of peaks with non-isotopically labeled pentoses. The different compounds detected in this experiment

are shown in Figure 2.19. A typical DIOS spectra of this experiment is shown in Figure 2.20.

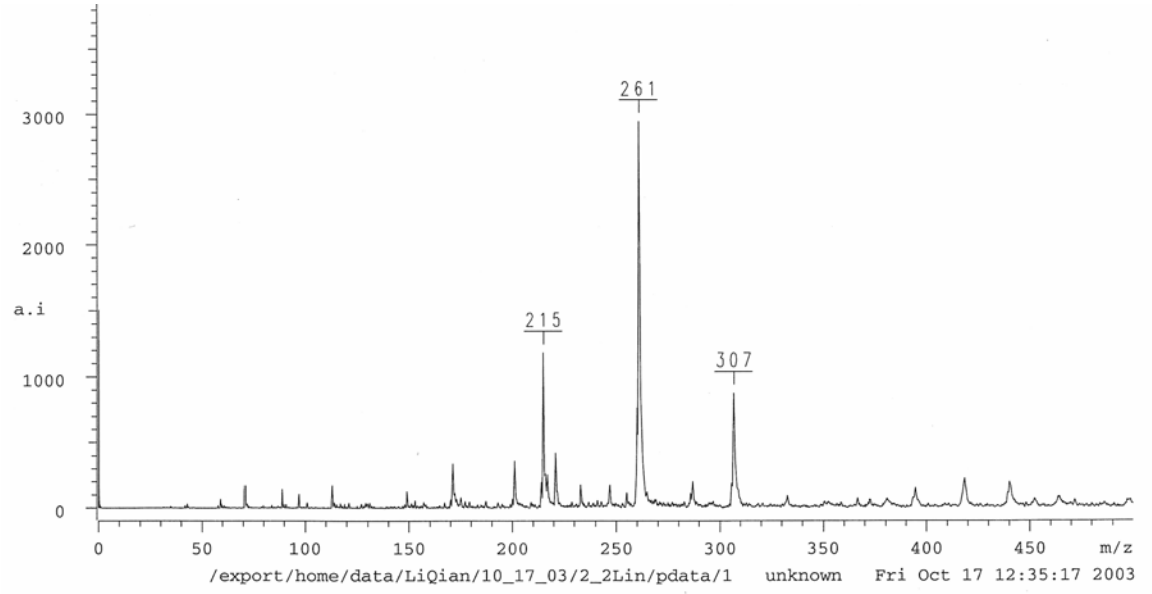


Figure 2.17. DIOS spectra of competition experiment D-arabinose vs AET. Boron (10 mM) was set as the limiting reagent to favor spirane formation.

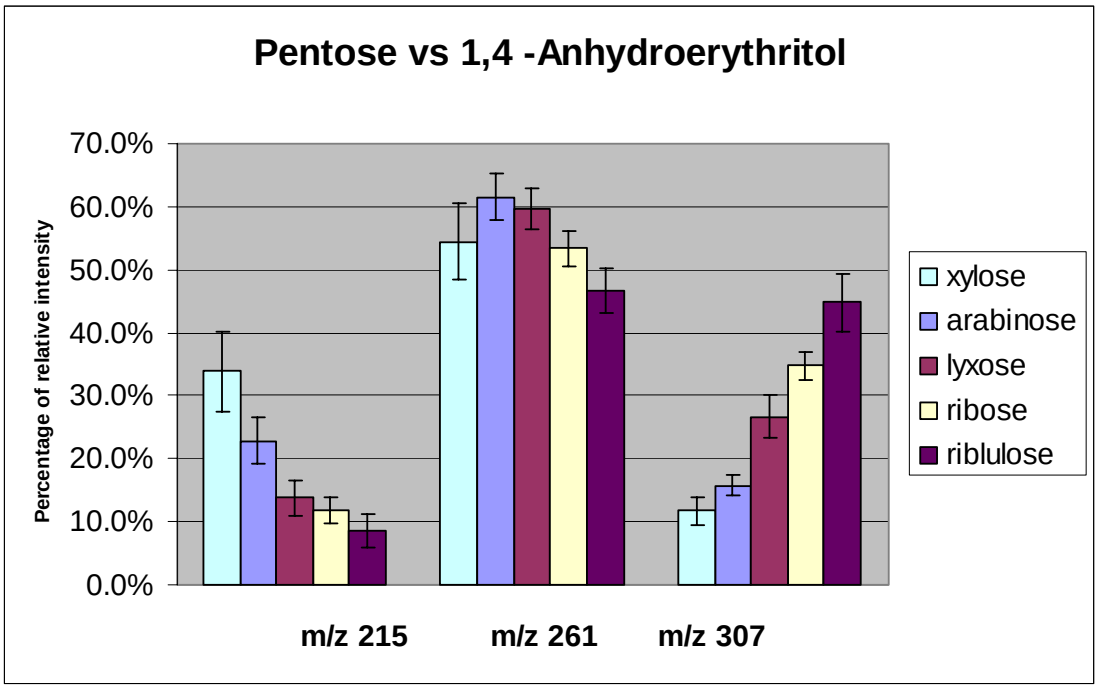
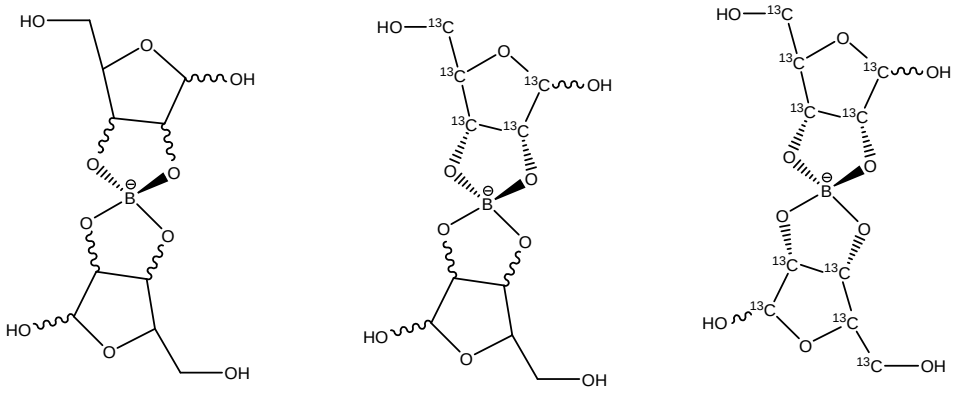


Figure 2.18. Competition experiments between the different pentoses and AET. Relativive intensities represent the molar fraction of the different borate complexes in solution.



m/z : 307 m/z : 312 m/z : 317

Figure 2.19. $^{13}C,^{12}C$-D-ribose borate ions detected by DIOS. Pentose borate complexes detected in the competition experiment using isotopically labeled ribose.

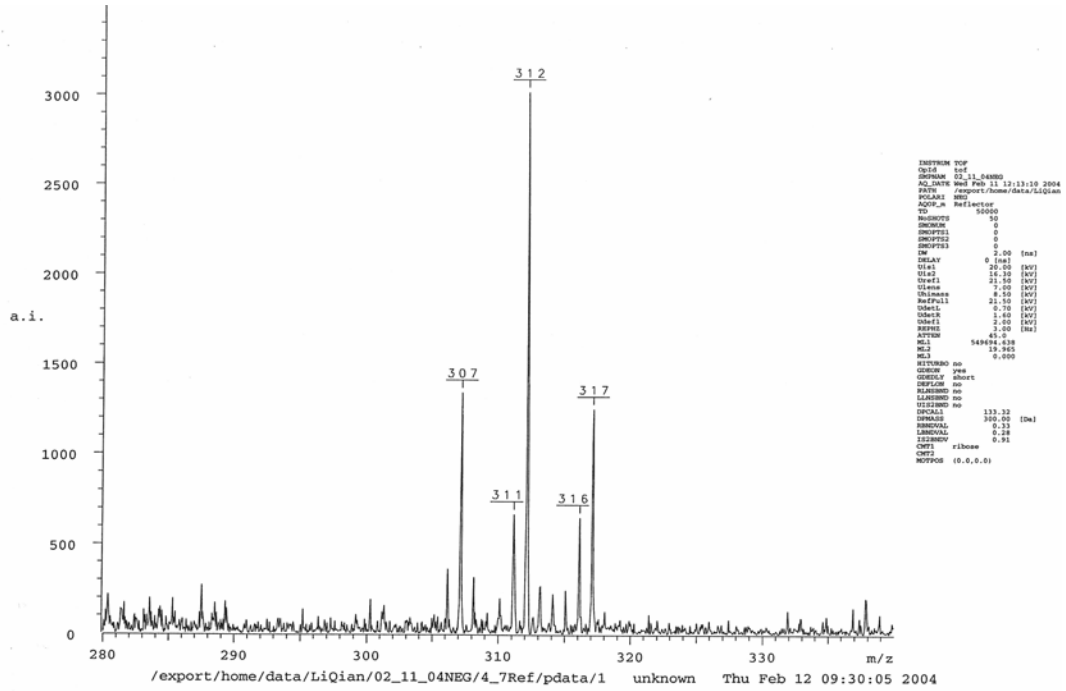


Figure 2.20. DIOS spectra of competition experiment ^{13}C ribose vs ^{12}C ribose. Boron (10 mM) was set as the limiting reagent to force the formation of spirane compounds.

Each pentose was analyzed independently in competition for boron binding against ^{13}C-ribose. Figure 2.21 shows the ratio of ^{13}C-ribose/pentose in equilibrium detected by DIOS. Error bars represent the standard deviation calculated from a set of 10 different measurements.

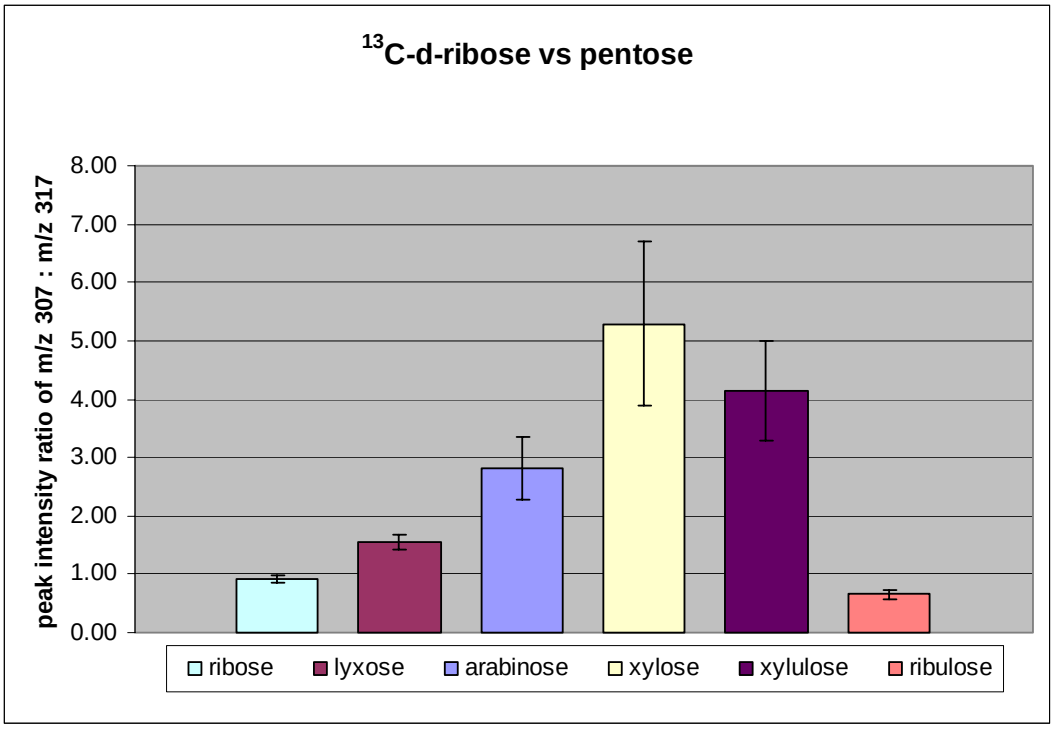


Figure 2.21. Ratio of borate complexes of ^{13}C-ribose vs pentoses.

^{12}C-D-ribose was included in the analysis to test the isotope effect in ionization-vaporization efficiency, however, no such effect was observed during the experiment. The slight deviation from the value of 1.0 in D-ribose is due to the presence of 1.1% unlabeled carbon in the ^{13}C-D-ribose used for the competition experiment.

Discussion

Reaction intermediates of the formose reaction were used as starting materials for the synthesis of pentoses under alkaline conditions. When the reaction was made in the presence of calcium hydroxide, a simple visual inspection gave some clue of the composition. The solution which was originally transparent, turned clear brown after 20 minutes and after one hour dark brown, which, in the case of sugars, is an indicator of decomposition with branching (know as caramelization or browning). Gas chromatography analysis showed that pentoses were indeed present at the early stages of the reaction in addition to other uncharacterized carbonaceous material as seen in Figure

2.13 (probably tetroses, hexoses and branched sugars). However, further incubation of the alkaline mixture, showed a decrease in the total amount of detectable carbon caused by degradation of pentoses to higher molecular weight material that it is not volatile and therefore undetectable by the GC analysis. These results agree with the expected reactivity of the formose mixture.

When boron was introduced as the mineral colemanite into the reaction mixture, visual inspection indicated that the decomposition process was either absent or slow. Browning was not observed even under long incubation periods. GC analysis of the reaction after a period of one hour shows the presence of pentoses as the major products, which indicates that boron is playing an active role in the control of the regiochemistry, and also stabilizing the pentoses product once these are formed. Because an equimolar amount of the fourth aldopentoses was detected, indicates that there was not stereoselectivity in the condensation reaction in the presence of borate.

A proposed mechanism that will account for these observations is depicted in Figure 2.22.

The mechanism depicted above, explains the regiochemistry obtained from the reaction between glyceraldehyde and glycolaldehyde. But it does not address the stabilization of the pentoses by boron, once formed, towards alkaline degradation.

Degradation experiments in the presence of colemanite, showed that all the aldo-pentoses are stabilized by boron, following the trend: ribose > lyxose > arabinose > xylose. Shifts in the NMR signal corresponding to the anomeric proton of the monosaccharides are consistent with previous literature reports indicating that boron coordinates to pentoses, by favoring the formation of a closed furanose conformation.

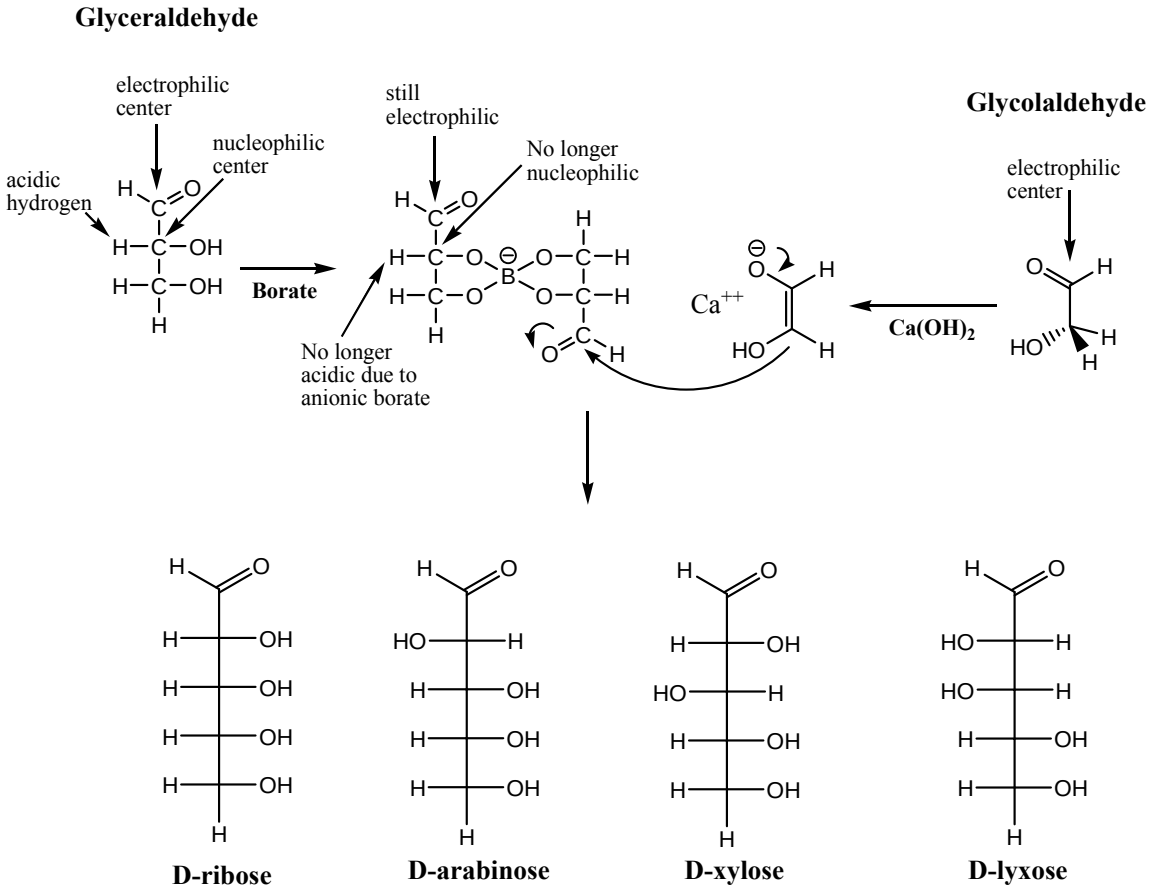


Figure 2.22. Suggested mechanism for pentose formation. The aldol condensation reaction between glyceraldehyde, and glycolaldehyde in the presence of borate is shown. (Pentose products in the Fischer representation are shown only in the D-form for simplicity).

Because the rate of degradation of these sugars is proportional to the amount of pentose in the open form, the half life should be directly proportional to the amount of pentose-borate ester in each case. Boron coordinates pentoses in different fashions to make spirane complexes; this is of course an obvious consequence of the spatial orientation of the different *cis*-hydroxyl groups.

DIOS experiments were focused in understanding these binding preferences. A qualitative description of the preferential binding of borate for the different pentoses shows that the preferential order for binding is: ribose > lyxose > arabinose > xylose.

This information confirms that the stabilization of pentoses (increased half life)

under alkaline conditions is a direct consequence of borate-pentose complex formation.

This synthesis of pentoses in the presence of the mineral colemanite is therefore plausibly prebiotic. Indeed, in the presence of borate, and given that ribose is the first compound in the formose product progression that offers a non-aldehydic cyclic form with unhindered *cis*-diols, the formation of ribose appears to be the natural consequence of the intrinsic chemical reactivity of compounds available from the interstellar medium under alkaline, calciferous conditions. As these conditions are not excluded from the early Earth, it is also not possible to exclude the availability of pentoses at the time when life originated.

This example of how minerals can productively control organic reactivity reminds us of the fact that prebiotic chemistry is occurring on a planet, in the context of a larger geology. Minerals must be considered as we constrain models for the origin of life.

CHAPTER 3
CATALYSIS AND THE RNA WORLD

Introduction

Ten years ago, Szostak, Joyce, Ellington, and others applied *in vitro* selection (IVS) to libraries of nucleic acids to extract nucleic acids that catalyze simple reactions, such as RNA ligation (Bartel & Szostak, 1993) and RNA cleavage (Breaker & Joyce, 1994). This work opened the possibility of using *in vitro* selection to ask quantitative questions about the performance of these catalysts. This includes questions concerning the mechanism of specific nucleic acid enzymes, as well as broader questions, such as how functional behavior is distributed in nucleic acid sequence "space", and whether adding chemical functionality to nucleic acids, either by modifying the nucleobases or by adding cofactors, can enhance the catalytic potential of a nucleic acid library (Breaker, 2000; Perrin *et al*, 2001).

Before broader questions can be addressed using *in vitro* selection, it is necessary to explore some of the specific features displayed by many nucleic acid catalysts that have emerged from selection experiments. For example:

(a) Partial conversion. Many studies of individual nucleic acid catalysts report that their reaction goes only partially to completion. Hammerhead ribozymes, for example, are frequently reported to cleave only 40 - 60% at plateau (Stage-Zimmermann & Uhlenbeck, 1998) (Kore *et al*, 2000). DNAzymes with ribonuclease activity similar to the ones studied here are also frequently reported with cleavage plateaus of 25 - 65% (Perrin

et al, 2001; Geyer & Sen, 1997; Faulhammer & Famulok, 1996; *ibid*, 1997). The same is seen for DNAzymes with DNase activity (Carmi *et al*, 1996).

(b) Intramolecular versus intermolecular reactions. Many reactions catalyzed by nucleic acid catalysts are selected to be intramolecular, making the term "catalyst" technically incorrect; the catalyst is not regenerated. Many of the intramolecular reactions have analogous intermolecular processes that are truly catalytic, however, and these are often accessible both during the selection and in the subsequent kinetic analysis.

(c) Loss of the most active catalysts during the set-up. In many selection schemes, a library of catalysts must be synthesized and folded before the selection step begins. This leads to the possibility that active catalysts in a pool will be lost before the selection system can extract them.

(d) Michaelis-Menten kinetic behavior. Many nucleic acid catalysts bind their substrate in a reversible step prior to the step where chemical bonds are made or broken. This should generate saturation kinetics similar to those seen in protein enzymes.

It is, of course, impossible to address these questions for the general reaction catalyzed by all nucleic acid catalysts. Rather, these questions must be addressed for individual nucleic acid catalysts working on specific reactions. Such studies follow the tradition in physical organic chemistry, where many detailed studies of specific reactions eventually generate a body of literature that addresses broader issues in catalysis.

For this work, we began with work of Breaker and Joyce, who selected for DNA enzymes that catalyze the cleavage of a ribonucleotidyl-3'->5'-deoxyribonucleotide linkage in an oligodeoxyribonucleotide (Breaker & Joyce, 1995). The reaction almost certainly proceeds with the attack on the phosphorus electrophilic center via the

ribonucleotide 2'-hydroxyl group (Figure 3.1). An adaptation of the Breaker-Joyce selection procedure led to several new catalysts for this reaction (Figure 3.2). One of these DNAzymes, designated **614**, was studied in detail to address some of the issues outlined above.

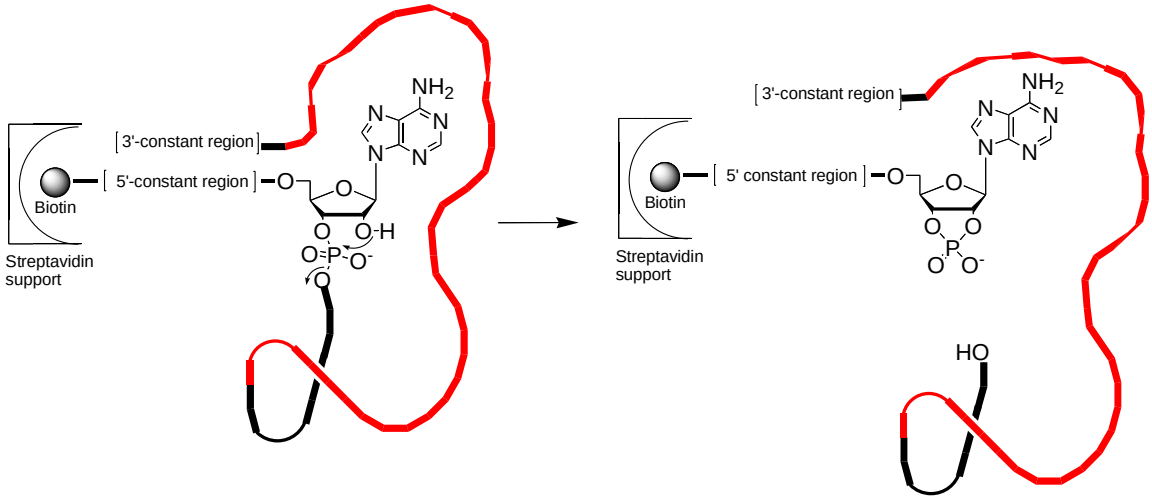


Figure 3.1. *In vitro* selection experiment representation. A library of DNA oligonucleotides containing an internal adenine riboside and a 40-nt random region (represented in red) is attached to a solid support. The DNA-catalyzed ribonuclease reaction proceeds with the attack on the phosphorus electrophilic center by the ribonucleotide 2'-hydroxyl group, cleaving the phosphodiester backbone and releasing the catalytic portion of the DNAzyme from the solid support.

Materials and Methods

Preparation of Precursor DNAzymes via PCR (Maniatis et al, 1982)

DNAzymes were prepared by PCR amplification of the template (synthesized by Integrated DNA Technologies (Coralville IA), or from a clone) using a catalytic strand primer (*cat+ribose* or *cat+deoxyribose*, see Table 3.1) and a complementary strand primer (*compl*, *compl+5'P*, or *compl+tail*). For *trans* cleaving assays, DNAzymes were generated without the internal ribose using the *cat+deoxyribose* primer, identical to the

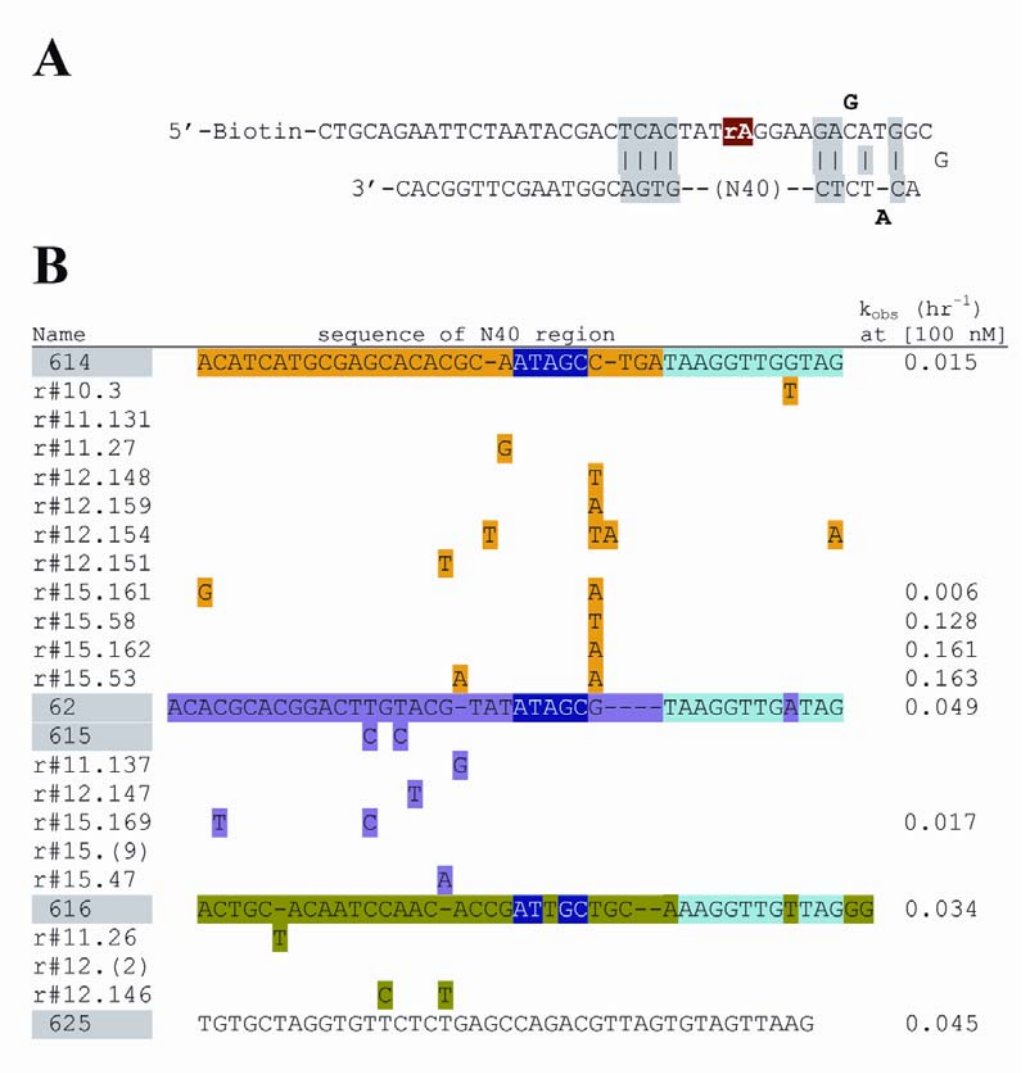


Figure 3.2. Sequence of the initial library and DNAzymes isolated from the *in vitro* selection. (A) The initial library was based on the sequence used by Breaker and Joyce, with an internal adenine riboside incorporated at position 28 to provide a cleavable linker. Two nucleotide substitutions were introduced to eliminate one of the clamps that were designed to hold the substrate and catalyst portions of the molecules in the library together. The original two nucleotides in the Breaker Joyce sequence are shown in bold above and below the sequence used in this study. Base pairing that could form binding clamps are highlighted in grey. (B) DNAzymes isolated from multiple rounds of IVS. Sequences **614**, **62**/**615**, **616**, and **625** represent the major sequences classes initially cloned from the seventh round of the initial IVS. Sequences variants of these major classes were isolated from additional rounds of selection and are grouped according to sequence class (only variations from the prototype sequence are shown). Colored boxes are shown to highlight common motifs. Sequences isolated more than once are shown with the number of isolates identified in parenthesis following the name.

cat+ribose primer except that the single ribo-adenosine in the *cat+ribose* was replaced with a 2'-deoxy*ribose*-adenosine. To obtain single stranded catalysts, DNAzymes were produced using a catalytic strand primer and either a 5'-phosphorylated complementary strand primer (*compl+5'P,* for use with lambda exonuclease) or a complementary strand primer with an 15-nt poly-deoxyriboadenosine tail appended to the 5' via an 18-atom hexaethyleneglycol-based linker (*compl+tail,* for use in asymmetric PCR). This linker prevents polymerase read-through. Templates all had the same 5'-constant region and 3'-constant region to which complementary and catalytic strand primers bind, separated by 40 nucleotides. *Ribose*-**614**, *deoxyribose*-**614**, and the **library** templates were synthesized and PAGE purified by IDT DNA technologies.

Typical conditions for a PCR contained up to 1 ng template, 100 nM catalytic strand primer, 100 nM complementary strand primer, 100 μM dNTPS, 10 mM KCl, 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% Triton X-100, 3-4 units polymerase (*Taq* or *Vent* exo-), and alpha-^{32}P-CTP (10 μCi, for internally labeled samples), for a total volume of 100 μL. The PCR amplification cycle consisted of an initial incubation (3 min, 96°C) followed by 20 PCR cycles of (45 sec at 96°C, 45 sec at 50°C, and 2 min at 72°C).

PCR for *in vitro* selections (IVS) used *Vent* exo- polymerase (3-4 units) and excess catalytic strand primer (400 nM), with up to 40 cycles of PCR in the early rounds.

Preparation of Single-stranded DNAzymes

Double stranded DNAzymes generated via PCR with a 5'-phosphorylated complementary strand primer (*compl+5'P*) were converted to single-stranded DNA by digestion of the complementary strand using lambda exonuclease (which is specific for

the 5'-phosphorylated strand of double stranded DNA). The DNA was recovered by EtOH precipitation, resuspended in exonuclease solution (25 μL, 5 units lambda exonuclease, 67 mM glycine-KOH (pH 9.4), 2.5 mM $MgCl_2$, 50 μg/mL BSA, for a 100 μL PCR). Samples were mixed and incubated (37°C 30 min). Reactions were terminated with formamide stop dye (1 mg/mL xylene cyanol, 1 mg/mL bromophenol blue, 10 mM EDTA, in 98% formamide) and heating (80°C 10 min). The single-stranded products were resolved by 8% PAGE/urea, and full-length ssDNA products excised. Gel slices of individual samples were crushed with individual disposable mortars and were eluted in buffer (350 μL, 500 mM NH_4OAc, 0.1 mM EDTA, 0.1% SDS, pH 7) overnight. Gel-purified samples were extracted with phenol/$CHCl_3$/isoamyl alcohol (25:24:1), then with $CHCl_3$/isoamyl alcohol (24:1), and the resulting DNA was precipitated in NH_4OAc and EtOH.

Single stranded DNAzymes were also purified via asymmetric PCR using a complementary strand primer with a 15-nt poly-deoxyriboadenosine tail connected to the 5'-end of the primer by a C18 linker (*compl+tail*), at which *Vent* and *Taq* polymerases terminate. Complementary strand molecules generated by extension of a complementary strand primer containing the 5'-tail are longer than the full-length catalytic strand, and were separated from the catalytic strand by PAGE/urea (8% acrylamide). Gel-purified samples were excised, extracted and recovered as before.

5'-End Labeling of DNA

Single-stranded DNA (20 mM) was 5'-labeled with gamma-^{32}P-ATP (20 μCi) using T4 polynucleotide kinase (10 units) in Tris-HCl buffer (70 mM, pH 7.6, 10 mM $MgCl_2$, 5 mM dithiothreitol, final volume 10 μL, 30 min 37°C). An equal volume TE

was then added, and the mixture heated (20 min 70°C). End-labeled DNA was separated from unincorporated nucleotides by spinning through a G-25 column (600 g, 3 min).

DNAzyme Kinetic Assays

DNAzymes and substrates were isolated via EtOH precipitation and re-suspended in HEPES buffer (50 mM, pH 7). Their concentration was estimated by Cherenkov counting. Samples were diluted to twice the desired final concentration with additional HEPES buffer. For *trans* assays, enzyme and substrates were typically mixed (unless otherwise noted) and diluted to twice the desired final concentration. Samples were then mixed with equal volume 2X reaction buffer (typically 2M NaCl, 2 mM $MgCl_2$, 50 mM HEPES pH 7). Mixtures were then heated (96°C, 3 min), and slowly cooled to 23°C (over 10 min) – collectively termed "slow cooling". The initial "time zero" point was the time at which the sample completed the slow cooling. Unless otherwise noted, reactions were run at 25°C and terminated at various times by diluting an aliquot of the reaction into the formamide stop dye, followed by freezing (-20°C). Samples were resolved using 8% PAGE/urea and the product cleaved quantified using a Bio-Rad phosphorimager. Data was analyzed using the GraphPad Prism 3.0a software package.

Cloning and Sequencing DNAzymes

Single-stranded DNAzymes were converted to duplex DNA by PCR amplification, usually with only the complementary strand primer *compl* (96°C for 3 min; followed by 3 cycles of 45 sec at 96°C, four min ramp cool to 50°C plus 45 sec at 50°C, and 1 min at 72°C; followed by 7 min at 72°C). When cloning cleaved and uncleaved **614**, the desired single-stranded DNA (either cleaved product or uncleaved reactant) was first isolated via PAGE/urea. PCR of the unpurified **614** reaction mixture containing both cleaved and uncleaved **614** with only *compl* primer yielded clones only of uncleaved **614**

molecules; cleaved **614** molecules were therefore cloned by PCR amplifying gel purified cleaved fragments using the *compl* primer with the 5'-truncated catalytic primer *cat.nt29-45*. Fresh double stranded PCR products were cloned using the TOPO TA Cloning System (InVitrogen) and plated on agarose plates containing ampicillin. Transformed cells were given only 15-30 min to recover in antibiotic-free media prior to plating on antibiotic containing plates (to prevent recently transformed clones from doubling prior to plating, favoring isolation of only unique species from the original pool).

Clones were transformed into the TOPO TA Cloning Vector and DNA was prepared from individual clones using the alkaline lysis protocol.

Clones were sequenced using the *1224* plasmid sequencing primer (3.2 pM, 5'-CGCCAGGGTTTTCCCAGTCACGAC) with 300-500 ng plasmid DNA template, 2X final concentration Big Dye reaction buffer and 2 μL Big Dye Terminator Sequencing mix in a final volume of 10 μL. Samples were overlaid with mineral oil and amplified by PCR (25 cycles, 96°C for 30 sec, 50°C for 15 sec, and 60°C for 4 min). Sequencing samples were analyzed on an Applied Biosystems Prism 310 Genetic Analyzer. Sequencing results were confirmed by examining the chromatograms manually using the "Sequencher" software package.

***In vitro* Selection**

DNAzyme libraries for the first round of selection were prepared by a single cycle of run-off PCR using the library template (5'-GTGCCAAGCTTACCGTCAC-N_{40}-GAGATGTCGCCATCTCTTCC (where N indicates equal molar concentrations of A, T, G, and C) and *cat+ribose* primer. Run-off reactions volumes ranged from 3 to 20 mL, each containing 1 ng/μL library template, 100 nM catalytic strand primer, 100 μM

dNTPs, 10 mM KCl, 20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 2 mM Mg_2SO_4, 0.1% Triton X-100, 20 units/mL *Vent exo-*, and 4.3 μCi/mL alpha-^{32}P-CTP. An aliquot (1 mL) of each sample mixture (minus polymerase) was placed in Eppendorf tubes (1.5 mL), heated (96°C, 8 min, then slowly cooled to 55°C over 30 min). Polymerase was added, and the samples were incubated (72°C, 15 min).

DNA was recovered by EtOH precipitation with NH_4OAc and glycogen as a carrier by storing overnight at -80°C and then centrifuging in Corex tubes (16,000 g, 40 min, 4°C). The EtOH was removed and the pellet re-suspended in water (200 μL). DNA was recovered by precipitation a second time with EtOH (NH_4OAc) and centrifugation (16,000 g, 20 min, 4°C). After removing the EtOH, the pellet was resuspended in 1X binding buffer (1 M NaCl, 1 mM EDTA, 50 mM HEPES pH 7) and bound to a streptavidin column. The unbound material was removed by flushing the column with wash buffer (50 mM HEPES pH 7). The complementary strand was removed by washing quickly with 0.2M NaOH, followed again by wash buffer.

The cleavage reaction was initiated by eluting the wash buffer and replacing it with reaction buffer (1M NaCl, 1 mM $MgCl_2$, 50 mM HEPES pH 7). The cleaved products were eluted from the column after two hours incubation. The eluted material was used in a PCR with the *cat+ribose* and *compl+5'P* primers to generate material for the next round of selection.

Table 3.1. Name, sequence, and description of oligonucleotides used throughout this chapter.

Names used in text	Description
ribose-**614**	5'– CTGCAGAATTCTAATACGACTCACTATrAGGAAGA CATGGCGACTCTCACATCATGCGAGCACACGCAAT AGCCTGATAAGGTTGGTAGTGACGGTAAGCTTGGC AC A DNAzyme isolated from the *in vitro* selection in this study. It catalyzes the cleavage of the ribo-adenosine embedded within the *cat*+*ribose* primer, either as a part of its own sequence, or in *trans*.
ribose-**614ΔC72T**	A mutant of *ribose*-**614** in which the cytosine at position 72 is replaced by a thymidine, resulting in a 30-fold lower rate of cleavage at 100 nM.
library template	5'– GTGCCAAGCTTACCGTCAC(n40)GAGATGTCGCCAT CTCTTCC The complementary strand template containing a region of 40 nucleotide randomized region flanked on both sides with constant regions, one complementary to *cat*+*ribose*, and the other identical to *compl*.
ribose-**library**	A random library generated by PCR amplification of the **library** template with *cat*+*ribose* and *compl* primers. Without selection, this library is predominately inactive. Selection of this library gave rise to DNAzyme **614**, and others. All sequences contain the *cat*+*ribose* primer and therefore have the potential to function as substrates for **614** cleavage.
ribose-**lib62**	An individual clone isolated from a the *ribose*-**library**. This sequence has no intrinsic ribonuclease activity of its own, but contains the *cat*+*ribose* primer and therefore can function as a substrate for **614** cleavage.
deoxyribose-**614,** *deoxyribose*-**lib62,** *deoxyribose*-**614ΔC72T**	The same sequences as *ribose*-**614,** *ribose*-**lib62,** and *ribose*-**614ΔC72T** but with the ribo-adenosine at position 28 replaced by a *deoxyribose*-adenosine (via PCR amplification with *cat*+*deoxyribose* primer). *deoxyribose*-**614** is still capable of cleaving the ribo-adenosine embedded within the *cat*+*ribose* primer, but only in *trans*.

Table 3.1 Continued

Names used in text	Description
compl-**614**	The complementary strand to *deoxyribose*-**614**.
cat+ribose	5'–CTGCAGAATTCTAATACGACTCACTATrAGGAAGACATGGCGACTCTC Primer used to generate full length molecules from template, thus incorporating a ribo-adenosine (rA) at position 28. This oligo has no catalytic activity of its own and was therefore used as a substrate in *trans* cleavage assays.
cat+deoxyribose	The same oligonucleotide sequence as *cat+ribose* but with a deoxyribonucleotide replacing the rA. This primer was used to generate full length DNAzymes without the ribo-adenosine moiety. DNAzymes incorporating *cat+deoxyribose* can not be cleaved and therefore act as a true enzyme.
compl	5'–GTGCCAAGCTTACCGTCA This primer is complementary to the 3'-end of the catalytic strand of the full length DNAzymes used in this study, and is used to generate double stranded DNAzymes via PCR.
compl+5'P	This primer is the same sequence as the *compl* primer but with a phosphate attached to the 5'-position. This primer is used to generate double stranded DNAzyme via PCR, followed by degradation of the complementary strand by lambda exonuclease (an enzyme that specifically degrades 5'-phosphorylated DNA).
compl+tail	This primer is the same sequence as the *compl* primer but with a 15-nt poly-adenosine attached to the 5'-position via an 18-atom hexaethyleneglycol-based linker. Polymerase can not efficiently read through this linker, so PCR using this primer genearates a double stranded DNAzymes with a complementary strand that is 15-nt longer than the catalytic strand, thus allowing purification of the catalytic strand via PAGE/urea.
chase	The *cat+deoxyribose* oligo was used as the chase substrate for cleavage assays with **614** functioning as the enzyme.

Table 3.1 continued

Names used in text	Description
28-nt product	The 28-nt product generated from cleavage of *ribose*-**614** (106-nt).
78-nt product	The 78-nt product generated from cleavage of *ribose*-**614** (106-nt).
cat.nt29-45	5'-GGAAGACATGGCGACTCTC A 5'-truncated version of *cat+deoxyribose* (containing nucleotides 29 - 45) used to generate dsDNA of the 78-nt *ribose*-**614** cleavage product for subsequent cloning.

Results

In vitro Selection

An *in vitro* selection experiment to select for a DNAzyme having ribonuclease activity was conducted using a procedure adapted from Breaker and Joyce (Breaker & Joyce, 1995). A library was constructed containing a 5'-primer constant region and a substrate segment, which contained an internal ribo-adenosine at position 28. This was connected to a 3'-primer constant region by a segment, 40 nucleotides in length, having random sequence. The 5'-constant region used by Breaker and Joyce was altered to minimize a base paired "clamp" engineered to favor association between the substrate portion and random portion (see Figure 3.2a) of the DNAzyme. Table 3.1 presents a summary of the many oligonucleotides discussed in this chapter.

A cycle of selection and amplification was applied for 7 rounds to the library. A sampling of the DNA molecules recovered was then cloned and sequenced. Several DNA molecules having distinct sequences were isolated. When inspected for catalytic activity, the molecules cleaved themselves with apparent first order rate constants ranging

from 0.015 to 0.049 hr^{-1} (Figure. 3.2b). These rate constants are low in comparison to optimized DNAzymes obtained in the presence of divalent cation (ca. 60 hr^{-1})(Santoro & Joyce, 1997; 1998), but comparable to Mg^{++}-independent catalysts reported by Geyer and Sen (0.17 hr^{-1}) and by Faulhammer and Famulok (0.006-0.024 hr^{-1}). Additional rounds of selection isolated variants of the sequences initially identified in round 7.

The behavior of one catalyst, termed **614**, isolated from the seventh round of selection was examined. DNAzyme **614** was chemically synthesized and purified by polyacrylamide gel electrophoresis (PAGE) by Integrated DNA Technologies (Coralville IA). Large quantities of internally radiolabeled **614** were generated by PCR using this **614** template. The catalytic strand of the double stranded PCR products, termed *ribose*-**614**, was separated from its complement by two methods. In some cases, *ribose*-**614** was generated with a 5'-phosphorylated complementary strand primer (*compl+5'P*). Following PCR, the complement was digested with lambda exonuclease. Alternatively, *ribose*-**614** was generated using a complementary strand primer with a 15-nt poly-adenosine tail attached to its 5' end via a PEG linker (*compl+tail*). By increasing the length of the complementary strand, the tail allowed PAGE separation of the catalytic strand from its complement.

Purified *ribose*-**614** was resuspended in HEPES buffer, and cleavage reactions were initiated by adding equal volume of 2X reaction buffer, immediate heating to 96°C for 3 min, and slow cooling to 23°C over 10 min (the set-up). The reaction was then followed by gel electrophoresis. *Ribose*-**614** (100 nM) cleaves itself at a rate of 0.015 hr^{-1}, slow in comparison to previously isolated DNAzymes, but still significantly faster than the

uncatalyzed reaction (the uncatalyzed cleavage rate is negligible in 300 hours, estimated from the cleavage of *ribose*-**library** shown in Figure 3.5)

Cleavage of 614 does not go completion

A progress curve for *ribose*-**614** cleavage shows that the product formation ceased after ca. 65% conversion; this is a "cleavage plateau" (Figure 3.3a). Such plateaus are frequently seen in selected DNAzymes and RNAzymes, but are rarely explained. We considered six possible explanations for the failure of the cleavage reaction to go to completion:

1. The complementary strand may have been incompletely removed in the set-up. This strand may inhibit the reaction, or may erroneously increase the estimate of the uncleaved product if the radiolabeled complementary strand is assigned to the reactant;
2. Part of the substrate may be missing the adenine riboside that offers the cleavable site, due to failure in the synthesis of the *cat*+*ribose* primer;
3. The reaction might be reversible, with the 65% cleavage representing the achievement of the equilibrium between uncleaved substrate and cleaved products;
4. The products of the reaction may inhibit **614**;
5. Part of **614** may adopt a conformation that is inactive;
6. Part of the **614** might no longer have the correct sequence, having incurred mutation during the synthesis of the template or amplification. We examined each possibility.

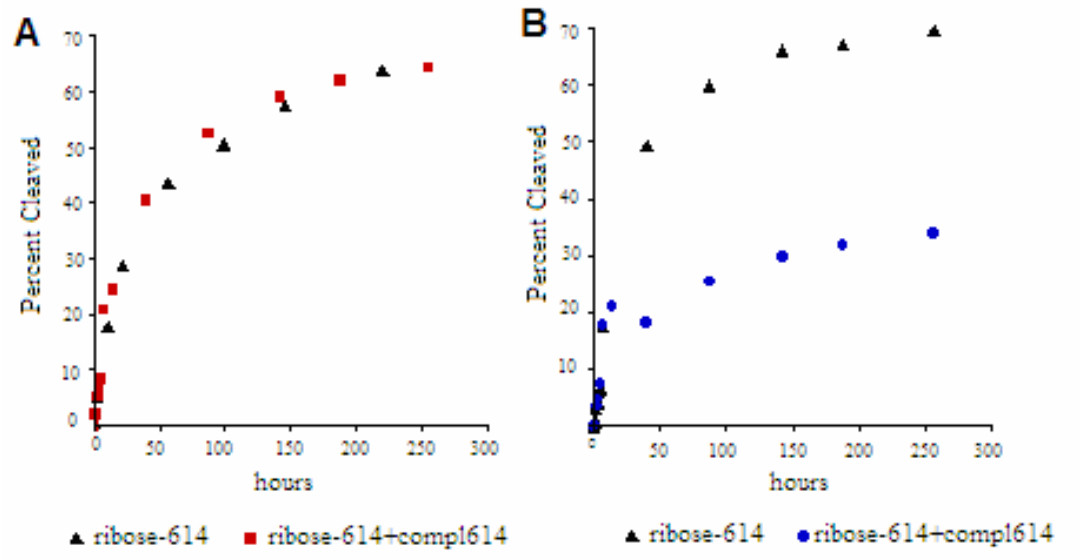


Figure 3.3. *Ribose*-**614** cleavage is unaffected by the addition of 10% complementary strand. Addition of equimolar amounts of the complement greatly lowers cleavage rate and plateau. (a) Cleavage profile for *ribose*-**614** (100 nM)(triangles), or *ribose*-**614** (115 nM) plus complementary strand (11.5 nM)(squares). (b) Cleavage profile for *ribose*-**614** (222 nM)(triangles), or *ribose*-**614** (200 nM) plus equimolar complementary strand (circles).

Inhibition by Incompletely Removed Complementary Strand does not Account for the Plateau

The complement of **614** (termed *compl*-**614**) was added in small amounts to *ribose*-**614** ([*ribose*-**614**] = 115 nM, [*compl*-**614**] = 11.5 nM). The initial cleavage rate and plateau were unchanged (Figure 3.3a). Similar results were seen at a lower concentration of *ribose*-**614** (11.5 nM) plus *compl*-**614** (1.15 nM). Experiments with equal amounts of *ribose*-**614** (200 nM) and *compl*-**614** (200 nM) found that the plateau was dramatically lower (Figure 3.3b), even though the initial cleavage rate was unchanged. Similar results were seen with *ribose*-**614** at 50 nM plus *compl*-**614** at 50 nM. This showed that the complementary strand inhibited cleavage by **614**, and the importance of removing the complementary strand prior to measurement of DNAzyme kinetics.

We then asked whether the presence of complementary strand could explain the incomplete cleavage of *ribose*-**614**. Two independent procedures (exonuclease digestion

and asymmetric PCR) were used to generate *ribose*-**614** lacking its complement. Experiments with each yielded approximately the same cleavage plateau. This suggested that the plateau is not the consequence of incomplete removal of *compl*-**614**, as it is unlikely that the amounts of complement remaining following the two procedures are the same. In fact, it is difficult to imagine that any *compl*-**614** remains following the purification using the asymmetric PCR procedure, as the complementary strand is 15-nt longer than the catalytic strand, and would almost certainly be removed by PAGE purification.

To assess the amount of complementary strand remaining after exonuclease treatment, we treated *ribose*-**614** generated via the exonuclease method with base (0.5 M NaOH, 80°C, 1 h). This cleaves all of the substrate at the adenine riboside site. The amount of cleaved products was similar to that with the 5'-^{32}P labeled *cat*+*ribose* primer (85-95% cleaved for *ribose*-**614** *vs.* 90-97% cleaved for *cat*+*ribose* primer following treatment with base). The incomplete cleavage of *cat*+*ribose* primer by base suggest a lack of chemical susceptibility of the primer (discussed below). Since contaminating complementary strand cannot be present in the *cat*+*ribose* primer, the ca. 5% difference between base cleavage of the *cat*+*ribose* primer and the full length *ribose*-**614** is an estimate of the upper limit of the amount of complementary strand that might remain from incomplete degradation by exonuclease.

The efficiency with which lambda exonuclease removes the complementary strand was also tested using a 5'-^{32}P labeled *cat*+*ribose* primer to synthesize full-length, double stranded *ribose*-**614** without 5'-phosphate on the complementary strand primer. The duplex PCR product was divided into two aliquots. One was subjected to standard

exonuclease treatment, while the other sample was untreated. The samples were resolved by PAGE-urea to determine the amount of full-length product. Less than 10% of the original phosphorylated strand (compared to untreated sample) remained following exonuclease digest. These results rule out contamination by residual *compl*-**614** as a major cause of incomplete **614** cleavage .

An Approach to Chemical Equilibrium does not account for the Plateau

The failure of a substrate to be completely transformed to product may result from an approach to chemical equilibrium between substrates and products in a reversible reaction. To test for this, *ribose*-**614** (200 nM) was incubated for 144 hours. At this time, plateau had been reached, and the reaction mixture was divided into three aliquots. One aliquot was reserved. The second was diluted 25-fold with reaction buffer. Diluting is expected to drive the equilibrium towards the cleaved product. An equal amount of a variant of *ribose*-**614**, *deoxyribose*-**614**, was added to the third aliquot. The compound *deoxyribose*-**614** is the same sequence as *ribose*-**614**, but with an uncleavable uncleaveable adenine 2'-deoxyriboside replacing the cleavable adenine riboside at position 28. As discussed below, *deoxyribose*-**614** is a true catalyst, acting in *trans*. Its addition in excess to *ribose*-**614** ensured that catalytic activity was sufficient to see cleavage at the plateau if it was occurring. Neither treatment significantly altered the plateau, excluding reversibility as its explanation.

In a second experiment, the 78-nt product of *ribose*-**614** cleavage was isolated via gel purification. This product was radiolabeled and, following gel purification, incubated (in excess) with full-length *ribose*-**614** (unlabeled). As the unlabeled *ribose*-**614** becomes cleaved, and if the reverse reaction (ligation) occurs, excess 78-nt ensures that it may capture by ligation the 28-nt product, converting 78-nt to full-length *ribose*-**614** over

time. The reaction was monitored for 400 hours. No conversion of 78-nt to full-length material was observed. This also excludes reversibility as an explanation for the plateau.

Testing if the Cleavage Products are Acting as Catalysts or Inhibitors

By incorporating a short self-complementary segment that encourages a hairpin to clamp together the substrate and catalytic portions of the DNAzyme, Breaker and Joyce hoped to increase the chance that their library would contain molecules that self-cleave rapidly in *cis*. Further, *cis*-cleavage might be expected to predominate over *trans* cleavage, because the substrate is covalently bonded to the catalyst. The possibility that either cleaved product continues to act as a ribonuclease in *trans* or as an inhibitor cannot be ruled out. This is especially true for **614**, where the clamp that might favor a hairpin, and therefore *cis*-cleavage, is not present.

To test this, the 28- and 78-nt products were gel purified. These products were added in equal amounts to a sample of uncleaved *ribose*-**614**. Cleavage of *ribose*-**614** occurred at the same rate in the presence or absence of the 28- and 78-nt fragments (Figure 3.4). This suggests that 28- and 78-nt do not act as inhibitors or catalysts, at least when both are present in stoichiometric amounts (approximately the highest concentration they reach under normal **614** cleavage conditions).

The 28- and 78-nt products were also added to a sample of the *ribose*-**library** used in the selection. The *ribose*-**library** contains the *cat+ribose* primer followed by a segment of random sequence. Thus, it contains substrates that are (for the most part) not catalysts, and serves as the opposite of *deoxyribose*-**614**, which is a catalyst but not a substrate. The *ribose*-**library** alone does self-cleave to a detectable extent, as expected given that an unselected, random library has very few active catalysts.

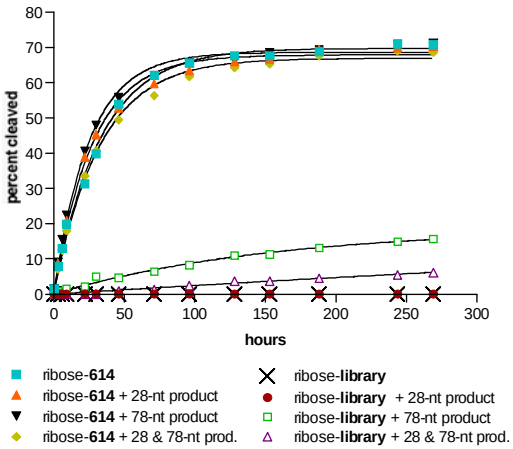


Figure 3.4. Cleavage products do not affect *ribose*-**614** cleavage, and only minimally alter *ribose*-**library** cleavage. *Ribose*-**614** or *ribose*-**library** were incubated either alone (400 nM), or with the addition of the 28-nt product of **614** cleavage (400 nM), the 78-nt product of **614** cleavage (400 nM), or both (400 nM each). Data are fit to a single exponential curve. Errors in percent cleaved are $\leq \pm 2$ percentage points.

The 28-nt product alone did not cleave *ribose*-**library**, showing that this product is not a catalyst. The 78-nt product alone did cleave *ribose*-**library**, although with a very low rate constant (k_{obs} = 0.0006 hr^{-1}, 35-fold lower than *ribose*-**614** under similar conditions). The 28-nt product with the 78-nt fragment reduced cleavage of the *ribose*-**library** below that observed with the 78-nt fragment alone. This inhibition is presumably because 28-nt product competes with the *ribose*-**library** substrate in binding to the 78-nt catalyst. Inhibition of **614** by 28-nt is not noticeable presumably because 28-nt does not compete as effectively for **614** as a second molecule of **614** does.

Improperly Folded ribose-614 Accounts for Part of the Plateau

We then asked whether *ribose*-**614** folds into active and inactive conformations, with the inactive form contributing to the cleavage plateau. Predictions of the conformation of *ribose*-**614** using *Mfold* (Zuker, 2003) showed several potential structures. Close inspection of *ribose*-**614** on a non-denaturing gel suggested two bands

present in the starting material, one converting to a third band upon incubation under reaction conditions, and the other remaining unchanged. This is consistent with the hypothesis that *ribose*-**614** adopts active and inactive conformations.

To test this, *ribose*-**614** was incubated until the plateau was reached (140 h). The mixture then was diluted into 2 volumes of formamide and heated (90°C, 2 min). The uncleaved *ribose*-**614** was purified by PAGE/urea, resuspended in buffer, re-folded using the slow cooling protocol, and subjected to cleavage conditions a second time. An additional 25% of the *ribose*-**614** sample was cleaved in the 300 hours following gel purification. This additional cleavage following gel purification suggests that some of the initially uncleaved *ribose*-**614** was in an inactive conformation. It is notable, however, that the gel-purified *ribose*-**614** reached a plateau of ca. 25%, significantly lower than the cleavage plateau of the original sample (Figure 3.5).

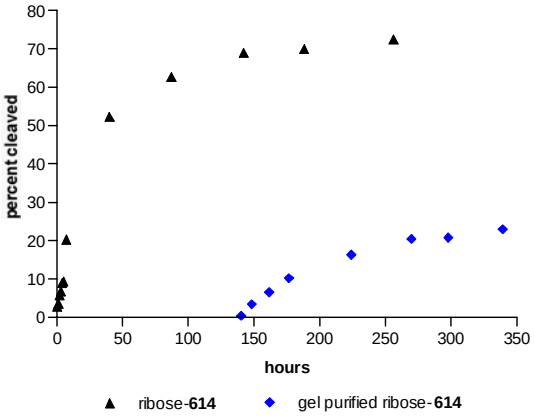


Figure 3.5 Gel-purification of *ribose*-**614** at cleavage plateau results in additional cleaveage, indicating a fraction of **614** is folded into an inactive conformation. *Ribose*-**614** (222 nM) was allowed to self-cleave for 140 hours, nearing cleavage plateau (triangles, top curve), at which point half the sample was resolved with denaturing PAGE. The fraction of *ribose*-**614** remaining uncleaved at 140 hours was gel-purified, resuspended in reaction buffer (to a final concentration of 100 nM) and incubated for additional time (diamonds, bottom curve; percentage cleaved as a fraction of the label in the gel purified product, not of initial substrate). Errors in percent cleaved are $\leq \pm 2$ percentage points

If misfolding were the only cause of incomplete cleavage, a plateau of approximately 65% would be expected in the second round of cleavage. The fact that the gel-purified sample reaches a lower plateau indicates that misfolding may account for only ca. 9% of total uncleaved fraction in the original sample (as an additional 25% of the sample cleaved following gel purification, and ca. 35% of the original sample remains uncleaved, the fraction of the initial sample that is misfolded is 25% of 35%, or 9%).

We considered the possibility that gel-purification removed an inhibitor of cleavage. *Ribose*-**614** was incubated to reach the plateau (141 h), and an aliquot was heated and slowly cooled. Any molecule in an inactive conformation was thus given another chance to adopt an active conformation (Figure 3.6). As in the gel-purification experiments, denaturing and refolding via this procedure increases the amount of material cleaved. The cleavage plateau over the first 150 hours following re-heating is about 10% higher than the untreated sample (in agreement with the estimate of 9% above), but still well below the ca. 90% cleavage when treated by strong base. When untreated samples are incubated beyond 300 hours, the cleavage levels approach those seen when samples are reheated at 140 hours, suggesting that re-folding into active conformations may occur slowly at room temperature. Thus, alternative conformations account for some, but not all, of the plateau seen between 150 and 300 h.

Mutations Introduced into 614 during Cloning and Sequencing Account for part of the Plateau

The ability of base to cleave ca. 94% of the *cat+ribose* primer suggests that about 6% of the *cat+ribose* primer, and therefore any full length DNAzyme made from the *cat+ribose* primer, may be missing the adenine riboside unit that is the site for cleavage.

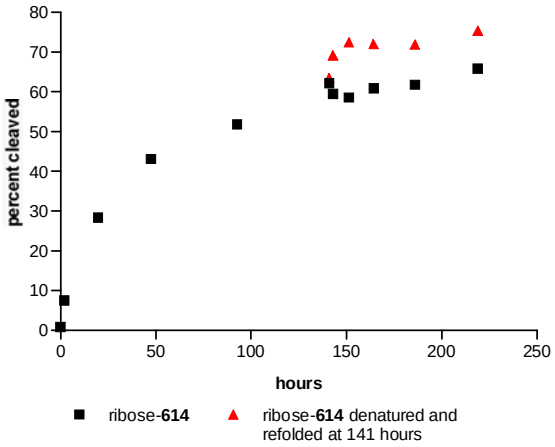


Figure 3.6. Reheating *ribose*-**614** results in additional cleavage, indicating a fraction of *ribose*-**614** is folded into an inactive conformation. *Ribose*-**614** (100 nM) was allowed to react until it reached cleavage plateau (141 h), at which point half of the sample was denatured by heating to 96°C for 3 min, and slowly cooled to 23°C over 10 min. Errors in percent cleaved are $\leq \pm 2$ percentage points

Other sequence variants may be present throughout *ribose*-**614** as well, either as a result of mutations introduced during the synthesis of the primers and template, or mutations introduced by the polymerase during PCR. These mutations may reduce or eliminate the catalytic power of a fraction of the *ribose*-**614** DNAzyme pool.

This possibility was examined by cloning and sequencing DNA molecules from the cleaved and uncleaved fractions of *ribose*-**614** after the plateau had been reached (Figure 3.7). Of the 29 sequenced clones of the 78-nt cleaved product of *ribose*-**614**, only four (14%) were found to be mutants. The mutations, at three sites, were in the N40 region between the primers.

In contrast, of the 66 clones from the portion of ribose-614 that remained uncleaved at the plateau, only 64% had at least one mutation (42 individuals). Some 18% had a mutation in the N40 region, 5% had a mutation in the complementary strand primer, and 48% had a mutation in the catalytic strand primer (including 20% of the uncleaved *ribose*-**614** molecules sampled that were missing the adenine riboside cleavable site).

Excluding molecules missing the adenine riboside cleavable site, 44% of the uncleaved molecules possessed a mutation. This corresponds to 15% of the total initial population (44% of ca. 35% of total initial population remaining uncleaved at plateau).

The extent of misfolding in **614** was estimated by assuming that all unmutated sequences from the uncleaved pool at the plateau remained uncleaved because they had adopted an inactive conformation. Given this, 13% of the total population is calculated to be misfolded (36% of the unreacted material is not mutated, and the unreacted material at plateau is ca. 35% of the total).

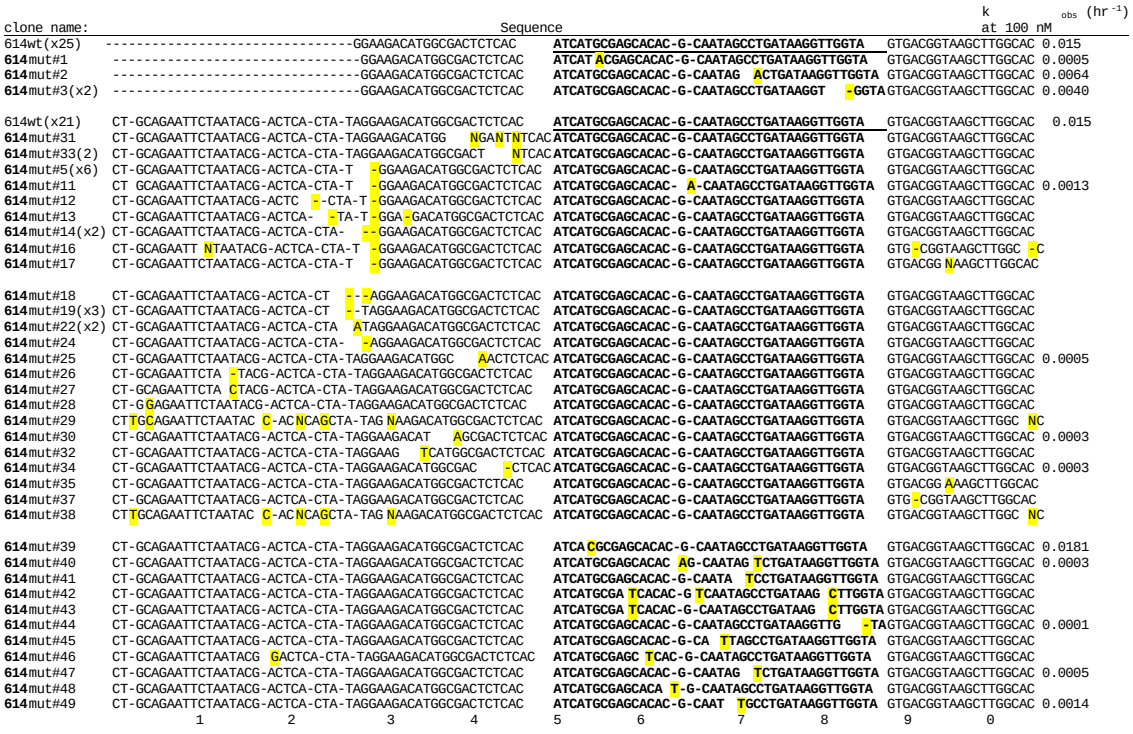


Figure 3.7. Sequence alignment of cleaved and uncleaved cloned **614** sequence variants. Mutations of **614** are highlighted in yellow; "N40" region between primers is bold in each sequence and underlined in the original (no mutation) **614** sequence. Apparent first order catalytic rate constants (at 100 nM, in units of hr^{-1}) are shown to right for a sample of mutant-**614** DNAzymes.

This is consistent with the estimate above based on the outcome of a cycle of unfolding and refolding.

Several of the **614** mutants were tested to see if the mutations in fact reduced the rate of self-cleavage. Eleven of the twelve **614** mutants that were tested showed cleavage rates reduced by 10 to 100 fold compared to *ribose*-**614**.

***Ribose*-614 Catalysis is not Mg^{++}-dependent**

The mono and divalent cation requirements for *ribose*-**614** activity were then examined. Reducing the concentration of NaCl from 1 M to 0.1 M eliminated cleavage. Only a modest change in the rate of cleavage of *ribose*-**614** (100 nM) was observed when the reaction was run in the absence of $MgCl_2$ (with and without 1 mM EDTA, initial rate 1.32×10^{-2} hr^{-1} and 1.37×10^{-2} hr^{-1}, respectively) or in the presence of $MgCl_2$ (1 mM, 1.53×10^{-2} hr^{-1}; 10 mM, 1.70×10^{-2} hr^{-1}; or 100 mM, 2.52×10^{-2} hr^{-1}). Catalysis by **614** therefore does not require Mg^{++} in significant amounts even though **614** was selected in the presence of $MgCl_2$ (1 mM). Requirement of a trace of Mg^{++} cannot, of course, be excluded. The rates of two other catalysts isolated from this selection (*ribose*-**62** and *ribose*-**616**) were also largely unchanged when the $MgCl_2$ (1 mM) of the reaction buffer was replaced with EDTA (1 mM). For *ribose*-**625**, another molecule isolated from the selection, cleavage was eliminated by replacement of $MgCl_2$ by EDTA at 25°C.

Ribose*-614 Cleaves in *trans

A series of experiments was then performed to examine the rate of cleavage as a function of concentration. Self-cleavage is expected to be kinetically first order. Cleavage in *trans* is expected to be kinetically second order, meaning that an apparent first order rate constant k_{obs} will not be independent of [DNAzyme], and fall to zero as [DNAzyme] falls to zero. A convenient way to separate simultaneous intra and intermolecular

processes is to plot the apparent k_{obs} as a function of [DNAzyme]. The *y*-intercept is k_{uni} for the unimolecular reaction, while the slope of the line, reflects a second order process.

Figure 3.8 shows a plot of the apparent k_{obs} versus [*ribose*-**614**]. At low [*ribose*-**614**], the rate is independent of [*ribose*-**614**]; primary cleavage profiles are shown in Figure 3.9. At higher concentrations, however, the apparent k_{obs} increases with increasing [*ribose*-**614**]. These results are consistent with a model that includes both unimolecular and bimolecular processes.

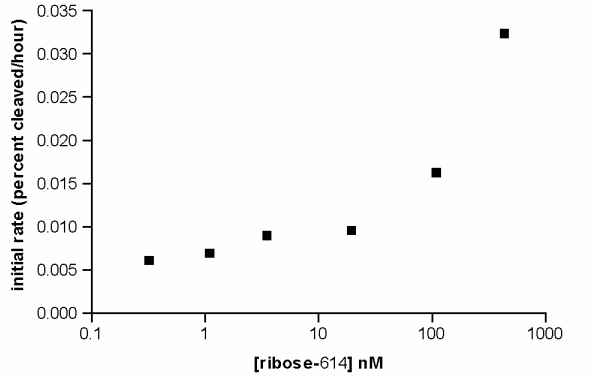


Figure 3.8. Initial rate of *ribose*-**614** cleavage as a function of [*ribose*-**614**]. At higher concentrations, a second order process is apparent. The rate constant for the apparent first order process (uncorrected for the cleavage plateau), from the *y* intercept, is ca. 0.006 hr^{-1}.

A best-fit line extrapolated to infinite dilution for **[***ribose*-**614**] 3.5 nM and below (the region where the process is mostly unimolecular) gives an intercept corresponding to a unimolecular rate constant of 0.006 hr^{-1}. This is an underestimate because the rate constant, based on initial rates, is uncorrected for the cleavage plateau (see below).

An alternative way to obtain a rate constant for a first order process plots the log of uncleaved [*ribose*-**614**] (substrate remaining) versus time. Here, the progress of cleavage at low [*ribose*-**614**] fit a linear model well, while the progress at higher concentrations did not (Table 3.2). This suggests that below 20 nM, a unimolecular rate process

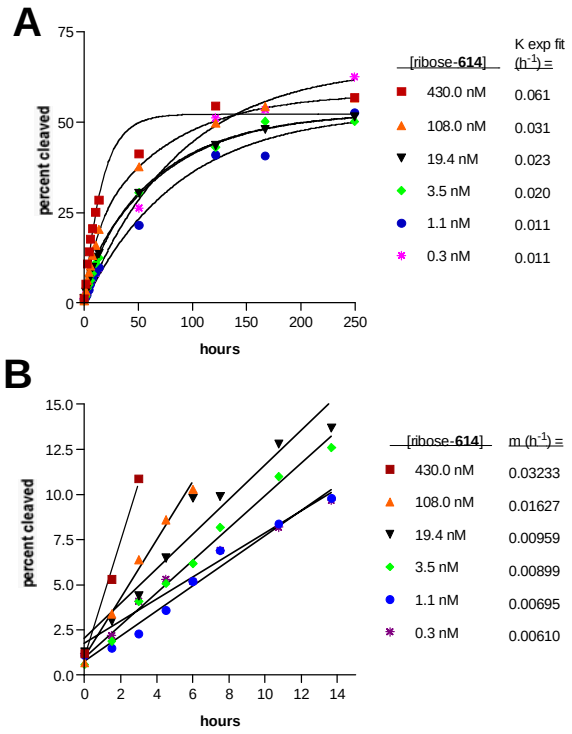


Figure 3.9. *Ribose*-**614** cleavage rate is concentration dependent. (A) Complete time course for *ribose*-**614** at various concentrations. Rates are estimated based on a fit to a single first-order exponential equation. (B) Linear initial phase of time course. Rates are estimated based on the slope of best fit line.

dominates, while a bimolecular rate process contributes above 20 nM. The rate constant estimated in this manner for [*ribose*-**614**] at 0.3 nM was 0.0042 hr^{-1} (uncorrected for the plateau), in good agreement with the estimate above.

Table 3.2. Data from plot $\ln[S]_t$ versus time for ribose-614 cleavage at various concentrations fit to a linear equation.

[*ribose*-**614**] nM	430	108	19.4	3.5	1.1	0.3
- (slope of $\ln[S]_t$)	0.0032	0.0035	0.0030	0.0031	0.0030	0.0042
R^2	0.78	0.89	0.92	0.90	0.97	0.97

We then asked if the progress curve for cleavage at high and low [*ribose*-**614**] fit best to a single exponential (in which both the rate and plateau can vary) or the sum of two exponential equations (in which both the rate and plateau can vary for each of two

equations). At *ribose*-**614** concentrations ≤3.5 nM, progress curves fit best to a single exponential. At higher concentrations, the progress curve is fit best by the sum of two independent exponentials. This suggests that a single process dominates transformation of *ribose*-**614** at low concentrations. If multiple active conformers exist at low [*ribose*-**614**], they are either in rapid equilibrium, or self-cleave with comparable rate constants. The first order constant was then corrected given information collected above about the plateau. After correction, a single exponential curve fit to the progress curve for 0.3 nM *ribose*-**614** cleavage under unimolecular conditions. A rate constant 0.011×10^{-2} hr^{-1} was estimated (with a plateau of 61%).

Various ribose-containing Substrates are Cleaved by *deoxyribose*-614

To establish *trans* cleavage in this system, internally labeled *ribose*-**614** (which yields the 78-nt fragment as the only labeled product) was incubated with 5'-^{32}P-labeled *cat*+*ribose* primer (100 nM). The *cat*+*ribose* contains the substrate domain of *ribose*-**614**, but is not catalytically active. Here, *cat*+*ribose* is cleaved, as is *ribose*-**614** (Figure 3.10, right panel). This establishes *trans* cleavage by *ribose*-**614**. The cleavage of *ribose*-**614** (100 nM) in the presence of *cat*+*ribose* primer (100 nM) was lower than *ribose*-**614** incubated alone (at 100 nM), suggesting that the *cat*+*ribose* primer competes with *ribose*-**614** for *trans* cleavage by *ribose*-**614** (Figure 3.11).

Trans cleavage was also tested by challenging *deoxyribose*-**614** (a modified form of *ribose*-**614** in which the adenine riboside was replaced by a non-cleavable deoxyadenosine) with various substrates, each containing the *cat*+*ribose* sequence. These included (a) the *cat*+*ribose* primer, (b) a pool of molecules incorporating the *cat*+*ribose* primer followed by a random region and a 18-nt constant region (*ribose*-**library**), which

collectively have no detectable catalytic activity, (c) a single clone from the *ribose*-**library**, (*ribose*-**lib62**), which also has no detectable catalytic activity, and (d) a mutant of **614** (**614ΔC72T**) that has ca. 30 fold reduced activity compared to **614**, at 100 nM.

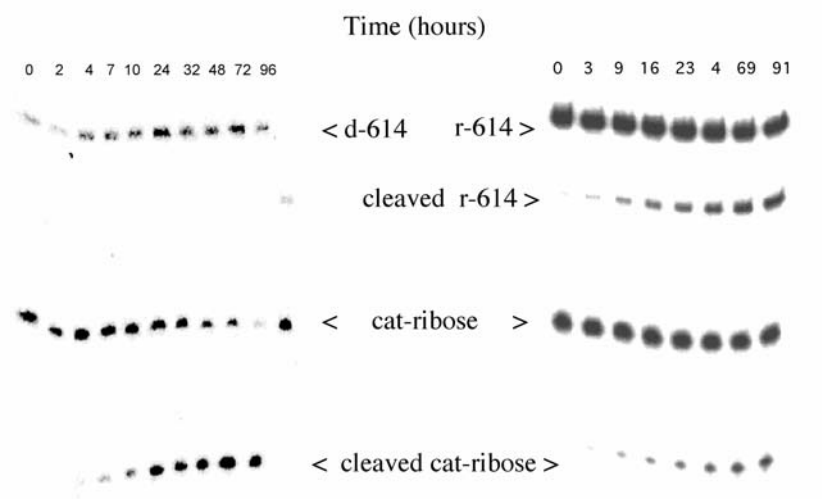


Figure 3.10. Both deoxyribose-614 (left panel) and ribose-614 (right panel) cleave cat+ribose primer in *trans*. Internally labeled deoxyribose-614 (100 nM) or ribose-614 (100 nM) were mixed with 5'-end labeled cat+ribose (100 nM) and incubated under reaction conditions. The deoxyribose-614 does not contain the ribo-adenosine and therefore does not cleave itself. Unless otherwise stated, all kinetics were performed under standard reaction conditions, namely 1M NaCl, 1 mM MgCl2, 50 mM HEPES, 25°C, and errors in percent cleaved are < ±2 percentage points.

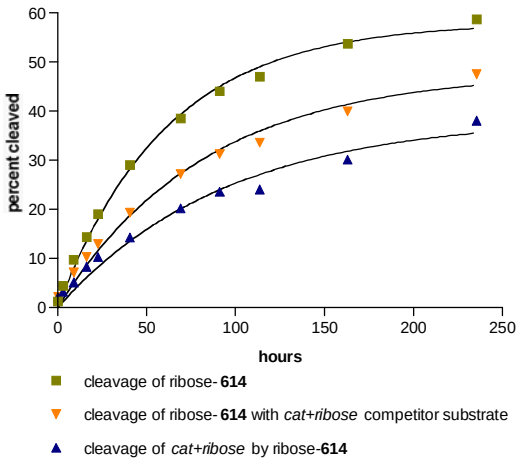


Figure 3.11. *Cat+ribose* competes with *ribose*-**614** for cleavage. *Ribose*-**614** (100 nM) was incubated with and without *cat+ribose* (100 nM). *Cat+ribose* is not catalytic and when incubated alone is not cleaved. Data are fit to a single first-order exponential equation.

Each substrate was cleaved by deoxyribose-614 (Figure. 3.12) although with different efficiencies. *Ribose*-**614ΔC72T** (which should fold like *ribose*-**614**) is cleaved by *deoxyribose*-**614** more slowly (suggesting that *ribose*-**614ΔC72T** may fold so as to make the substrate domain inaccessible to catalysts), while the *cat+ribose* primer is cleaved faster. The fact that the small *cat+ribose* primer is better than full length substrates is consistent with the view that folding in longer substrates inhibits *trans* cleavage by *deoxyribose*-**614**. The unimolecular and bimolecular mechanisms by which **614** can cleave are summarized in Scheme 3.1.

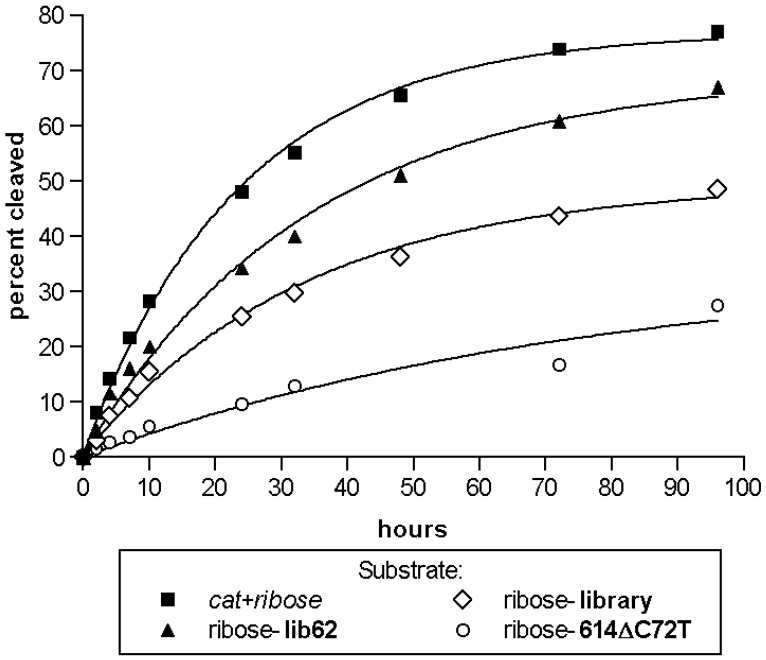


Figure 3.12. Compound *deoxyribose*-**614** can cleave various substrates in *trans*. Unlabeled *deoxyribose*-**614** (270 nM) was incubated with various radiolabeled substrates (*cat+ribose*, closed squares; *ribose*-**lib6lo2**, closed triangles; *ribose*-**library**, open diamonds; or *ribose*-**614ΔC72T**, open circles) each at 7.5 nM). Each substrate possessed the same 5'-primer sequence containing the adenine riboside. Compound *deoxyribose*-**614** does not cleave itself, and percentage cleaved is the fraction of the substrate converted to product. Data are fit to a first order exponential curve. Data points being accurate to $\leq \pm 2$ percentage. Errors in the measurement of time are ±1 minute.

Unimolecular Kinetic Scheme:

$$614_{unfolded} \underset{k_{-1\ (uni)}}{\overset{k_{1\ (uni)}}{\rightleftharpoons}} 614_{folded} \xrightarrow{k_{cat\ (uni)}} Products$$

Bimolecular Kinetic Scheme:

$$E + S \underset{k_{-1\ (bi)}}{\overset{k_{1\ (bi)}}{\rightleftharpoons}} E{\cdot}S \xrightarrow{k_{cat\ (bi)}} E{\cdot}P$$

E = *ribose*-**614** or *deoxyribose*-**614**
S = *ribose*-**614**, *ribose*-**lib62** or *cat*+*ribose*

Scheme 3.1. **614** can cleave via a unimolecular and bimolecular mechanism.

Competition Studies of *Ribose*-614 Cleavage

We then tested the ability of three of the above substrates, each with their ribose replaced by 2'-deoxyribose (*cat*+*deoxyribose* primer, *deoxyribose*-**614ΔC72T**, and *deoxyribose*-**lib62**), to compete with *ribose*-**614** for self-cleavage. These competitors were added in 9-fold excess over *ribose*-**614**. For comparison, 9-fold excess of *deoxyribose*-**614** was added to *ribose*-**614**.

The addition of *deoxyribose*-**614** (270 nM) to *ribose*-**614** (30 nM) increased the rate of *ribose*-**614** cleavage to the same level as seen for *ribose*-**614** at 300 nM (Figure 3.13).

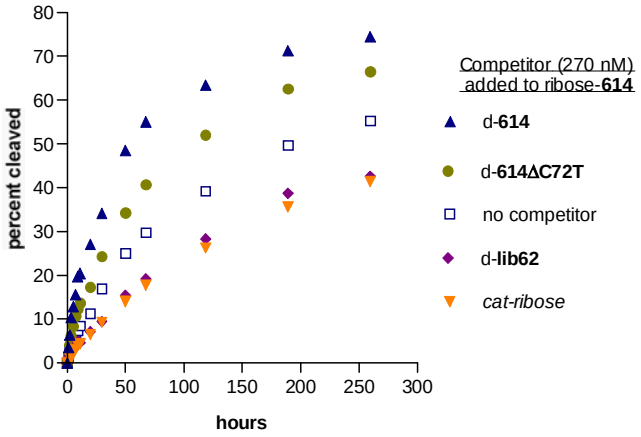


Figure 3.13. Various substrates can compete with *ribose*-**614** for self-cleavage. A nine fold excess of unlabeled *deoxyribose*-**lib62**, *deoxyribose*-**614**, *deoxyribose*-**614ΔC72T**, or *cat*+*deoxyribose* was added to radiolabeled *ribose*-**614** (33 nM) at time zero. Errors in percent cleaved are < ±2 percentage points.

The cleavage of *ribose*-**614**, however, was reduced by excess *cat*+*deoxyribose* or *deoxyribose*-**lib62**. This showed that these molecules compete for *ribose*-**614** self-cleavage.

Curiously, addition of a 9-fold excess of *deoxyribose*-**614ΔC72T** to *ribose*-**614** did not reduce the rate of *ribose*-**614** cleavage, but rather increased it. The ability of *deoxyribose*-**614ΔC72T** to cleave *ribose*-**614** may indicate that the mutation alters the ability of *deoxyribose*-**614ΔC72T** to function as a substrate without inactivating its catalytic domain, consistent with the observation above that *ribose*-**614ΔC72T** is itself a poor substrate for *deoxyribose*-**614** cleavage.

The competition study above was done at high [*ribose*-**614**] to favor *trans* cleavage. A parallel study at low concentrations of *ribose*-**614** (favoring unimolecular cleavage) showed no inhibition of *ribose*-**614** self-cleavage by *deoxyribose*-**lib62** competitor.

Saturation Kinetics in *trans* cleavage by *deoxyribose*-614

Two mechanisms for the *trans* bimolecular reaction were considered. In the first, *deoxyribose*-**614** cleaves its substrate on every encounter. This gives a linear slope in a plot of an apparent rate constant vs. [catalyst] or [substrate] over the entire concentration range.

An alternative second order process is possible, however. Here, substrate and catalyst form a complex that can dissociate before it reacts. This gives saturation Michaelis-Menten kinetics.

To search for saturation kinetics, we exploited *deoxyribose*-**614** as a DNAzyme that cannot act upon itself, but must act in *trans*. Concentrations of *deoxyribose*-**614** were scanned to find a range where the apparent second order rate constant reaches a plateau reflecting saturation. Various substrates were used, including *cat*+*ribose*, *ribose*-**lib62**,

and *ribose*-**614**, at low amounts relative to *deoxyribose*-**614**. Cleavage rates did not increase (2000 to 6000 nM *deoxyribose*-**614**), suggesting saturation of the bimolecular process.

Assuming that unbinding is fast compared to reaction, a calculated fit gave a dissociation constant (K_d) of 29.0, 37.3, and 25.5 nM for *cat*+*ribose* primer, *ribose*-**614** and *ribose*-**lib62** substrates, respectively. The $k_{cat(bi)}$, determined from the maximum rate of cleavage with saturating enzyme, was 0.056, 0.048, and 0.049 hr^{-1} for *cat*+*ribose*, *ribose*-**614** and *ribose*-**lib62** substrates, respectively.

Compound *deoxyribose*-614 Cleaves with Multiple-turnovers

To show that *deoxyribose*-**614** catalyzes cleavage with multiple turnovers, either *cat*+*ribose* or *ribose*-**lib62** substrate were incubated in 4-fold excess over *deoxyribose*-**614** (133:33 and 400:100 nM substrate:enzyme). With 400 nM enzyme and 100 nM substrate, two turnovers of substrate were observed at 100 hours for *cat*+*ribose* substrate and 200 hours for the *ribose*-**lib62** substrate.

As seen under single-turnover conditions, the *cat*+*ribose* substrate is cleaved faster under multiple-turnover conditions than the full-length *ribose*-**lib62** substrate (at equivalent concentrations). This may be attributable to a better ability of the *cat*+*ribose* substrate to bind with or dissociate from the enzyme, or a lower proclivity for forming nonproductive interactions.

Catalytic Power in *trans* is Unaffected by Annealing Protocol

The impact of the slow cooling protocol on *trans* cleavage was tested by mixing *deoxyribose*-**614** with substrate (*ribose*-**lib62** or *cat*+*ribose*) either before or immediately

after the slow cooling. This change in the annealing procedure did not alter noticeably the cleavage profile.

At high concentrations (200 nM), the rate of cleavage of *ribose*-**614** was unchanged by omitting the slow cooling. This also shows that intermolecular folding at the start of the reaction is not significantly effected by the annealing protocol, even though denaturation by heating allows molecules in inactive conformations to refold into active conformations. Similar experiments at low [*ribose*-**614**] favoring *cis*-cleavage showed a slight decrease in initial rate with the omission of slow cooling.

The Commitment Step for *deoxyribose*-614 Cleavage

If the rate limiting step is dissociation of the product-enzyme complex, a burst is expected in the initial phase of multiple turnover kinetics. To seek a burst, the concentration of *deoxyribose*-**614** as a catalyst was held constant at 20 nM while the concentration of *cat+ribose* as a substrate was varied from 100 to 2000 nM. Turnover of substrate to product remained linear well beyond the initial turnover for all substrate concentrations, revealing no burst phase and suggesting that the overall rate is not limited by DNAzyme•Product dissociation.

To estimate the relative magnitudes of $k_{cat(bi)}$ and $k_{-1(bi)}$, a chase was performed with unlabeled *cat+deoxyribose* added after four hours. The *cat+deoxyribose* chase is a substrate analog that cannot be cleaved, and therefore should compete with the labeled substrate. If all of the substrate-catalyst complex proceeds to product (which is the case if $k_{cat(bi)} >> k_{-1(bi)}$), then addition of the unlabeled chase will not influence the subsequent rate of appearance of labeled products. If, however, the rate of substrate dissociation from catalyst is faster than cleavage ($k_{cat(bi)} << k_{-1(bi)}$), then unlabeled chase molecules

lacking a cleavable site should consume the newly disassociated catalyst, and the production of labeled products should largely cease.

In the chase, *deoxyribose*-**614** enzyme (2 μM, saturating) acted in *trans* on small amounts (4 nM) of *ribose*-**614** substrate. When the *cat*+*deoxyribose* chase was present at 5-fold excess over enzyme (10 μM added at t = 4 h), a decrease in the rate of cleavage of *ribose*-**614** after addition of the chase was observed (Figure 3.14a). The cleavage levels of *ribose*-**614** following the addition of chase is reduced below that seen for a predominately *cis*-cleavage reaction of 1.1 nM *ribose*-**614**. This suggests that $k_{cat(bi)} << k_{-1(bi)}$.

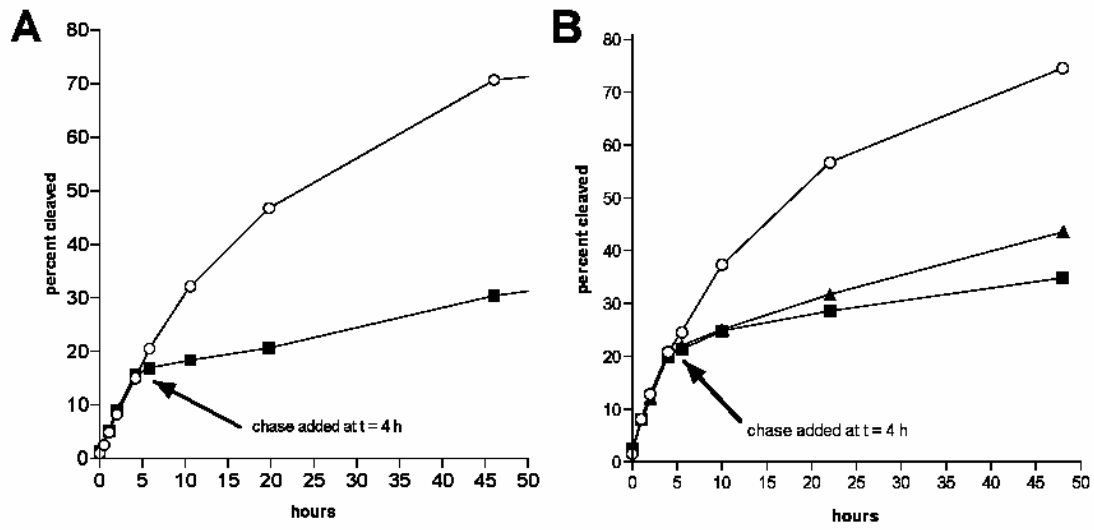


Figure 3.14. Cleavage of various substrates by **614** is reduced by the addition of unlabeled chase.This indicates that the rate of E•S dissociation is faster than the chemical step of cleavage. (A) Saturating amounts of *deoxyribose*-**614** (2 μM) cleaving *ribose*-**614** (4 nM), with (closed squares) and without (open circles) chase (10 μM) added at 4 h. (B) Saturating amounts of *deoxyribose*-**614** (2 μM) cleaving *ribose*-**lib62** (4 nM), without chase (open circles) and with chase (5 μM, closed triangles or 10 μM, closed squares) added at 4 h.

Similar results were seen following the addition of chase to *ribose*-**614** under *trans*-cleavage conditions. Cleavage of *ribose*-**614**, however, was not completely eliminated by the chase. This suggested several explanations, including the persistence of *cis* cleavage

in the presence of chase, or insufficient chase. Chase experiments conducted with *ribose*-**614** cleaving in *cis* (2 nM) showed no significant change in cleavage following the addition of 30 nM chase (data not shown).

If $k_{cat(bi)} << k_{-1(bi)}$, the enzyme-substrate complex will dissociate to give reactants more rapidly than it will form products. If association is also relatively rapid, however, with insufficient chase, the enzyme may bind and release a number of chase molecules until it finds a labeled substrate. The reduction in cleavage should therefore depend on the relative amounts of chase, enzyme, and substrate molecules.

To test this directly and without self-cleavage, chase experiments were performed with saturating levels of *deoxyribose*-**614** enzyme (2 μM**)** cleaving *cat+ribose* substrate (4 nM) and chase in either 5 or 10 μM. Cleavage is greatly reduced, but not eliminated, following the addition of the chase (Figure 3.14b). The reduction of cleavage was slightly greater with chase in 5-fold excess over enzyme as compared to only a 2.5-fold excess. Together, these experiments indicate that the rate of E•S dissociation is greater than the rate of chemical step ($k_{cat(bi)} << k_{-1(bi)}$).

Dependence on Temperature of *deoxyribose*-614 Cleavage

The temperature dependence of initial rates for ribose-**614** cleavage under *trans* cleavage conditions is shown in Fig. 3.15. The rate is lower by a factor of 2.5 at 25°C than at 15°C, and falls dramatically at higher temperatures. Thus, as true for many biological catalysts, it appears as if the catalyst-substrate complex required for *trans* cleavage unfolds (or dissociates) at higher temperatures.

The rate of cleavage of *ribose*-**614** under *trans* cleavage conditions is essentially the same between 4°C and 15°C, and 2.5 fold greater than the rate at the selection

temperature (25°C). Similarly, the rate for *cis* cleavage of *ribose*-**614** is essentially the same at 4°C and 25°C (data not shown). The rate of the chemical step is expected (as an approximation) to increase two-fold for every 10°C increase in temperature (Laidler & Peterman, 1979). Thus, it appears as though the overall rate of catalysis is largely affected by the folding/association step and not the chemical step.

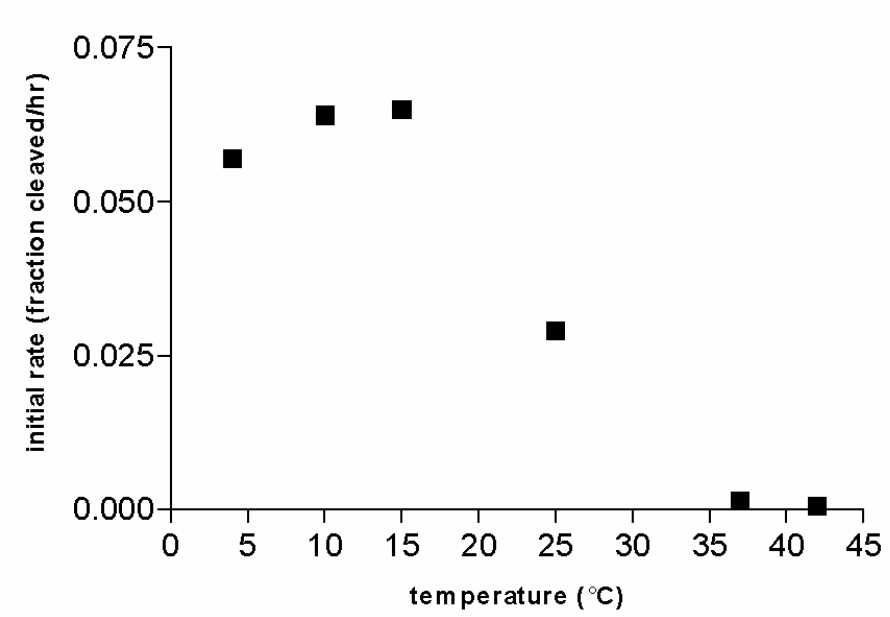


Figure 3.15. *Ribose*-**614** rate of self-cleavage in *trans* is increased at lower temperatures. *Ribose*-**614** (200 nM) was incubated at various temperatures from 4°C to 42°C. Initial rates were calculated and plotted vs. temperature.

The rate of cleavage of *ribose*-lib62 in *trans* by *deoxyribose*-614 in single-turnover experiments is comparable at 15 and 25°C, however. Similar experiments performed under multiple-turnover conditions reveal a rate enhancement at lower temperatures during cleavage of the first substrate, followed by cleavage of additional substrates at a slower rate (Fig. 3.16). This "burst phase" seen under multiple turnover conditions at 15ºC is not seen at 25ºC. This indicates that lower temperatures increase the initial rate by

stabilizing intermolecular association between enzyme and substrate or, perhaps, through greater stability of the folded form of the enzyme. This stabilization presumably also slows the dissociation of product from enzyme to within the range of $k_{-1(bi)}$, and therefore reduces the rate for multiple-turnover subsequent to the first turnover.

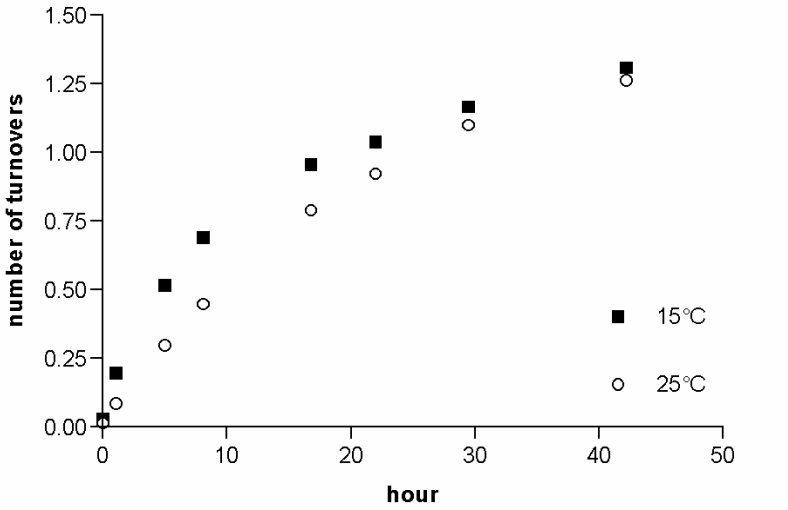


Figure 3.16. Burst kinetics (a higher cleavage rate during the first turnover compared to subsequent turnovers) is enhanced at lower temperatures. *Ribose*-**lib62** substrate (400 nM) was incubated under multiple turnover conditions with *deoxyribose*-**614** (100 nM) at 15°C (closed squares) and 25°C (open circles). The number of substrate turnovers was calculated by multiplying the percent of substrate cleaved by initial substrate concentration and dividing by [*deoxyribose*-**614**].

Predictions of the Energetically Favored Structure are not Supported by Experimental Data

The ability of **614** to cleave in *cis* is not particularly surprising, because this is the function for which the DNAzyme was selected. It is therefore interesting that **614** is able to cleave various substrates in *trans,* and that the apparent first order rate constant at saturation for *trans* cleavage is 4-fold higher than the first order rate constant for *cis*-cleavage. The ability of **614** to cleave the *cat+ribose* primer, as well as a library of

molecules containing the *cat+ribose* primer, suggests that **614** has a binding motif that pairs with a part of the *cat+ribose* primer common to all substrates, thus positioning a separate catalytic motif near the adenine riboside cleavage site.

Chemical modifications specific for single-stranded DNA were used to probe secondary structure *Deoxyribose*-**614** (300 nM) was folded overnight with and without excess unlabeled *cat+deoxyribose* substrate (10 μM). The mixtures were then treated with either potassium permanganate (which modifies thymine) or dimethyl sulfate (which, at high salt concentrations, modifies guanine) for 2 and 5 min. Recovered samples were resolved by PAGE-urea next to a 10-bp ladder.

All thymine and guanine nucleotides were somewhat sensitive to their respective reagents, indicating that either some **614** is not folded at all, or adopts multiple conformations. Nonetheless, T44, T46, T74, G33, G41, G65, G71, and G75 all demonstrated some degree of differential protection (numbers refer to the sequence of **614** shown in Figure. 3.7). The pattern of protected and sensitive positions observed under chemical modifications is not simultaneously compatible with any single structure predicted for **614** folded in *cis* or *trans*. This suggested that either a mixture of conformers exists, or none of the predicted structures accurately reflect the single true structure of **614**.

Likewise, each of the eight lowest free energy structure predictions made by *Mfold* was examined by generating mutations that would disrupt critical helixes. In no case could a loss of function arising from a nucleotide replacement be rescued by a compensatory mutation predicted by *Mfold*. This suggested that none of the predicted conformations dominated (data not shown).

It is interesting to note that many of the structures reported for *in vitro* selected DNAzymes are based on unverified predictions using programs such as *Mfold*. Although such programs have been useful for predicting structures for many RNA and DNA molecules, this does not appear to be the case with **614**.

Discussion

In vitro selection experiments offer the possibility of learning how molecular behavior is distributed within a sequence space defined by the building blocks of a biopolymer, in this case DNA. Structures within DNA sequence space are countable. For standard DNA, 4^n sequences exist, where n is the length of the biopolymer. This distribution is relevant to issues as diverse as the origin of life and biomedicine. For the first, we wish to know how large a hypothetical prebiotic pool of random nucleic acid sequences must have been to contain the physical and catalytic requirements for life. For the second, we may want therapeutic or diagnostic DNAzymes, and need to know the likelihood of obtaining these from selection experiments. In both, we may ask whether added compounds (including divalent metal ions or other cofactors) increase the likelihood of obtaining catalysts.

The answers to such questions might exploit the language of univariate statistical analysis (Johnson *et al*, 1994; Nelson, 1982). A distribution function $P(k_{obs})$ relates the probability of finding a molecule with a particular catalytic power, k_{obs}, to k_{obs} itself. Because good catalysts are presumed to be scarce in the initial library relative to poor catalysts, and excellent catalysts are presumed to be scarce relative to good ones, this distribution is expected to be a decreasing function of k_{obs}. Using such analyses, a better

pool is defined as one having a distribution shifted to the right (Fig. 3.17). Adding a useful cofactor to the pool should shift the distribution to the right as well.

To address such "big" questions, catalysts being examined must be sufficiently well behaved that their behavior can be quantitatively analyzed. Many DNAzymes are not, including (upon first inspection) **614**. In particular, the reaction of **614** did not go to completion. Ca. 35% of the starting material is unreacted even after prolonged incubation. This makes quantitative analysis difficult.

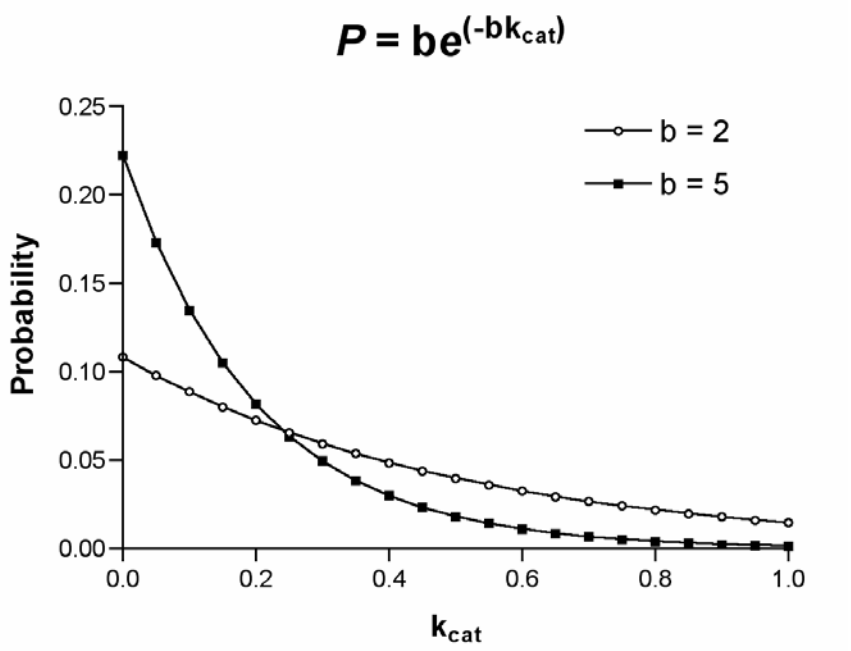


Figure 3.17. A univariate statistical distribution plotting the probability of encountering a catalyst with a particular kobs as a function of kobs, for two catalysts, one with a better pool (open circles, exponential function with b = 2), the other with a poorer (closed squares, exponential function with b = 5). In each case, the area under the line sums to unity. Because good catalysts are scarce in the initial library relative to poor catalysts, and excellent catalysts are scarce relative to good catalysts, this distribution is a decreasing function of kobs. The distribution shown is exponential; the true nature of the distribution in any particular space is, of course, unknown. Key questions in *in vitro* selection ask whether this curve is exponential, zipfian, or is best captured by mathematical approximations having multiple parameters.

Similar behaviors are well known in the literature of nucleic acid catalysts. Hammerhead ribozymes, for example, are frequently reported to cleave to only 40 - 60% at the plateau. DNAzymes with ribonuclease activity and deoxyribonuclease activity are also frequently reported with self-cleavage plateaus of 25 - 65%. This behavior must be understood before any quantitative analysis can begin.

This work accounts for the incomplete cleavage (the plateau) displayed by DNAzyme **614**. Approximately 6-7% of the total *ribose*-**614** sample does not cleave because it is missing the adenine ribonucleoside linkage, which is the site of cleavage. Approximately 9-13% of the total sample appears to fold into an unreactive conformation, which repopulates the active conformation via a cycle of denaturation and renaturation. As much as 15% of the unreacted material at the plateau appears to be sequences containing mutations introduced during amplification or the synthesis of template. These numbers sum to 30-35%, approximately the amount of DNAzyme left uncleaved at the plateau.

Less than 5% of the plateau in our experiments can be attributed to the presence of DNA complementary to the catalyst. Nevertheless, these results show the importance of efficient removal of the complementary strand. Contamination by 10% of the complementary strand has little effect on initial cleavage rate or cleavage plateau, while greater contamination can greatly reduce the cleavage plateau and therefore estimates for cleavage rates. For this reason, we believe both asymmetric PCR and exonuclease degradation are superior to column purification (under strong base) as a way of eliminating the complementary strand.

Given this understanding of the progress curve of the reaction undergone by **614**, we asked whether the features of this DNAzyme reflect the environment under which the DNAzyme was selected. We expected that the presence of Mg^{++} in the selection would lead to catalytic molecules that used Mg^{++} as a cofactor. In part, this expectation arose because Breaker and Joyce themselves reported a Mg^{++}-dependent DNAzyme emerging from their selection. Indeed, the Breaker-Joyce IVS protocol, by using Mg^{++} as the trigger to begin the selection process, seems to require that the DNAzyme be active *only* with Mg^{++}.

In addition, this expectation is based on a general view that the catalytic potential of DNA is poor. DNA lacks a range of functional groups, in particular, many of those present in proteins (Benner *et al*, 1987). In this view, the probability of finding a good catalyst that exploits a useful cofactor (such as Mg^{++}) is greater than the probability of finding a good catalyst that does not. In terms of the distributions shown in Figure 3.17, adding Mg^{++} should shift the distribution to the right.

For both reasons, we were surprised to discover that *ribose*-**614** is as active in the absence of Mg^{++} as it is in its presence. This behavior was not unique to *ribose*-**614** in this selection. It was also observed for two other deoxyribozymes isolated from the same selection, termed *ribose*-**62** and *ribose*-**616**. A fourth DNAzyme generated in this work, *ribose*-**625**, was found to be Mg^{++}-dependent at the temperature used for the selection, but evidently not at 4 °C. While this sample is far from statistical, it suggests that the distribution of catalysts in the pool examined does not greatly favor Mg^{++}-dependent catalysts over Mg^{++}-independent catalysts. This, in turn, implies that adding Mg^{++} to the library does not shift the distribution in Figure 3.17 greatly to the right.

The emergence from a selection of catalysts that fail to take advantage of available cofactors is not unprecedented. For example, Faulhammer and Famulok report a selection performed in the presence of histidine (20 mM) at a low concentration of Mg^{++} (0.5 mM). These authors hoped to select a DNAzyme that exploited histidine as a cofactor. They isolated, however, a DNAzyme that was highly active with Ca^{++} in the absence of histidine. This behavior was selected despite the fact that Ca^{++} was excluded from the selection. Mg^{++}, present in the selection, was a poor substitute for Ca^{++}. At the same time, Roth and Breaker, also seeking a histidine-dependent nucleic acid enzyme, found one (Roth & Breaker, 1998).

Our results differ from those reported by Breaker and Joyce, who analyzed a single, Mg^{++}-dependent catalyst that emerged from an analogous selection. Breaker and Joyce did not survey their catalysts to determine the ratio of Mg^{++}-independent and Mg^{++}-dependent catalysts. Thus, it is conceivable that their study of a Mg^{++}-dependent catalyst differs from our study of a Mg^{++}-independent catalyst because of the stochastic nature of the selection experiment.

Other explanations are conceivable, however. Although the concentrations of Mg^{++} were the same in our selection and the Breaker-Joyce selection, the sequences used by Breaker and Joyce possessed two pairs of self-complementary sequences (each referred to as a "clamp"). These clamps were introduced by design into the library so as to anchor the substrate and enzyme regions of the molecule together with a four base pair clamp on one side of the cleavage site, and a six base pair clamp on the other. The combined length of these clamps increased to 15 base pairs following selection.

In the selection experiments reported here, a binding clamp of this length was not introduced by design. This difference, which was not considered to be significant at the outset, may have had consequences. The absence of extensive clamps may transfer selective pressure onto the folding/association step, and away from the catalytic step. Thus, selection within a library lacking clamps may not generate as strongly catalysts that exploit cofactors as selection within a library having clamps. This may explain why the catalysts that we examined are Mg^{++}-independent, while the catalyst examined by Breaker and Joyce is Mg^{++}-dependent.

A curious parallel is found when comparing the Faulhammer/Famulok selection experiment cited above, and the analogous experiment by Roth and Breaker. The former failed to isolate histidine-dependent DNAzymes; the latter generated several histidine-dependent DNAzymes. Two differences in their selection protocols appear relevant. The Faulhammer/Famulok protocol included low amounts of $MgCl_2$ (0.5 mM) in addition to the histidine cofactor during selection; the Roth/Breaker selection did not. Further, the Roth/Breaker libraries included engineered binding clamps, analogous to the clamps in the Breaker-Joyce selection for ribonucleases. The Faulhammer/Famulok libraries did not.

It may be coincidental that selections with clamped libraries in two cases generated catalysts that exploited an external cofactor, while analogous selections with libraries lacking the clamp generated catalysts that did not. These results suggest, however, that the presence of base pairing between the substrate and enzyme portions of the nucleic acid enzyme may be important to the experimental outcome.

Including clamps between the substrate and enzyme motifs in the DNAzyme appears also to give faster catalysts. The catalyst that Breaker and Joyce examined in detail has a rate constant of 0.12 hr^{-1} before reselection; reselection improved this to 1.2 – 4.6 hr^{-1}. Likewise, the catalyst reported by Santoro and Joyce had rate constants of 6 hr^{-1} and 600 hr^{-1} before and after reselection. The metal-cofactor independent catalyst reported by Sen and Geyer had rate constants of 0.17 hr^{-1} and 0.4 hr^{-1} before and after reselection. In contrast, the rate constants for the metal-cofactor independent catalysts selected here are considerably smaller (0.015 to 0.05 hr^{-1}, both before reselection).

This work suggests that if one wishes to obtain a fast catalyst, one should engineer clamps into the sequence. Indeed, Santoro and Joyce found that reducing the length of the clamps (for a DNAzyme selected with clamps) to six base pairs or fewer (on both sides of the clamp) reduced k_{cat} 10 fold for an intermolecular reaction and increases K_M 100 fold. On the other hand, to understand catalytic potential in a truly random pool, one is advised not to use engineered clamps. The information in an engineered clamp containing a 14 base pair duplex is substantial, even considering that the exact bases paired in not important. The specification of 14 nucleotides in a 4 letter alphabet is $4^{14} \approx 10^8$. This is significant relative to the size of a typical library (10^{13} molecules).

What do these experiments tell us about sequence landscape and evolvability within DNA libraries, where the reaction sought is the cleavage of a ribonucleotide linkage? Many sequence variants of **614** were isolated that differed at only one or two positions. Most of these exhibited cleavage rates that were more than an order of magnitude slower than that for **614** itself. In contrast, variants of **614** changed at one or two sites were isolated following additional rounds of selection. These generally had

rates that were an order of magnitude faster than those for the original **614**. The fact that a 100-fold variation in rates arising from only minor changes in sequence suggests a degree of ruggedness to the landscape relating sequence space to catalysis in this system (Yomo, 2003).

What do these experiments tell us about the origin of life? The hypothesis that life emerged as a nucleic acid capable of self-replication (Rich, 1962) remains disputed (Shapiro, 1988). Our recent discovery of a plausibly prebiotic route to ribose is encouraging the re-examination of this hypothesis (Ricardo *et al*, 2004). In no case, however, does DNA appear to be a candidate for the first living biopolymer.

To the extent that conformation and catalysis in RNA and DNA is analogous, however, we may offer a few observations. First, binding clamps of sufficient length are unlikely to occur frequently in a prebiotic soup. If binding clamps of 14 base pairs are required (the optimal length estimated by Santoro and Joyce), it would increase the required size of a prebiotic library by a factor of 10^8. Without extensive clamps, it appears (if this work can be taken as a model) that catalysis is limited by folding. Folding, in the absence of extensive clamps, occurs more favorably at low temperatures. To the extent that the behavior of **614** is representative, cold temperatures would appear to be more favorable for a nucleic acid origins of life than high temperatures.

Last, *trans* cleavage in this system is incidental to selection. It is, however, a likely result of the promiscuous nature of nucleic acid binding. Under the conditions in which a nucleic acid catalyst was evolving, selection likely acted upon a small set of nucleic acid catalysts isolated within a compartment. Significant *trans* interactions would be possible

within the compartment. If *trans* reaction is generally associated with a nucleic acid catalyst that is capable of a *cis* reaction, this may be relevant to the origin of life.

CHAPTER 4
DETECTING ORGANIC MOLECULES ON MARS

Introduction

Mars has become a prime target in human space exploration during the last decade. This renewed interest is in part due to the nascent field of astrobiology and the suggestion that a vestige of ancient Martian life was found in the interior of the Allan Hills meteorite (Mc Kay, *et al*. 1996). Several landing missions have since been directed to the red planet but about half of these were unable to descend safely to the Martian surface. Unfortunately, many of the instruments directly relevant to the fields of chemistry and astrobiology were lost in the unsuccessful landings. Therefore, the knowledge of the chemistry of the Martian surface remains limited to the experimental results obtained by the Viking missions nearly 30 years ago and the Pathfinder rover in 1999. As this dissertation is being completed, new data is also emerging from the opportunity rover.

The Viking 1976 missions to Mars performed several experiments designed to assess the potential for life on the planet. The results were puzzling. Samples of Martian soil taken from the top 10 cm of the Martian surface released dioxygen when exposed to humidity (Oyama & Berdahl, 1977). At least one compound in a set of radiolabeled organic compounds (formate, D,L-lactate, glycollate, glycine and D,L-alanine) released radiolabeled carbon dioxide when placed in aqueous solution on the Martian surface, evidently via oxidative processes (Levin & Straat, 1979). Last, a gas chromatography-mass spectrometry (GC-MS) experiment looking for volatile products from a sample of

soil heated for 30 seconds at 500 °C did not detect any organic molecules (Biemann et al, 1977).

The failure to detect organic molecules by GC-MS was especially surprising since some 2.4 x 10^8 g of unaltered carbon comes to Mars each year via meteor (Table 4.1)(Flynn, 1996; Hayatsu & Anders, 1981; Mullie & Reisse, 1987). Many meteoritic organic compounds are volatile and should have been detected by GC-MS (Sephton, Pillinger, & Gilmour, 1998). Pyrolysis should have generated volatile products from many of the non-volatile compounds, including the polymeric organic substance known as "kerogen", which accounts for the majority of organic material coming to Mars via meteorite, and as much as 1-3% of the weight of some meteorites (Hayes & Biemann, 1968). These too should have been detected by Viking, but were not.

These results have been interpreted as evidence that the Martian surface contains no organic molecules of any kind, presumably because the Martian regolith carries an oxidant powerful enough to convert all organics to carbon dioxide. Coupled with the absence of liquid water on the surface of Mars, and the irradiation of the surface by ultraviolet light, the failure to detect organic substances led many to conclude that one must dig deeply (and perhaps very deeply) below the Martian surface to have a chance of encountering any organic molecules that may have arisen from life on Mars (perhaps present several billion years ago, when the surface of Mars was more like the surface of Earth at that time), or organic molecules that may have been delivered to Mars via meteorite (McKay et al, 1998; Kieffer et al, 1992).

As this conclusion was influencing the design of missions to Mars, Benner *et al*, 1999 re-examined it in light of what is known about the oxidation of organic compounds,

the nature of organic substances likely to come to Mars, and the features of the Viking 1976 analysis that limited the kinds of organic molecules that it could have detected. The examination suggested that organic compounds that arrive to Mars via meteorite are most likely to be converted to carboxylic acid derivatives that would not be easily detected by GC-MS. Organic molecules generated on Mars itself by non-biological (Hubbard et al, 1973; Chyba & Sagan, 1992; Hubbard, Hardy, & Horowitz, 1971; Horowitz & Hobby, 1977) or (entirely hypothetical) biological synthesis (Levin, 1977) should suffer similar fates at or near the surface.

The hypothesis of Benner *et al*, started with the fact that Mars is exposed to ultraviolet radiation with sufficient energy to cleave water to give H• and HO• radicals. Some of the H• radicals must recombine to give dihydrogen (H_2), which escapes into space (Hunten, 1974; ibid, 1979), leaving behind the HO• radical, which could react directly with organic substances, dimerize to give H_2O_2,(McDonald, de Vanssay, & Buckley, 1998) or generate peroxides or other species through combination with minerals in the Martian soil.

The HO• radical reacts directly with most organic molecules. In aqueous solution under terrestrial conditions, the second order rate constants range from 10^7 to 10^{10} $Lmol^{-1}sec^{-1}$ for reactions that include hydrogen atom abstractions and additions to double bonds. The concentration of HO• on Mars is 1-2 x 10^5 cm^{-3}, a number similar to the concentration of HO• radical in the atmosphere at the surface of Earth (Hunten, 1979).

Benner *et al*. considered five types of organic compounds (Table 4.1) known to come to Mars via meteorites: alkanes, alkylbenzenes, naphthalene and higher polycyclic

aromatic hydrocarbons, kerogen, and amino and hydroxy acids. They then formulated a hypothetical mechanism of how HO• and H_2O_2 might transform these compounds in generic oxidation pathways, and whether metastable intermediates in these pathways might accumulate.

Table 4.1. Expected metastable products from organic substances in the Murchison meteorite (Hayatsu & Anders, 1981; Mullie & Reisse, 1987)

Substance	Concentration (parts per million)	Metastable Products
Acid insoluble kerogen	14500	benzenecarboxylic acids
Aliphatic hydrocarbons	12-35	acetate
Aromatic hydrocarbons	15-28	benzenecarboxylic acids
Monocarboxylic acids	~330	acetate/oxalate
2-Hydroxycarboxylic acids	14.6	acetate/carbonate
Alcohols (primary)	11	acetate
Aldehydes	11	acetate
Ketones	16	acetate, benzenecarboxylic acids
Amines	10.7	acetate
Urea	25	carbonate
Heterocycles	12	carbonate, other products

Oxidation of Alkanes under Martian conditions

Alkanes react generically with the HO• radical via abstraction of a hydrogen radical (H•) at a tertiary center preferentially, then a secondary center, and last a primary center. This relative reactivity reflects the relative stability of the radical products, and is displayed under a wide range of conditions. Thus, a straight chain alkane would lose an internal hydrogen in the generic mechanism (Figure 4.1a).

The resulting secondary radical is extremely reactive, and will be trapped by almost anything available. It will react with another HO• radical to yield a secondary alcohol. It can transfer an electron to (for example) Fe^{+3}, generating a carbocation that can be

trapped by water (for example, from a hydrated mineral), also generating the secondary alcohol. Other products are possible, but a secondary alcohol is the generic intermediate in the oxidative degradation of *n*-alkanes (Figure 4.1a).

HO• abstracts an H• from the carbon attached to the alcohol oxygen more readily than it abstracts H• from the parent alkene. Thus, the secondary alcohol (under generic conditions) is expected to react faster than the parent. It will therefore not accumulate, but yield a ketone. The ketone, in turn, should undergo further oxidation to generate an ester, which will be cleaved to give a carboxylic acid and a primary alcohol, which will be oxidized directly to another carboxylic acid. Alternatively, the ketone might enolize, suffer oxidation, and then lead to a fragmentation to generate two carboxylic acids (Figure 4.1a).

By these steps, the generic oxidation pathway for alkanes leads to carboxylic acids. These are, of course, subject to further oxidation. The abstraction of an H• from the carbon attached to the COOH group is expected to be an important mode of oxidation involving HO•. This will ultimately generate the next shorter carboxylic acid. An alternative mode involves the one electron oxidation of the carboxylate anion to give the carboxylate radical, which would lose carbon dioxide to give the radical of the shorter-by-one-carbon alkane. Depending on the trap, the product would be an alkane (and the process would resume), or another more easily oxidized derivative.

In this cascade of intermediates, the carboxylic acid is the first that is slower to degrade than to be formed. Under typical Fenton conditions, for example, acetic acid reacts with the HO• radical 100 times more slowly than does ethanol (Walling, 1999). Carboxylic acids are therefore likely to accumulate. Further, acetate is more stable to

further reaction under generic conditions than propanoic acid and longer alkylcarboxylic acids. Thus, acetic acid accumulates especially effectively.

Exemplifying the "generic oxidation" pathway are some "brand name" oxidations. In a Kuhn-Roth oxidation, for example, an alkane is refluxed in a solution of concentrated chromic acid (Kirsten & Stenhagen, 1952). Insignificant amounts of ketone or alcohol products can be isolated as intermediates in the oxidation cascade that follows; these are too unstable with respect to further oxidation. Organic alkanecarboxylic acids (butanoic acid and propanoic acid, for example) can be isolated as metastable intermediates, however. Upon incubation for longer times, these are further degraded to acetic acid. Acetic acid too can be oxidized, to give carbon dioxide. Nevertheless, acetate is more stable than longer chain alkanecarboxylic acids, and accumulates. This makes the Kuhn-Roth oxidation useful for elucidating the structure of natural products. The amount of acetate produced from a known amount of alkane corresponds to the number of methyl groups in the alkane.

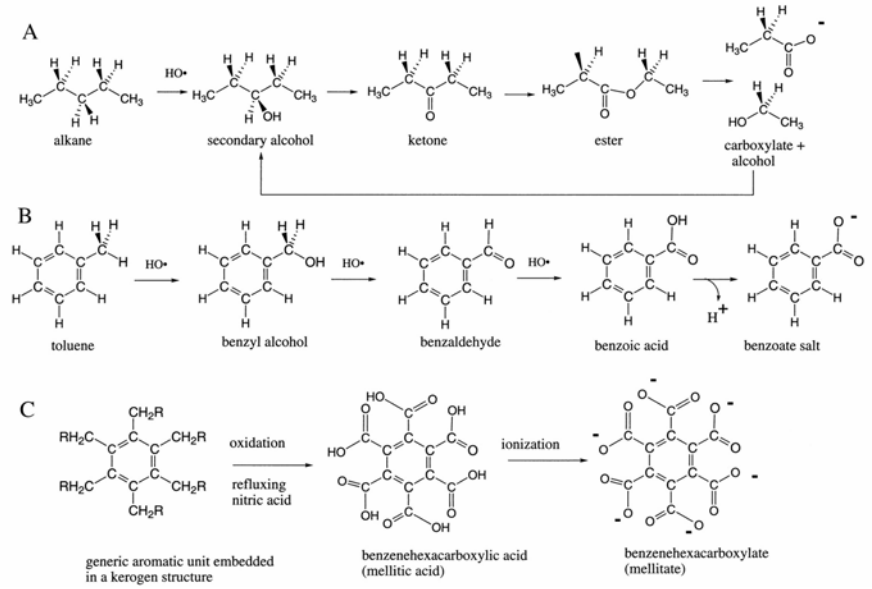


Figure 4.1 Oxidative degradation of the generic alkane (represented here by pentane) to acetic acid (a), toluene to benzoic acid (b), and kerogen to (mellitic acid) (c).

Oxidation of Alkylbenzenes Under Martian Conditions

The HO• radical abstracts a benzylic H• from the alkyl group of alkylbenzenes (such as toluene) to give a rather stable benzyl radical (Figure 4.1b). This may trap HO•, or lose an electron to Fe^{3+} and then trap water, in each case forming benzyl alcohol. Benzyl alcohol is more reactive than toluene under generic conditions. It is not expected to accumulate, but rather be converted to benzaldehyde. Benzaldehyde is also unstable with respect to further oxidation, and also should not accumulate in the generic process. Rather, it should be converted to benzoic acid.

Benzoic acid no longer has a benzylic hydrogen to lose to a radical oxidant. It is still thermodynamically unstable in the presence of oxidants to conversion to carbon dioxide. But it is metastable, resistant to further oxidation, and accumulates. Because benzoic acid has no hydrogen on the carbon adjacent to the COOH group, it also lacks a path available to alkanecarboxylic acids for further oxidative degradation. Further generic oxidative degradation involves a one electron oxidation of the benzoate anion, which decarboxylates to yield the phenyl radical, which then can be converted to benzene or phenol.

This generic pathway can be illustrated by a specific oxidative process with commercial importance. Benzoic acid is synthesized on ton scales via the oxidation of toluene. The stability of benzoic acid under these oxidizing conditions is sufficient to allow benzoic acid to accumulate in the industrial process (Heberger et al, 1994).

Oxidation of Polyaromatic Hydrocarbons (PAHs) Under Martian Conditions

The generic oxidation of polycyclic aromatic hydrocarbons involves the addition of the HO• radical to give a hydroxycyclohexadienyl radical. This suffers further oxidation to give eventually a single core aromatic rings to which carboxylic acids are attached

wherever a second ring was fused. Thus, naphthalene, phenanthrene and anthracene all give phthalic acid in the generic oxidation process (Barbas, Sigman & Dabestani, 1996; Theurich et al, 1997). Higher PAHs give benzenetricarboxylic, tetracarboxylic, pentacarboxylic, and hexacarboxylic acids (Figure 4.1c)(Juettner, 1937).

The generic pathway can be exemplified with laboratory reactions of naphthalene, which is 1-6 ppm in some carbonaceous chondrites (Pering & Ponnamperuma, 1971). The pseudo first-order rate constant for the first step in the reaction between naphthalene and the HO• radical (Figure 4.2) is 0.035 min^{-1} (Bunce et al, 1997). The rate constants for further oxidation 1- and 2-naphthol are higher (0.88 and 0.27 min^{-1}). This implies that neither 1- nor 2-naphthol will accumulate. The metastable end products are phthaldehyde and phthalic acid.

This is true for a variety of other conditions (oxidation catalyzed by TiO_2, by SiO_2, (Barbas et al, 1993) and by Fe_2O_3 (Guillard et al, 1993). This uniformity in outcome argues that the oxidation of naphthalene will generically yield phthalic acid as a metastable intermediate (Lane et al, 1996). The metastability of phthalic acid to further oxidation has commercial significance. An important industrial synthesis of phthalic acid begins with the oxidation of naphthalene (Lowenheim et al, 1975). Phthalic acid is also produced from naphthalene under simulated Martian conditions (Oró & Holzer, 1979).

Oxidation of Kerogen Under Martian Conditions

Polymeric organic material ("kerogen") has no defined structure. On Earth, kerogen (coal, for example) comes via metamorphosis of biological matter. Under generic oxidation

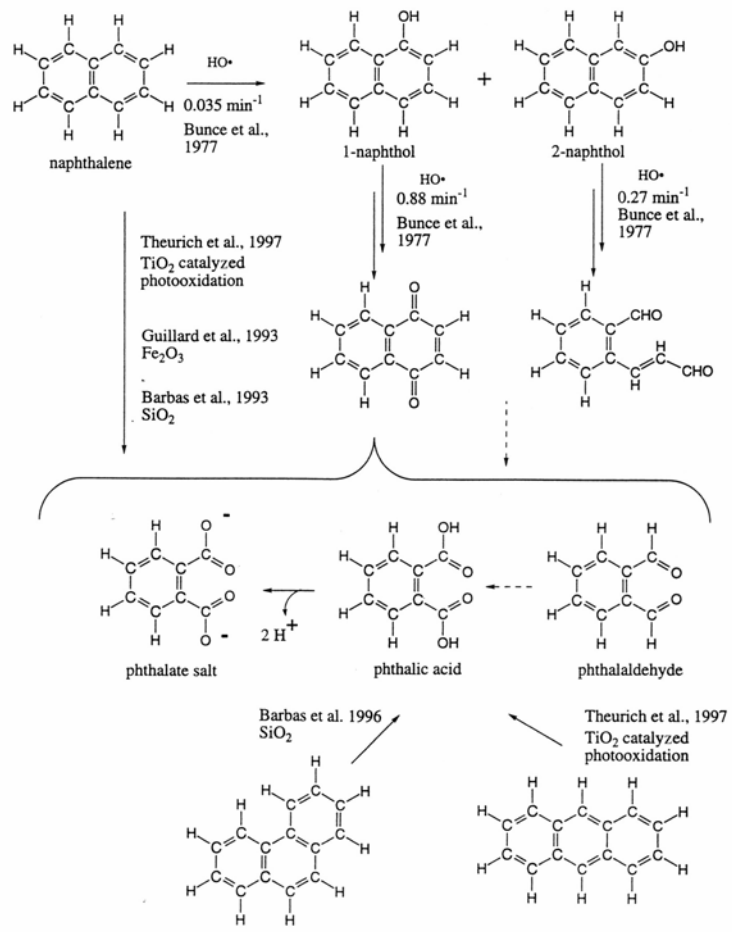


Figure 4.2. Oxidative degradation of naphthalene to phthalic acid. Solid arrows indicate reactions documented in the literature with citations. Dotted arrows indicate transformations presumed to occur but without documentation in the cited literature.

conditions, the aromatic portions of kerogen generate benzenecarboxylic acids, with one carboxylic acid group for every position on the core benzene ring that was attached to a carbon in the parent structure. These are metastable, accumulate, and are isolated and quantitated when defining the structure of kerogens. For example, treating coal with alkaline permanganate oxidized its carbon to carbonic acid (H_2CO_3, 42%), acetic acid (CH_3COOH, 2%), oxalic acid (HOOC-COOH, 7%) and benzenecarboxylic acids (48%), with a trace of succinic acid (HOOC-CH_2-CH_2-COOH) (Bone, Horton & Ward, 1930). Kerogen is the most abundant organic substance in meteorites. As with terrestrial

kerogen, the kerogen from the Murchison meteorite gives benzenecarboxylic acid products when oxidized (Hayatsu et al, 1977; Hayatsu et al, 1980). These are stable in refluxing nitric acid for 27 hours.

Oxidation of Amino and Hydroxyacids Under Martian Conditions

Polyfunctional molecules are easier to oxidize than unfunctionalized carboxylic acids. Thus, hydrogen peroxide (a mild oxidant) will, in the presence of iron salts, catalyze the oxidative decarboxylation of alpha-hydroxyacids to give carbon dioxide and the shorter aldehyde. This reaction, well known in sugar chemistry, has a brand name (the "Ruff degradation")(Wieland & Franke, 1927).

Because of their multiple sites of reactivity, polyfunctionalized compounds are not expected to survive the generic oxidation conditions; they are the most likely to be converted entirely to carbon dioxide on the surface of Mars. One intermediate in the oxidative degradation of polyfunctional organic compounds is oxalic acid, HOOC-COOH. A report from 1928 suggests that the iron salt of oxalic acid is stable to further oxidation (Walton & Graham, 1928). Thus, oxalic acid may be a metastable intermediate in the generic oxidation pathway on Mars, where iron is abundant.

The Amounts and Fates of Organic Carboxylic Acids

This discussion makes the case that aromatic and aliphatic carboxylic acids are the metastable products of generic oxidation of meteoritic organic compounds. The generic oxidation pathway is exemplified by so many specific (admittedly terrestrial) reactions, and is so well supported by Organic Structure Theory, that it seems plausible that these products are found on Mars as well.

Assuming that meteorites bring 2.4×10^8 g/year of organic carbon to Mars (Flynn, 1996), that kerogen in the Murchison meteorite is representative of this organic carbon,

that Mars-bound kerogen generates the same proportion of benzenecarboxylic acids as the Murchison kerogen, and that the mass yield of these is 10%, then ca. 2 x 10^{17} grams of benzenecarboxylic acids should have been generated on Mars since its surface dried three billion years ago. This corresponds to 2 kg of benzenecarboxylic acids per square meter of the Martian surface (the surface area of Mars is ca. 10^{14} m^2).

It is possible that these compounds were diluted by wind and impact into the Martian regolith. If mixed in the regolith to a depth of one meter, 2 kg of benzenecarboxylates would contribute ca. 500 ppm by weight of the first meter of surface of Mars (the density of Mars is ca. 4 gm/cm^3). If gardening mixes the material to a depth of 1 km, benzenecarboxylates will be present at a concentration of 500 ppb. The Viking MS had a sensitive to the ppb level, and should have detected these had it had access to them.

Other processes might have removed organic carboxylic acids from the immediate surface. Carboxylic acids react with metal oxides to form salts (Martell & Smith, 1977; Avdeef, 1993). These often display some solubility in water. If the Martian surface has been exposed episodically to water, organic salts may have been removed from the surface by leaching, and concentrated in sub-surface environments. Examples on Earth include highly soluble salts (e.g., halite) and poorly soluble salts (e.g. gypsum). This process is almost certainly less important than gardening in the recent past, as surface water on Mars has been scarce for billions of years, and iron salts of benzenecarboxylic acids are poorly soluble (Giammar & Dzombak, 1998; Wu et al, 1996).

Most important, of course, are chemical reactions that would degrade the carbon skeleton. The Fenton reaction serves as a model, even recognizing that it is best known as

an aqueous process (Walling, 1997). The Fenton reaction is believed to involve HO• radical generated from H_2O_2 (Chen et al, 1998). Hematite and goethite (both iron oxides believed to exist on the Martian surface) are effective catalysts (Lin & Gurol, 1998; Watts et al, 1997).

The Fenton reaction is known to degrade organic molecules ranging from 2-methylnaphthalene (an aromatic compound) to n-hexadecane (an aliphatic compound) entirely to carbon dioxide, if given sufficient time. Even relatively resistant molecules can be degraded. For example, trinitrotoluene (TNT) is converted by H_2O_2 in water to trinitrobenzoic acid, from there to trinitrobenzene, and from there to oxalic acid as the primary organic end product. The oxalic acid is removed only when the mixture is exposed to light (Li, Comfort & Shea, 1997). Fenton chemistry converts benzoic acid into hydroxybenzoic acid and guanosine into 8-oxoguanosine (Sandstrom et al, 1997). A UV-accelerated Fenton reaction is also known in aqueous solution, and is proposed to generate Fe^{2+} by photoreduction (Sun & Pignatello, 1993). The efficiency of the Fenton reaction depends on the ligands around iron (Pignatello & Baehr, 1994; Dean, & Nicholson, 1994). For example, ferric oxalate initiates the destruction of other molecules (Safarzadeh-Amiri , Bolton & Cater, 1997; *ibid,* 1996). Benzoic acid inhibits the Fenton reaction in certain terrestrial experiments (Zakharov & Kumpan, 1996). Aromatics are protected against degradation by more easily oxidized species (Owen et al, 1996). Thus, it is difficult to predict the consequences of Fenton chemistry on Mars, even if we assume that the process is analogous to the aqueous process known in the laboratory.

An alternative path for the oxidative degradation of carboxylic acids involves the one electron oxidation of their anions to give the corresponding carboxylate radicals

(Lamrini, 1998). These will lose carbon dioxide to generate an organic radical, which will then be trapped as part of the oxidative cascade. Photons can accelerate this process, and are likely to be important on the UV-irradiated surface of Mars (Oró & Holzer, 1979; Bullock et al, 1994). Alkanecarboxylic acids are particularly susceptible to photochemical degradation. Many benzenecarboxylic acids are quite stable to photochemical degradation, however (Jeevarajan, & Fessenden, 1992). Phthalic acid derivatives, for example, yield phthalic anhydride under prolonged irradiation (Balabanovich, & Schnabel, 1998; Balabanovich, Denizligil & Schnabel, 1997), but no further degradation of the phthalic acid core.

Failure of Viking 1976 to Detect Organic Carboxylic Acids

If laboratory reactions are taken as examples of the generic oxidation pathway, the rates for the destruction of benzenecarboxylates are at least 10^3 to 10^6 fold slower than their rates of formation. Depending on the tempo of chemistry overall on Mars (and remembering that the billion years available for the accumulation of meteoritic organics is also available for the destruction of the derived benzenecarboxylates), substantial amounts of the kilogram of benzenecarboxylates expected to be generated per square meter should have survived. Their concentration would fall below the nominal sensitivity of the Viking 1976 mass spectrometer only if more than 99% of these were destroyed *and* if gardening diluted these to an average depth of 1 km or greater.

To gain access to the Viking mass spectrometer, however, the organic molecule first pass through a gas chromatograph. Only volatile molecules can do so. Salts of organic carboxylic acids are not volatile. Thus, the salts of benzenecarboxylic acids, oxalic acid, and acetic acid would not be *directly* detectable by the Viking GC-MS experiments, even if they had been present.

The ability of the Viking experiments to detect organic carboxylates therefore depends critically on the ability of these carboxylates to generate volatile products in the sample preparation (pyrolysis for 30 seconds at 200, 350, and 500°C). The three generic oxidation products, benzenecarboxylic acids, acetic acid, and oxalic acid, could be detected only with difficulty by the Viking GC-MS. Oxalic acid generates carbon dioxide, carbon monoxide and water under pyrolysis. These were in fact detected, but are all also components of the Martian atmosphere.

Higher benzenecarboxylic acids also do not easily yield volatile pyrolysis products. Benzenehexacarboxylate, a non-volatile compound, will eventually release carbon dioxide upon pyrolysis and become benzenepentacarboxylate (Manion, McMillen & Malhotra, 1996) and then benzenetetracarboxylate. The salts of these, however, are also not volatile.

Acetic acid and its salts may be pyrolyzed to give volatile products. At high concentrations, acetone is formed (Davis, & Schultz 1962). However, the iron (II) acetate and iron (II) propionate salts are reported to be "amazingly stable up to 400-500°C" (Granito & Schultz, 1963).

For these reasons, the Viking experiments do not exclude the possibility that the soil being tested contained organic carboxylic acids, especially benzenecarboxylic acids in substantial amounts. To examine experimentally this conclusion, Benner *et al* 2000, synthesized the iron (III) salts of phthalate, mellitate and benzene-1, 2, 4-tricarboxylate. These were then subjected to thermolysis-mass spectrometry following heating from 25° to 400° in 30 seconds. The phthalate, mellitate and oxalate salts were separately heated to 400° in a quartz capillary in a direct insertion probe. Under these conditions, which give

the mass spectrometer better access to the probe than the Viking MS had, no signal was observed from the iron (III) salts of mellitic acid and benzene-1,2,4-tricarboxylic acid. Iron (III) phthalate yielded a peak that might have been detected by the Viking GC-MS, corresponding to phthalic anhydride. Iron (II) acetate generated acetone, acetic acid and acetic anhydride in addition to water and carbon dioxide. Iron (III) oxalate releases carbon monoxide as well as carbon dioxide and water. This suggests that the Viking experiments can rule out substantial amounts of acetate on the surface (the top 10 cm), and modest amounts of phthalate, but not higher benzenecarboxylates, which are the principal products of the generic oxidative degradation of organic materials arriving via meteorite.

The Infrared Spectra of the Martian Surface

The infrared spectra of the Martian surface (Figure 4.3) was recently obtained by using data obtained with the Termical Emission Spectrometer (TES) instrument on board of the Mars Global Surveyor (MGS) (Bandfield, et al, 2003).

The assignment of the signals on the spectra by Banfield *et al*, was done with the assumption that carbonate minerals are present in the surface of Mars (since not organic molecules had been detected on the planet). Several carbonate minerals mixtures were prepared in different proportions until the infrared spectra of the prepared sample closely resemble the most distinct features of the Martian surface spectra (Figure 4.4).

From this comparison, the author concluded that magnesite (MgCO3) was the mineral which adsorption lines fit better the Martian IR spectra (Figure 4.5). The carbonate mineral was found to be distributed uniformly around the planet with not indication of a concentrated source and calculated to have an abundance in the surface of

~2-5% weight on Mars surface. The authors however recognize that many other combinations of minerals will probably give a better match to the spectra.

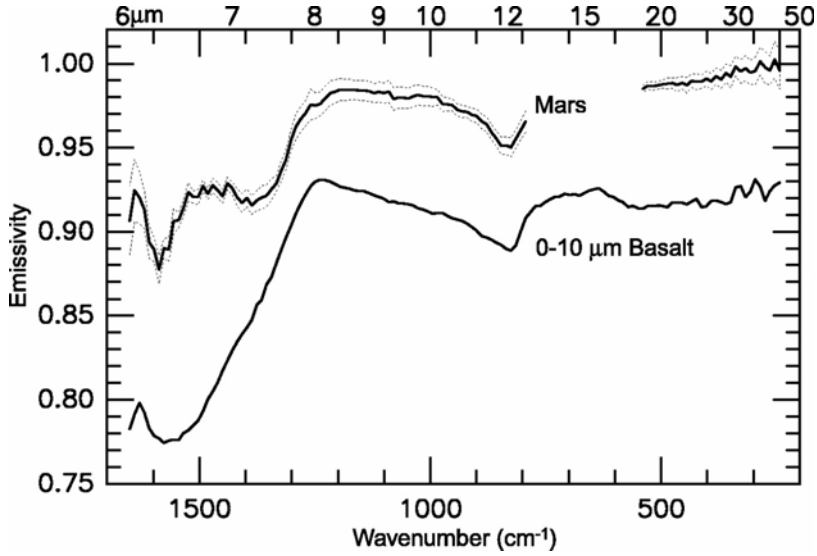


Figure 4.3. Martian surface IR spectrum. The top spectra corresponds to the high- albedo surface dust spectrum, dashed lines represent standard deviation. The spectra at the bottom corresponds to terrestrial particulate basalt, shown to illustrate the resemblance in spectral features between 300-1200 cm-1.

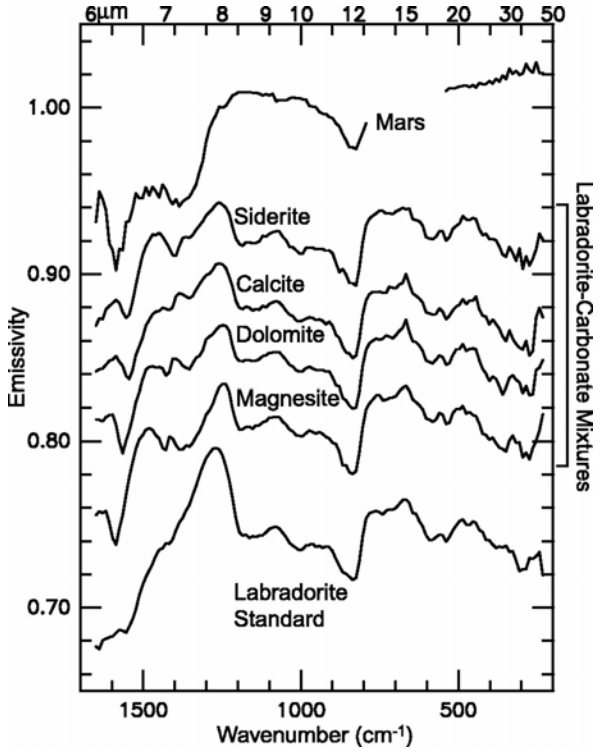


Figure 4.4. Mars dust, labradorite standard and labradorita + minerals mixture spectra. (see absorptions in the regions of 1350 cm-1 to 1580 cm-1 and a carbonate absorption at ~880 cm-1)

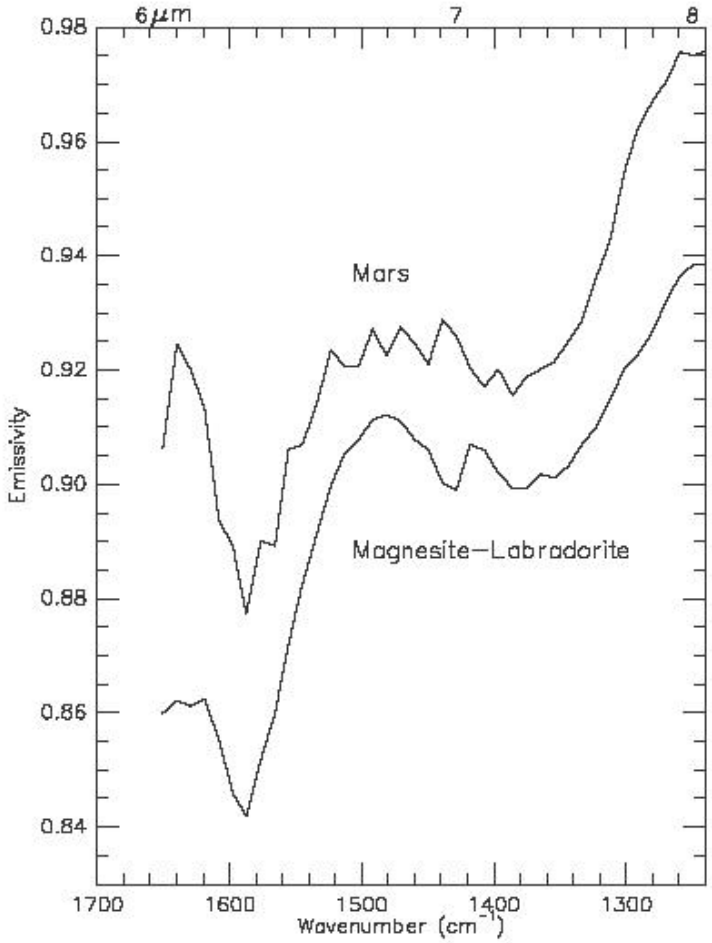


Figure 4.5. Mars dust and magnesita-labradorite mixture spectra.

The predominant abundance of magnesite over siderite ($FeCO_3$) or calcite ($CaCO_3$) was not explained in the paper, also details in the 1380-1520 cm^{-1} region were leave unexplained (the elemental abundance in the martian soil for the metals Mg, Fe and Ca are: magnesium~5%, iron ~13%, calcium ~4% ; Rieder et al, 1997).

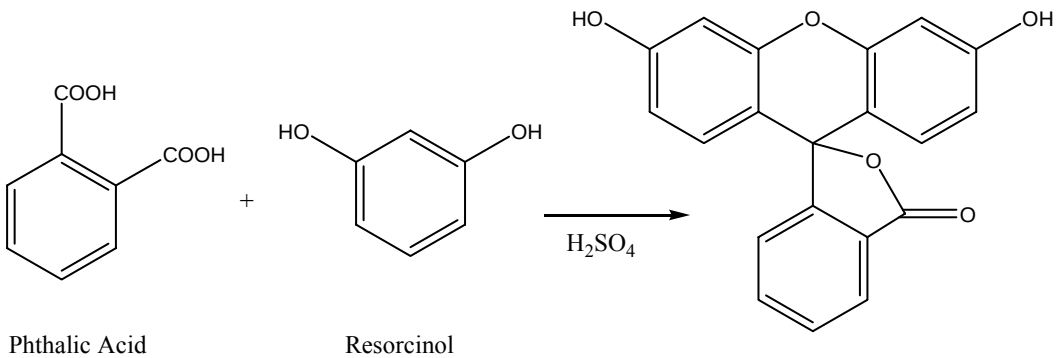
In the present chapter salts of benzenecarboxylates will be synthesized by the general method described by Galwey (1965), and their infrared spectra will be compared to that of the martian soil to evaluate if these salts could account for the unexplained features in the spectra.

Detecting the Missing Organics on Mars

Any attempt to perform qualitative and quantitative chemistry on Mars using a robotical lander, will be conditioned to the non-chemical aspects associated with space exploration (i.e: payload, complexity of the task, etc.). With this in mind, an experimental scheme for the detection of benzenecarboxylates is presented and evaluated on this chapter.

Fluorescence derivatization of the different benzenecarboxylates will be attempted by using a one step reaction known as the "Fiegl's test" (Feigl, 1961) which has been traditionally used in classical organic chemistry for the qualitative identification of 1,2-dicarboxylic acids. Fiegl's test requires the addition of an excess 1,3-benzenediol (resorcinol) and catalytic amounts of acid (sulfuric acid) to the dicarboxylic acid containing molecule under analysis. The reaction occurs in a few minutes if the mixture is heated at 120 °C and the product formed contains the xanthene ring structure (Figure 4.6).

Fluorescein is the reaction product of phthalic acid (or anhydride) under Fiegl's conditions and is of widely use in chemical and biological fluorescence sensing systems due to the high quantum yield and chemical stability that allows its detection at the single molecule level (Figure 4.6). The addition of carboxylic units to the phtalic acid core does not appear to affect the fluorescent ability of the fluorescein derivatives. Carboxy-fluorescein the product of resorcinol condensation with 1,2,4 benzene tricarboxylic acid is also a common fluorophore in molecular biology applications due to its enhanced solubility in water as a result of the additional carboxylic residue.



Phthalic Acid Resorcinol Fluorescein

Figure 4.6. Phthalic acid yield fluorescein when heated with resorcinol in the presence of acid.

The reactivity of the iron salt of benzenecarboxylates hypothesized to be present on Mars surface, will be also tested towards fiegl's reaction.

Materials and Methods

Chemicals

All reagents and materials were purchased from Sigma-Aldrich Co. in their highest quality and used without any further purification.

Analytic Instrumentation

Fluorescence analysis

Flourescence spectra (emission and absorption) were recorded in a SpectroFluorometer Fluorolog 3. Courtesy of Professor Weihong Tan at the University of Florida.

Infrared analysis

Infrared spectra (KBr pellets) were recorded in a Perkin Elmer 1600 series FT-IR. Courtesy of Professor Kirk Schanze at the University of Florida.

High performance Liquid chromatography- mass Spectrometry (HPLC-MS)

HPLC-MS (ESI) analysis were carried out by Dr. Jodie Johnson at the University of Florida.

Synthetic Preparations

Synthesis of Mellitic Acid Salts

Crystalline aluminium mellitate (mellite) was obtained as a gift from Dr. Steven Benner mineral collection.

Synthesis of manganous mellitate (1)

Manganous carbonate ($MnCO_3$; 0.55 g, 4.3 mmol) and mellitic acid (0.5 g, 1.4 mmol) were mixed in water (10 mL) and boiled under stirring to complete evolution of gas, the solution was cooled, filtered and the precipitate was washed thoroughly with Milli-Q water. The resulting solid was dried in air at 80 ºC.

Synthesis of zinc mellitate (2)

The procedure for the preparation of this salt was the same as the one used for the synthesis of compound (1). The amounts used were: zinc carbonate-hydroxide hydrate ($ZnCO_3.2Zn(OH)_2.H_2O$; 0.68 g, 2.1 mmol) and mellitic acid (0.25 g, 0.7 mmol).

Synthesis of cupric mellitate (3)

The procedure for the preparation of this salt was the same as the one used for the synthesis of compound (1). The amounts used were: basic copper carbonate ($CuCO_3.Cu(OH)_2$; 0.96 g, 4.3 mmol) and mellitic acid (0.5 g, 1.4 mmol).

Synthesis of nickel mellitate (4)

Basic nickel carbonate ($2NiCO_3.3Ni(OH)_2.4H_2O$; 0.61 g, 1.05 mmol) and mellitic acid (0,25 g, 0.7 mmol) were mixed in water (5 mL) and heated for 30 min at 90 °C. The product was filtered off, washed with water, ethanol and ethyl ether, and dried at 60 °C.

Synthesis of cobalt mellitate (5)

The procedure for the preparation of this salt was the same as the one used for the synthesis of compound (4). The amounts used in the experiment were: cobalt carbonate

($CoCO_3$; 0.52 g, 4.3 mmol) and mellitic acid (0.5 g, 1.4 mmol).

Synthesis of magnesium mellitate (6)

The procedure for the preparation of this salt was the same as the one used for the synthesis of compound (5). The amounts used in the experiment were: magnesium carbonate hydroxide pentahydrate ($4MgCO_3.Mg(OH)_2.5H_2O$; 0.25 g, 0.5mmol) and mellitic acid (0.25 g, 0.7 mmol).

Synthesis of calcium mellitate (7)

The procedure for the preparation of this salt was the same as the one used for the synthesis of compound (5). The amounts used in the experiment were: calcium carbonate ($CaCO_3$; 0.42 g, 4.3 mmol) and mellitic acid (0.5 g, 1.4 mmol)

Synthesis of iron mellitate (8)

Ferric hydroxide, was precipitated by the addition of an excess concentrated ammonium hydroxide (1,5 mL) to a aqueous solution of ferric chloride ($FeCl_3$, 5 mL, 1 M). The resulting mixture was boiled for 30 minutes, and the precipitated solid washed with hot water and removed by decantation. Mellitic acid (0.56g, 1.6 mmol) and water (5 mL) were then added to the solid ferric hydroxide and this solution was heated for 30 minutes at 90 °C. The product was washed by decantation, filtered and dried in air at 60 °C.

Synthesis of fluoresceins: Fiegl's test

In a typical reaction, the poly-benzecarboxylic acid (phtalic acid, benzene tricarboxylic acid, pyrometllitic acid and mellitic acid) (1 eq) was mixed with resorcinol (2 eq) and heated by 5 min at 120 °C at which time sulfuric acid (0.1 mL) was added. The resulting brown mixture was heated by additional 5 min, then cooled and made alkaline by addition of a concentrated solution of sodium hydroxide (5 M).

Results

Synthesis of mellitic acid salts

The infrared spectra of each of the mellitic acid salts are presented below:

Manganous mellitate (Figure 4.7), zinc mellitate (Figure 4.8), cupric mellitate (Figure 4.9), nickel mellitate (Figure 4.10), cobalt mellitate (Figure 4.11), magnesium mellitate (Figure 4.12), calcium mellitate (Figure 4.13), iron mellitate (Figure 4.14) and aluminium mellitate (Figure 4.15).

Fluorescence spectra Analysis

The Fiegl's test derivatives of the different benzenecaboxylate showed strong fluorescence in alkaline medium. Table 4.2 shows the different excitation and emission signals obtained during the analysis.

Table 4.2 Fluorescence analysis of the fluoresceines of benzenecarboxylates.

Acid	Excitation (nm)	Emission (nm)
Blank	475	----
1,2 benzene dicarboxylic acid (phtalic acid)	475	512
1,2,4 benzenetricarboxylic acid	475	517
1,2,4,5 benzene tetracarboxylic acid (pyromellitic acid)	475	517
Benzene hexacarboxylic acid (mellitic acid)	475	512
Iron phtalate	475	512

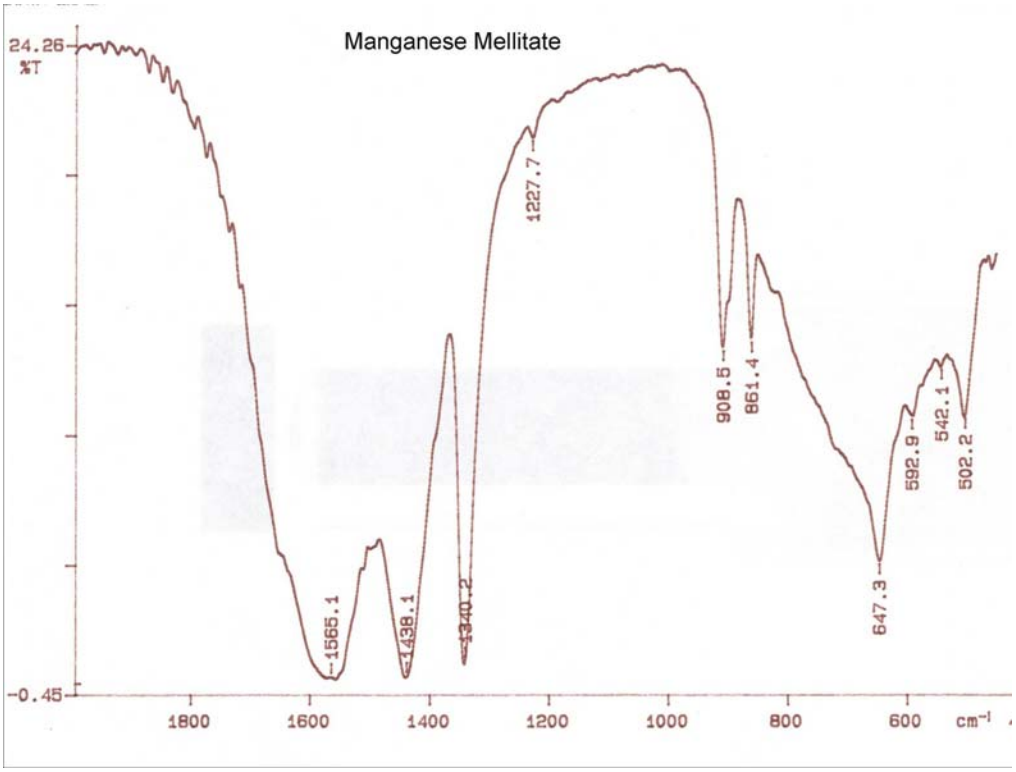


Figure 4.7. Infrared spectra of manganous mellitate (KBr).

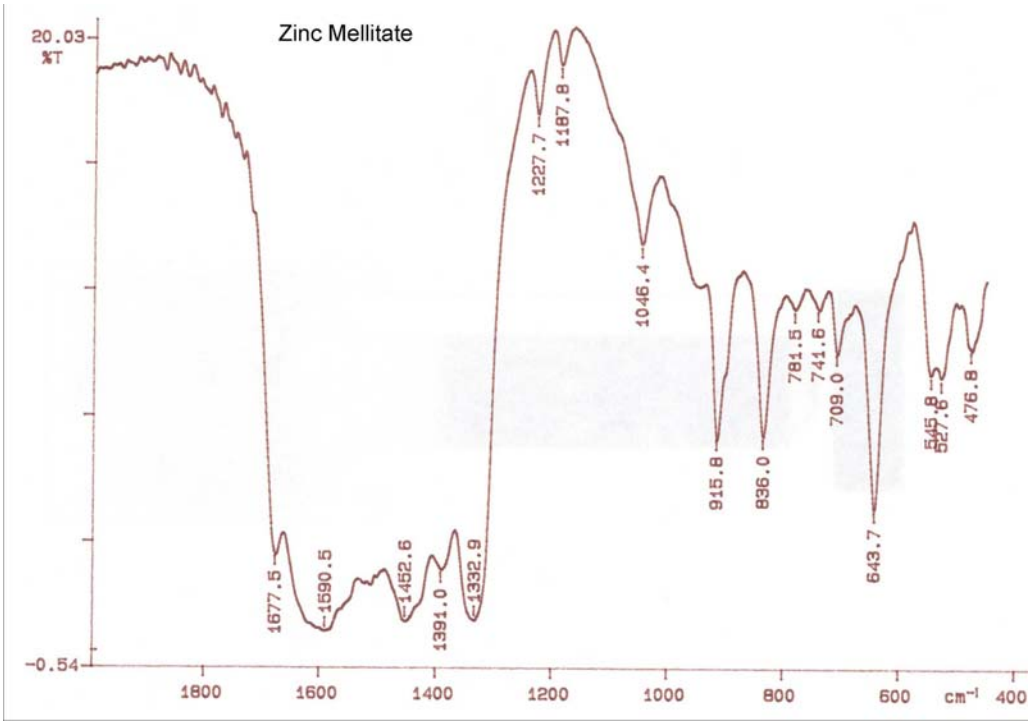


Figure 4.8. Infrared spectra of zinc mellitate (KBr).

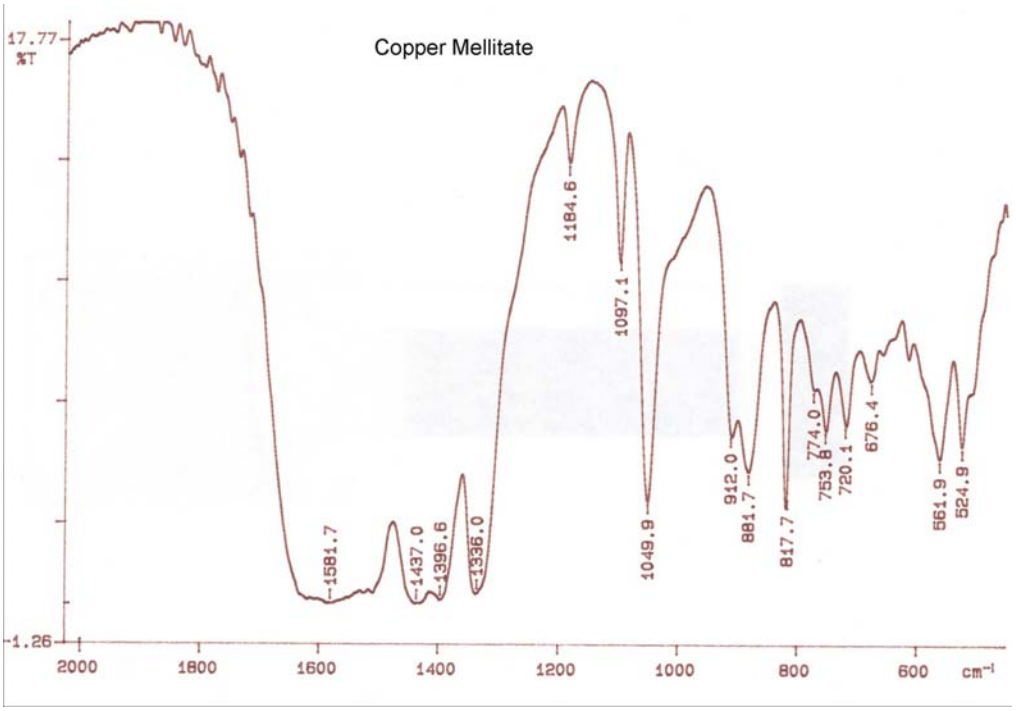


Figure 4.9. Infrared spectra of cupric mellitate (KBr).

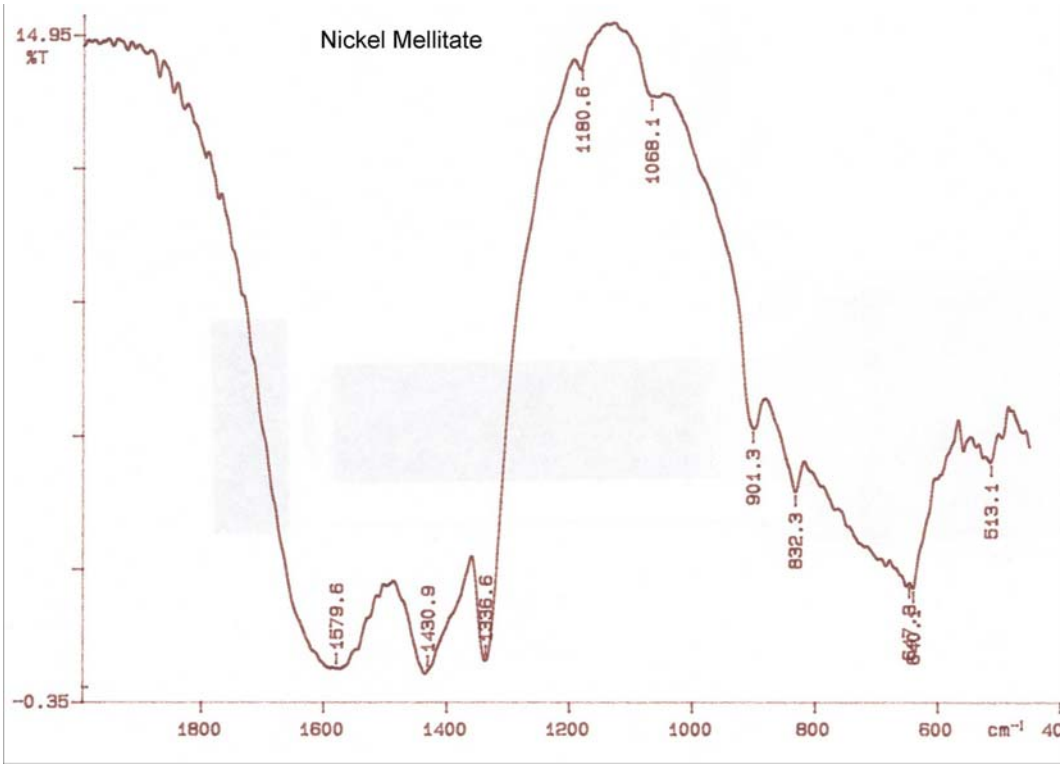


Figure 4.10. Infrared spectra of nickel mellitate (KBr).

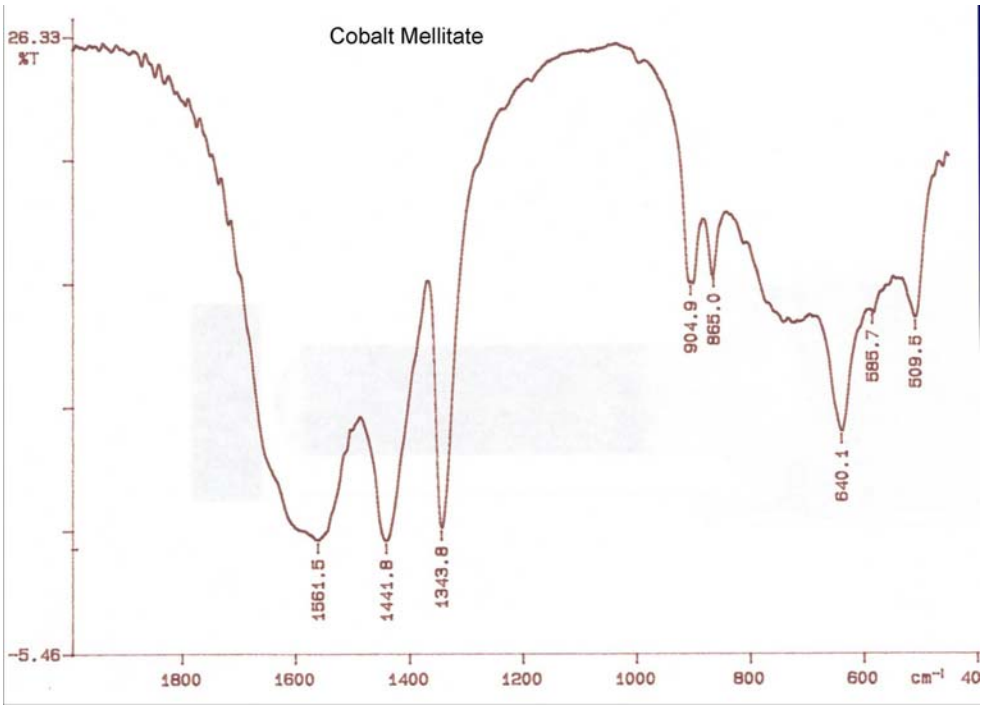


Figure 4.11. Infrared spectra of cobalt mellitate (KBr),

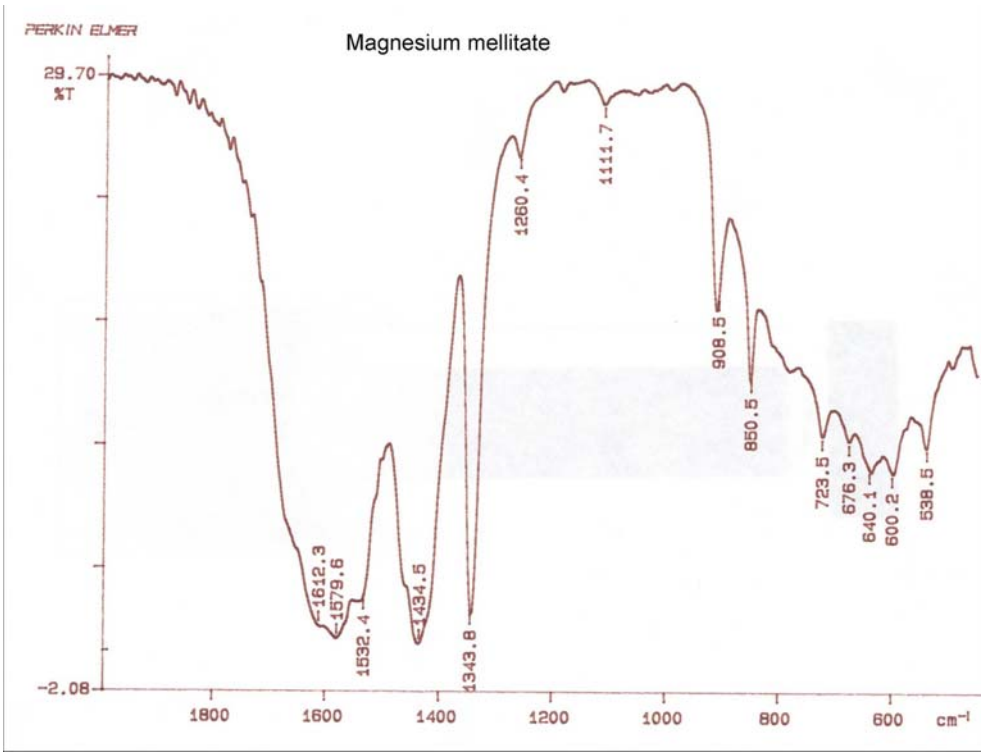


Figure 4.12. Infrared spectra of magnesium mellitate (KBr)

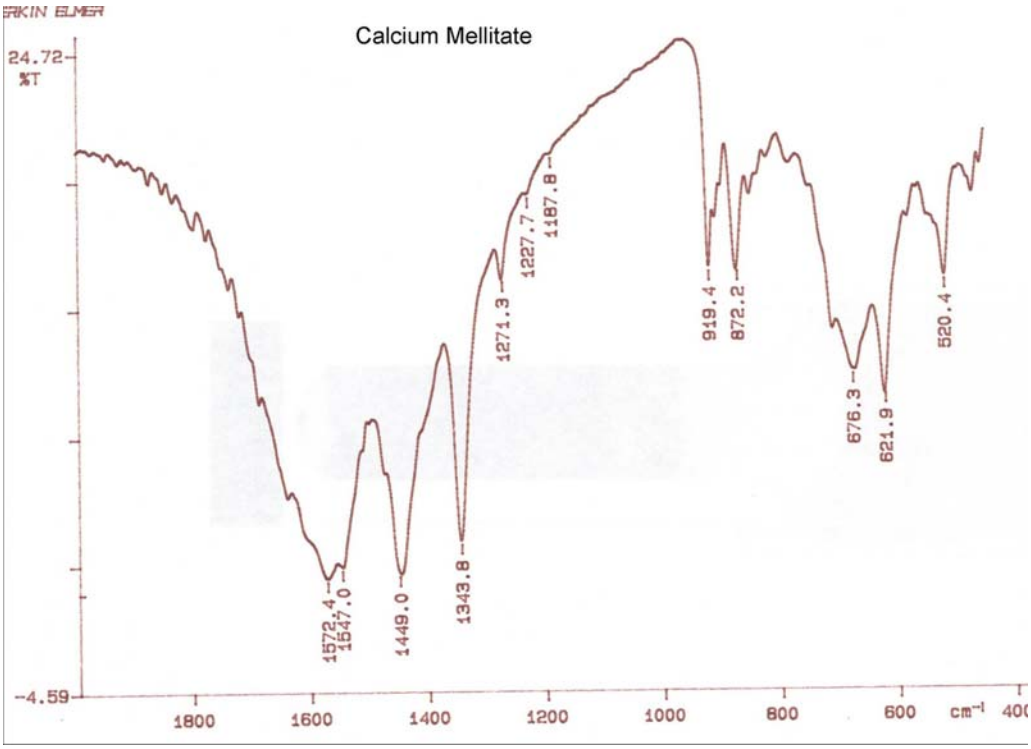


Figure 4.13. Infrared spectra of calcium mellitate (KBr)

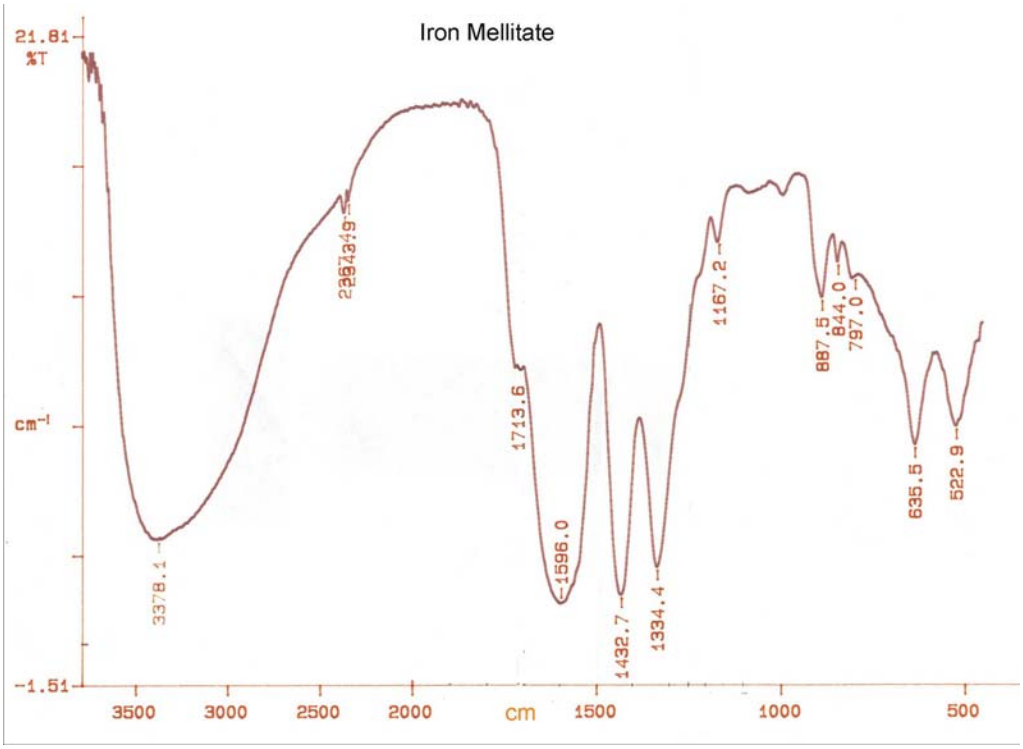


Figure 4.14. Infrared spectra of iron mellitate (KBr)

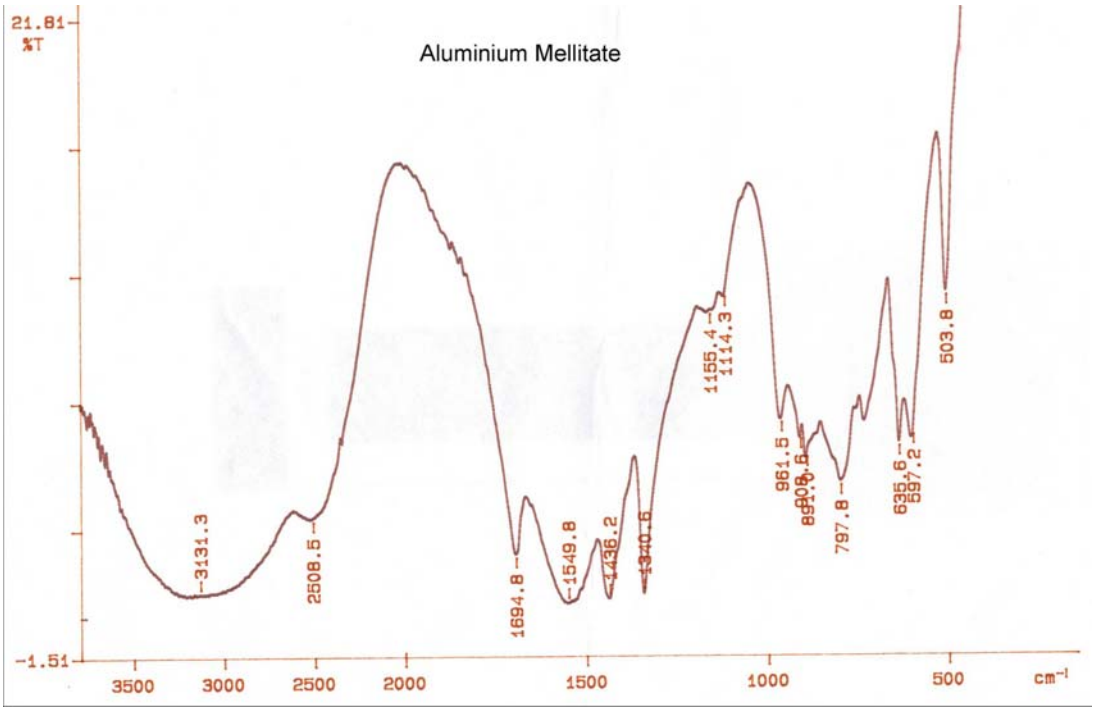


Figure 4.15. Infrared spectra of aluminium mellitate (KBr)

Discussion

A comparison of the Infrared spectra of the different synthetic mellitic carboxylates salts, indicates that all the salts show absorption bands in the region of 1350-1600. A superimposed IR spectrum of the mellitates of the most abundant elements on Mars, with the IR of the surface of Mars is shown in Figure 4.16. Not conclusive arguments can be done about the presence of mellitates in the martian surface in the absence of Martian samples. At the very least, we can say that the resemblance in the IR spectra keeps open the possibility for contemplating the presence of benzene carboxylates on Mars and therefore detection systems that target these compounds should be included when designing future missions to the red planet.

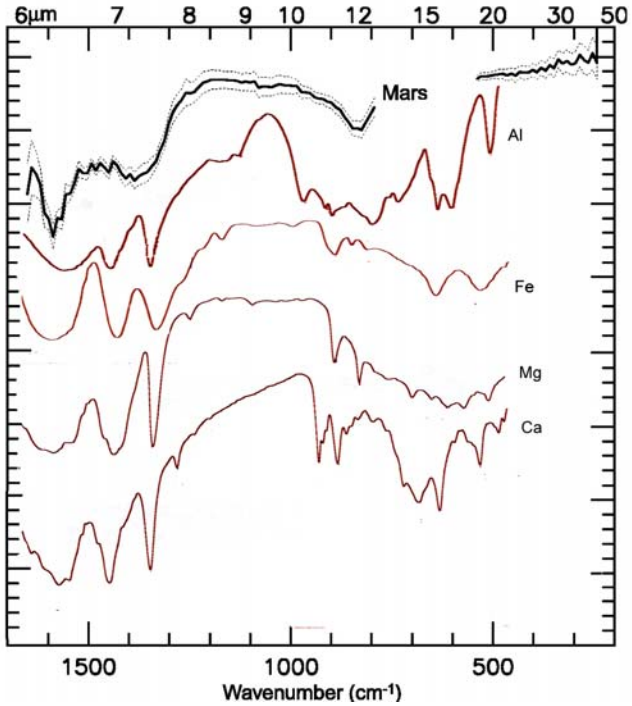
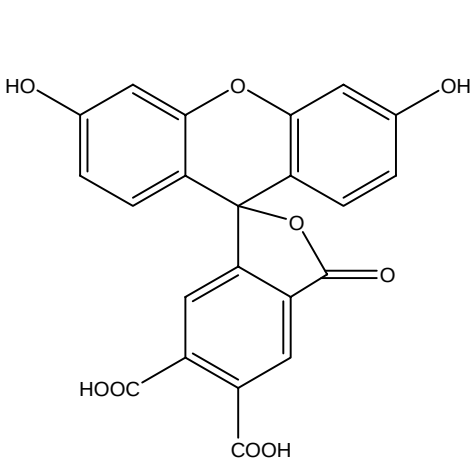


Figure 4.16. Infrared spectra of the mellitate salts of aluminium, iron, magnesium and calcium. Note the appearance of signals in the 1350-1500 region unexplained in previous reports by Banfield et al.

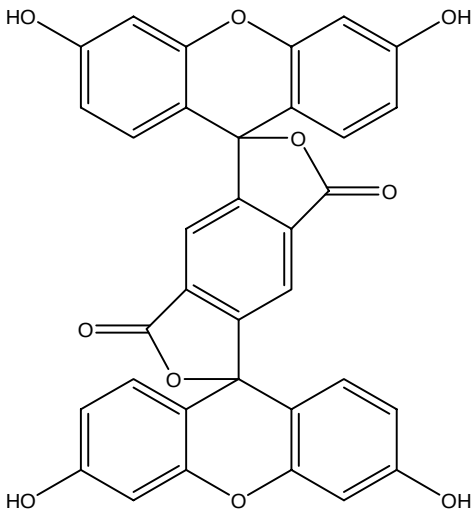
The Fiegl's test analysis shows that all the benzenecarboxylic acids studied were able to produce fluorescent derivatives, allowing identification even when coordinate to metals as in the carboxylates. The structure of the different fluoresceines was confirmed by mass-spectrometry and is consistent with the addition of two resorcinol molecules to the dicarboxylic acid moiety in benzenecarboxylates.

When resorcinol was used in large excess over the 1,2,4,5 tetracarboxylic acid, two major products were observed after HPLC-ESI analysis. These products correspond to the expected mono and di-fluorescein derivatives of the tetracarboxylic acid (Figure 4.17).

In general all the fluorescein derivatives show a similar spectrum. The intensity of the emission decreases as successive carboxyl substituents are added to the benzene core .



1,2-pyromellitic fluorescein



pyromellitic di-fluorescein

Figure 4.17. Structures of the fluorescein derivatives of pyromellitic acid (1,2,4,5-benzenetetracarboxylate) detected by HPLC-MS (ESI).

The absorption and emission spectra for the different derivatives (Table 4.2) show that the specific maximum emision spectral line for each compound makes it possible to distinguish and identify each individual fluorescein. The sensitivity of the method allows it to be a powerful tool for the search of oxidized organic molecules rising from kerogens.

This approach is suitable for both Earth–bound research on SNC meteorites and *in situ* studies on Mars.

APPENDIX A
NMR DEGRADATION EXPERIMENTS

Half life of the different isomeric aldopentoses was calculated by measuring the ratio of anomeric proton/ internal standard (integral value) at different time points.

In some cases several half life times are reported for the same aldopentose; this is caused by different rates of decay of the different anomeric proton signals. These phenomena could be explained by considering preferential protection of the different borate complexes possible for each aldopentose.

Because the rate of decomposition in each case is proportional to the amount of pentose in the aldehydic form, those boron-pentose complexes that stabilize the closed ring furanose form (locking it) will be more stable towards decomposition.

Another possible explanation is that the aldopentoses are interconverting (into another pentose or the corresponding ketose) and this alkaline isomerization is catalyzed by borate.

^{1}H NMR spectra were superimposed using the software package MestRe-C 2.3a obtained free of charge at: http://www.mestrec.com/.

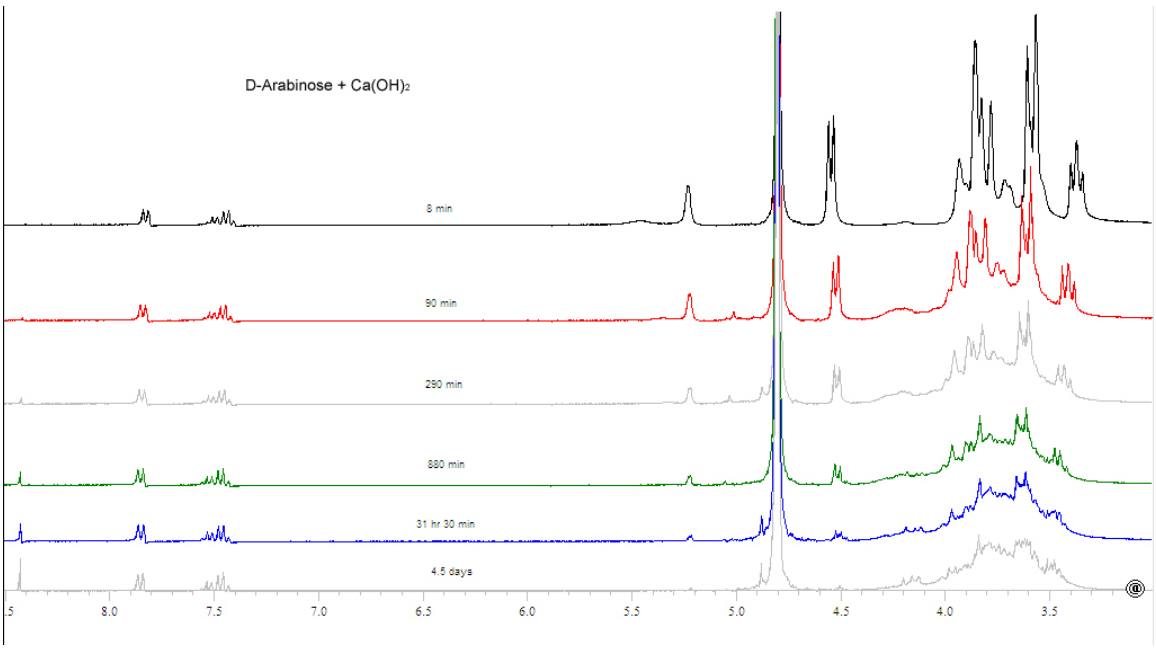


Figure A1.1. D-arabinose incubation in the presence of calcium hydroxide, pD:12.

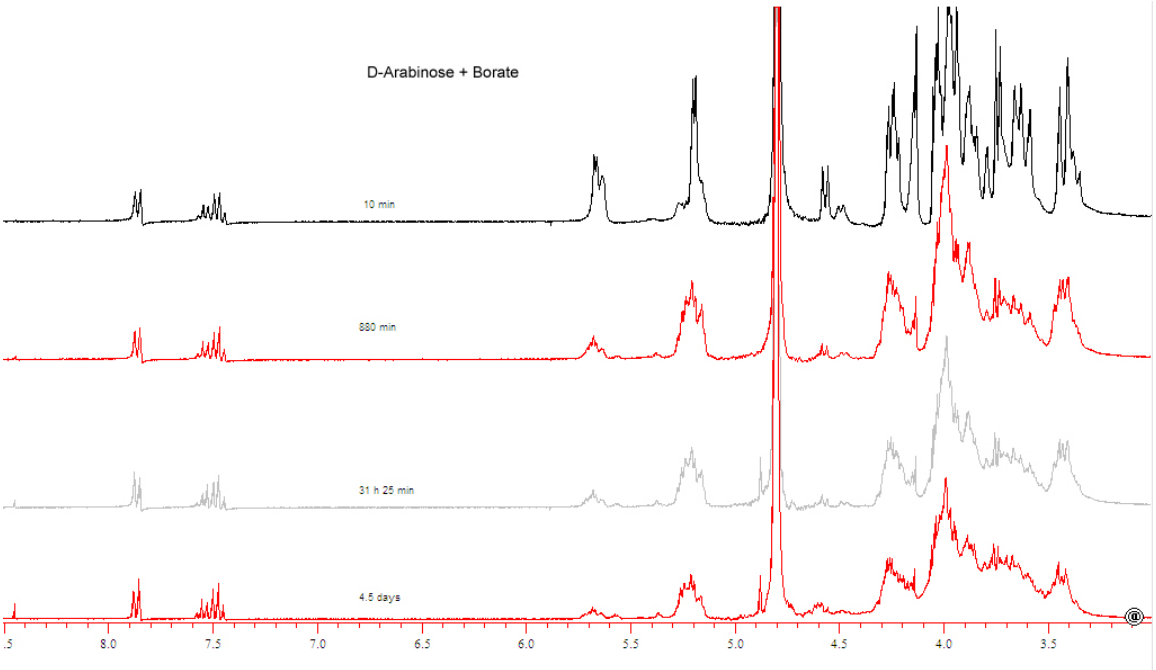


Figure A1.2. D-arabinose incubation in the presence of calcium-hydroxyde and borate, pD:12

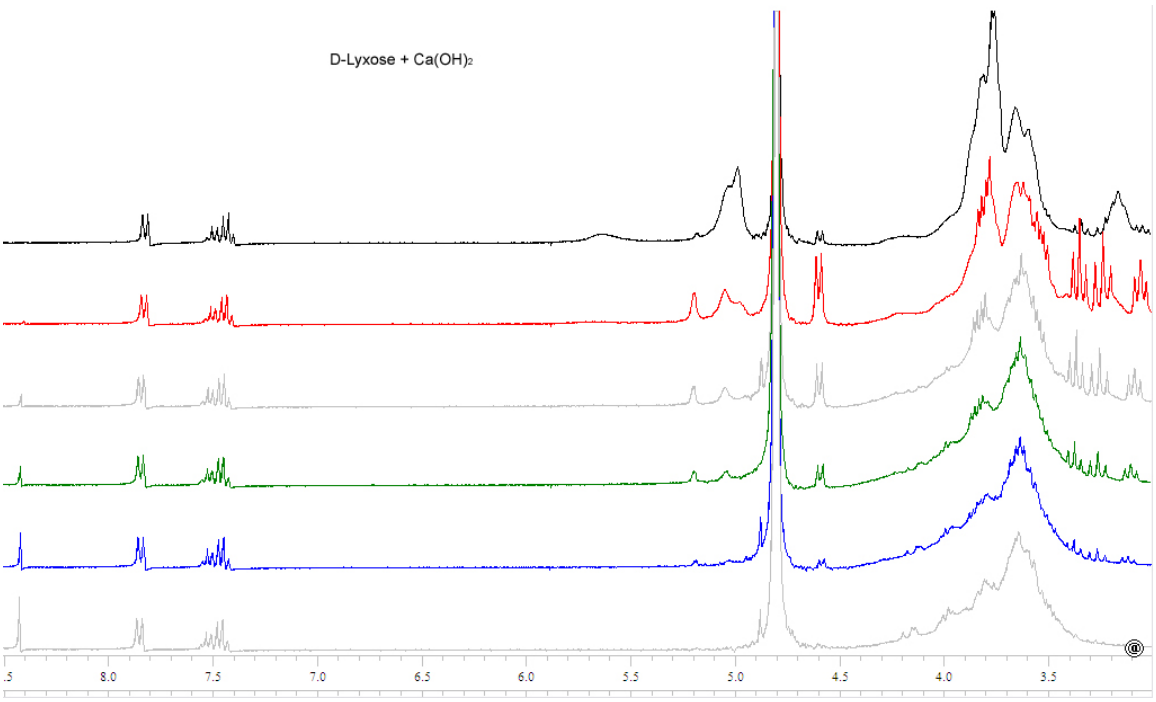


Figure A1.3. D-lyxose incubation in the presence of calcium hydroxide, pD:12.

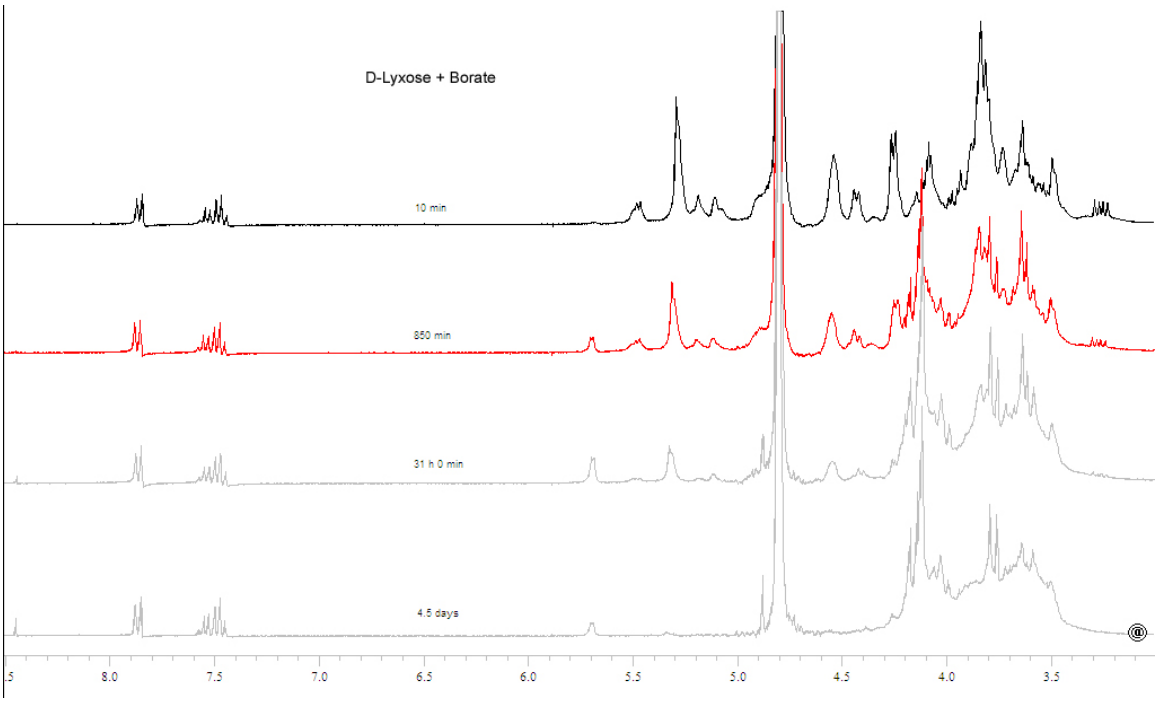


Figure A1.4. D-lyxose incubation in the presence of calcium hydroxide + borate , pD:12.

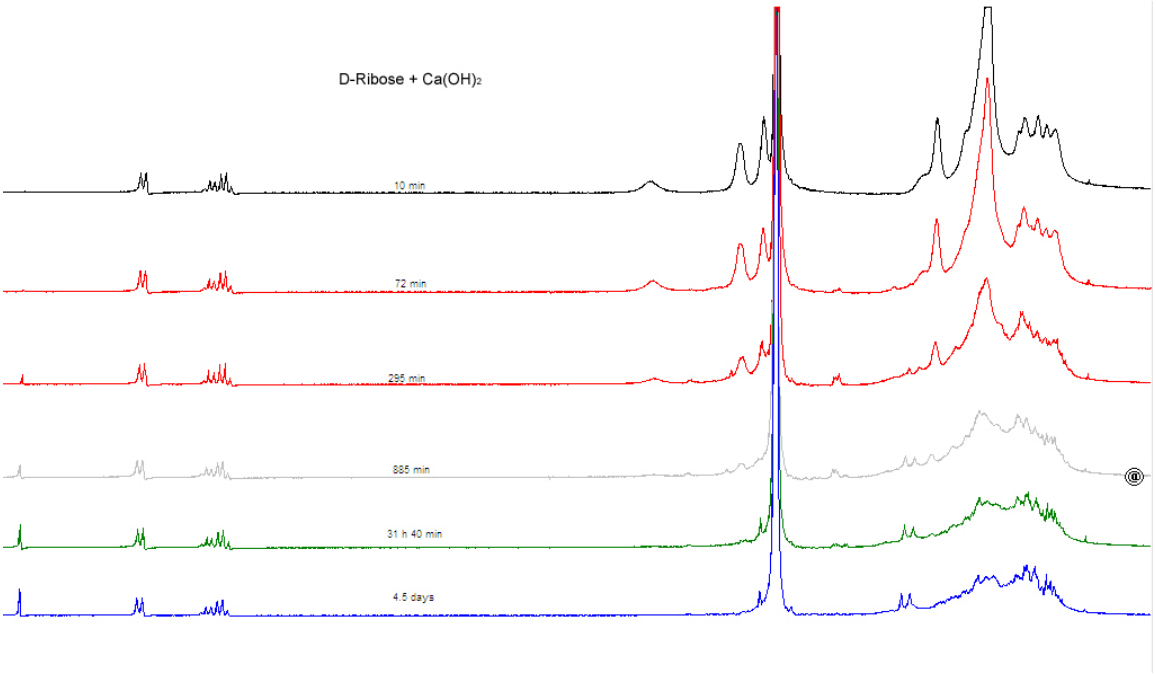


Figure A1.5. D-ribose incubation in the presence of calcium hydroxide, pD:12.

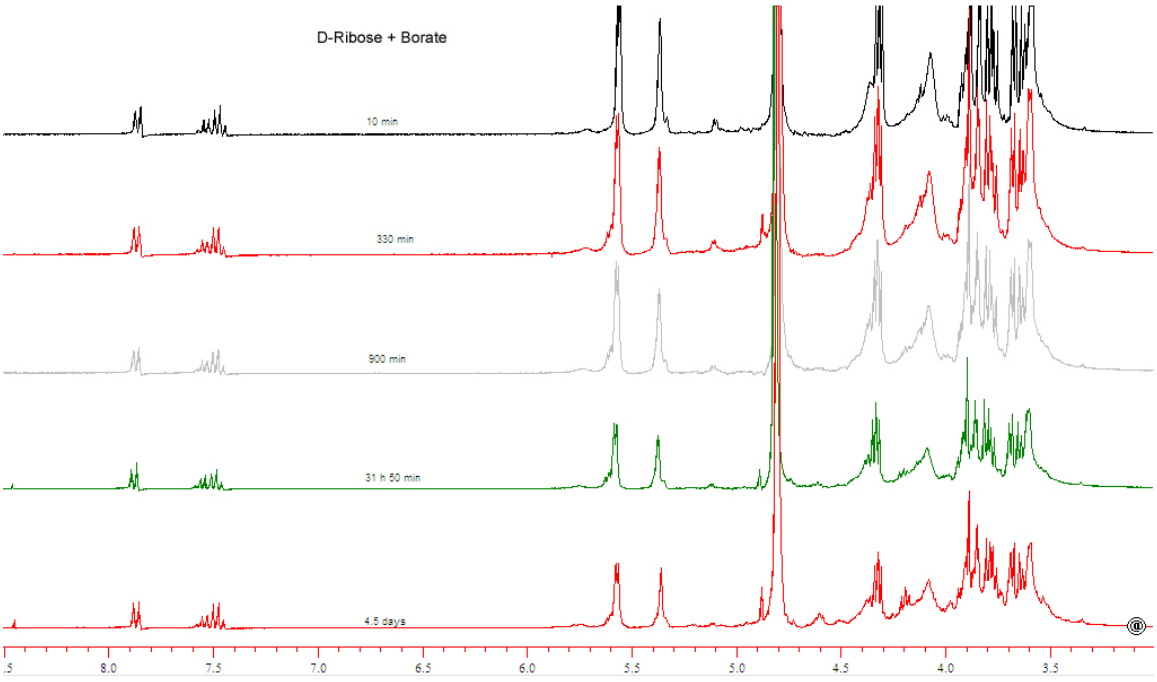


Figure A1.6. D-ribose incubation in the presence of calcium hydroxide + borate, pD:12.

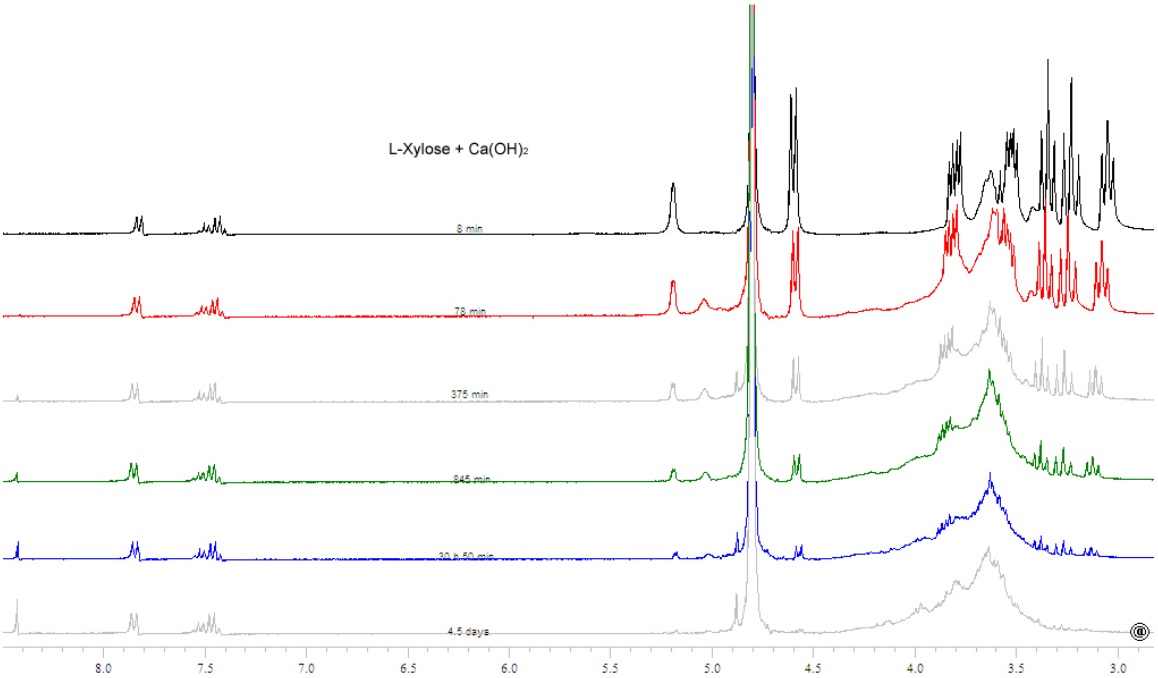


Figure A1.7. L-xylose incubation in the presence of calcium hydroxide, pD:12.

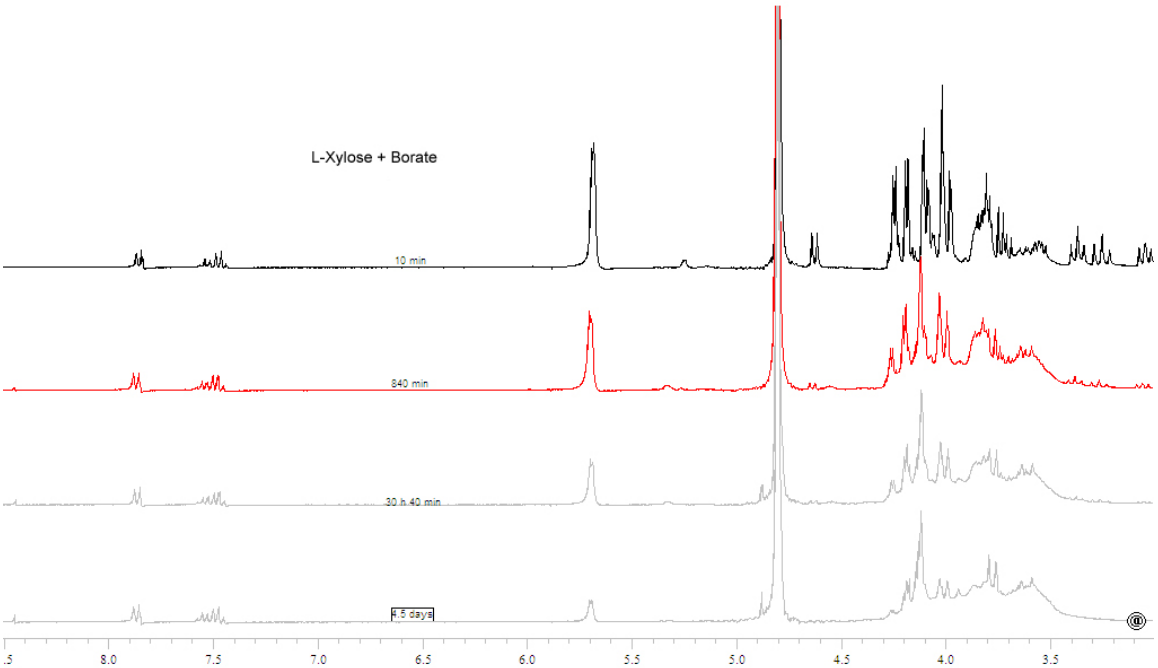


Figure A1.8. L-xylose incubation in the presence of calcium hydroxide + borate, pD:12.

APPENDIX B
DIOS COMPETITION EXPERIMENTS

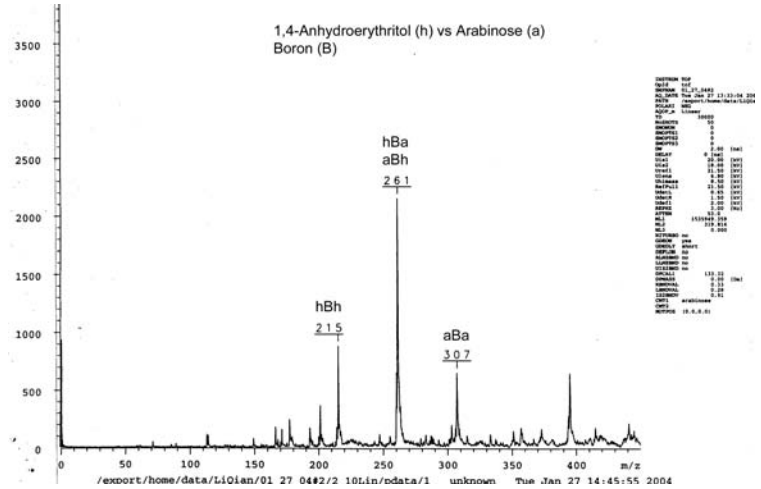


Figure A2.1. DIOS spectra of 1,4-Anhydroerythritol vs arabinose.

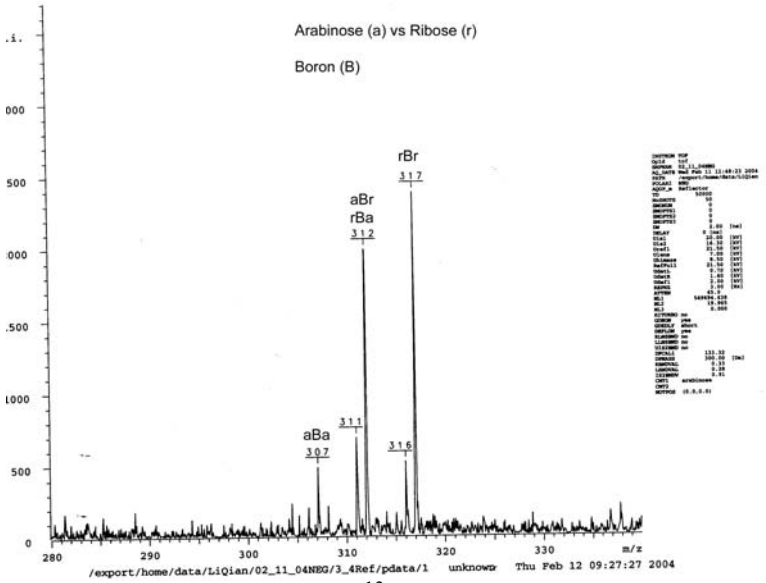


Figure A2.2. DIOS spectra of arabinose vs ^{13}C-ribose.

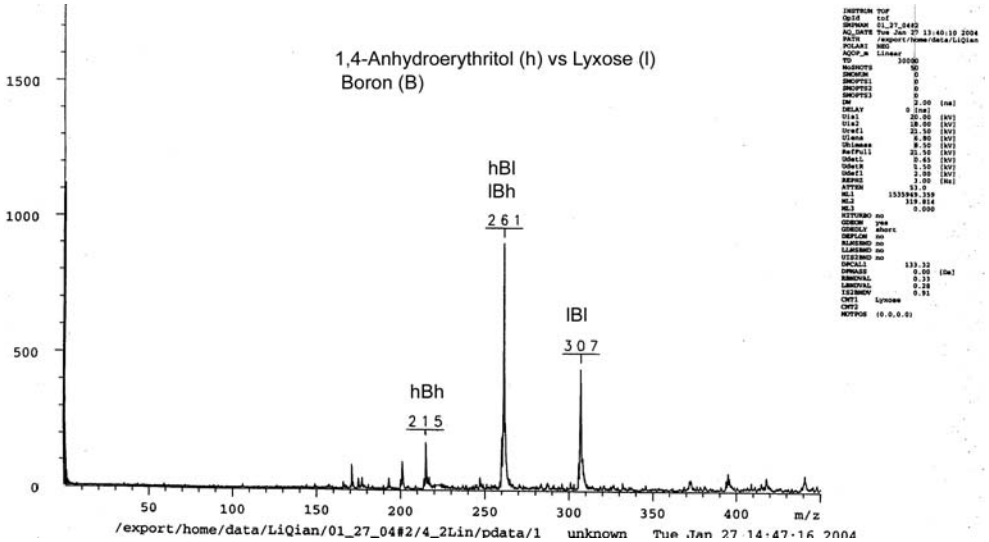


Figure A2.3. DIOS spectra of 1,4-Anhydroeythritol vs lyxose

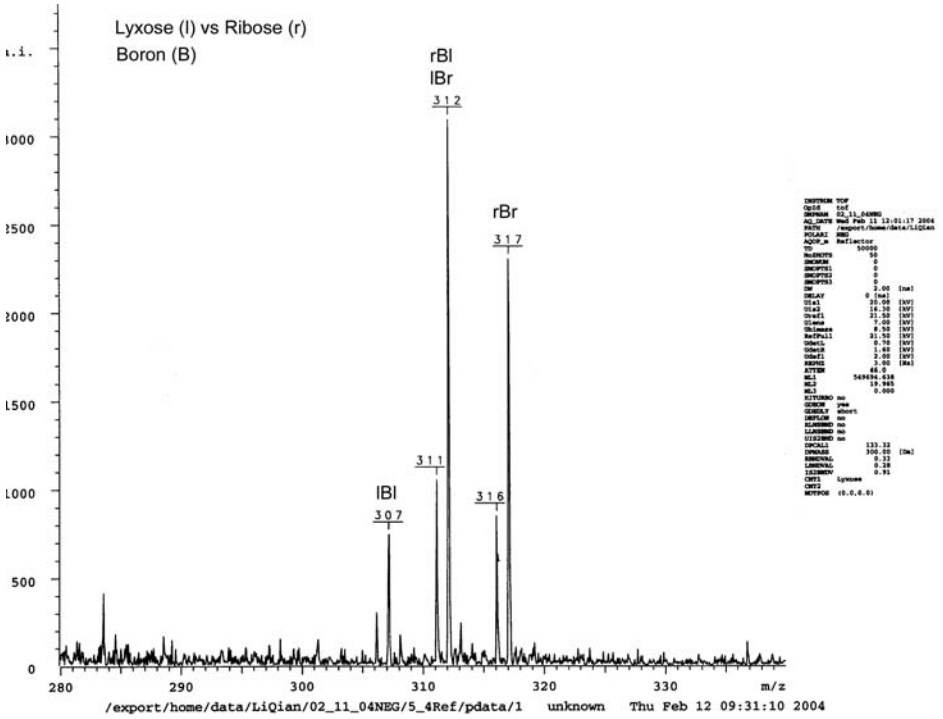


Figure A2.4. DIOS spectra of lyxose vs ^{13}C-ribose

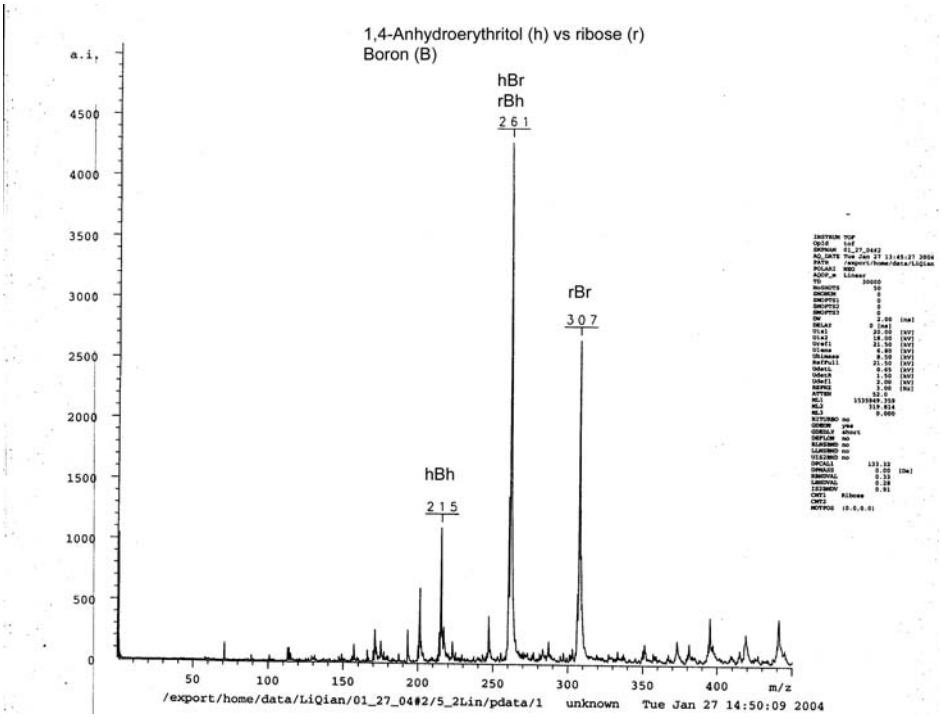


Figure A2.5. DIOS spectra of 1,4-Anhydroerythritol vs ribose

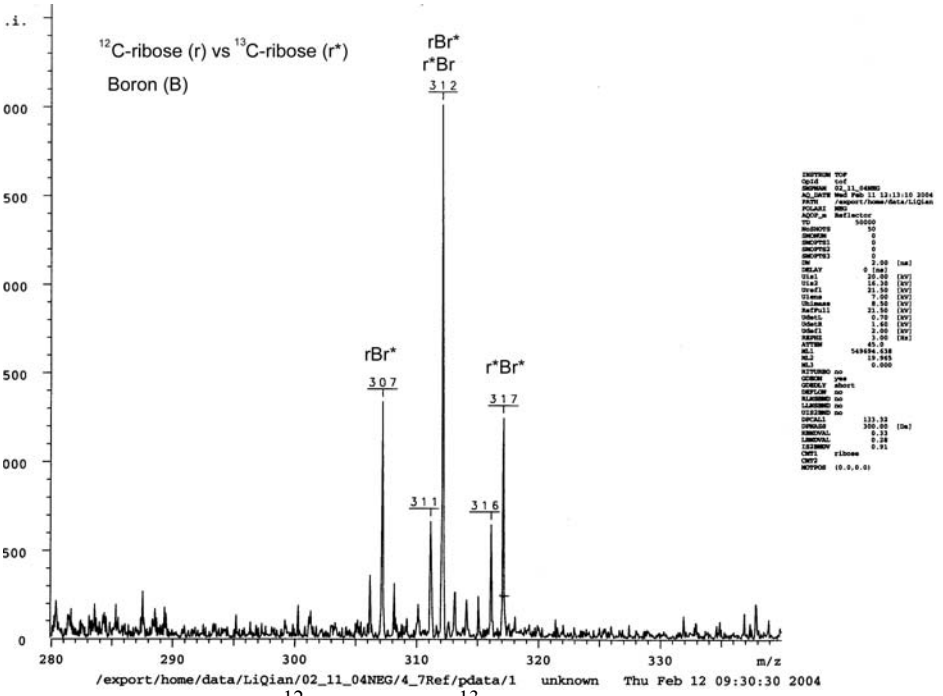


Figure A2.6. DIOS spectra of ^{12}C-ribose vs ^{13}C- ribose

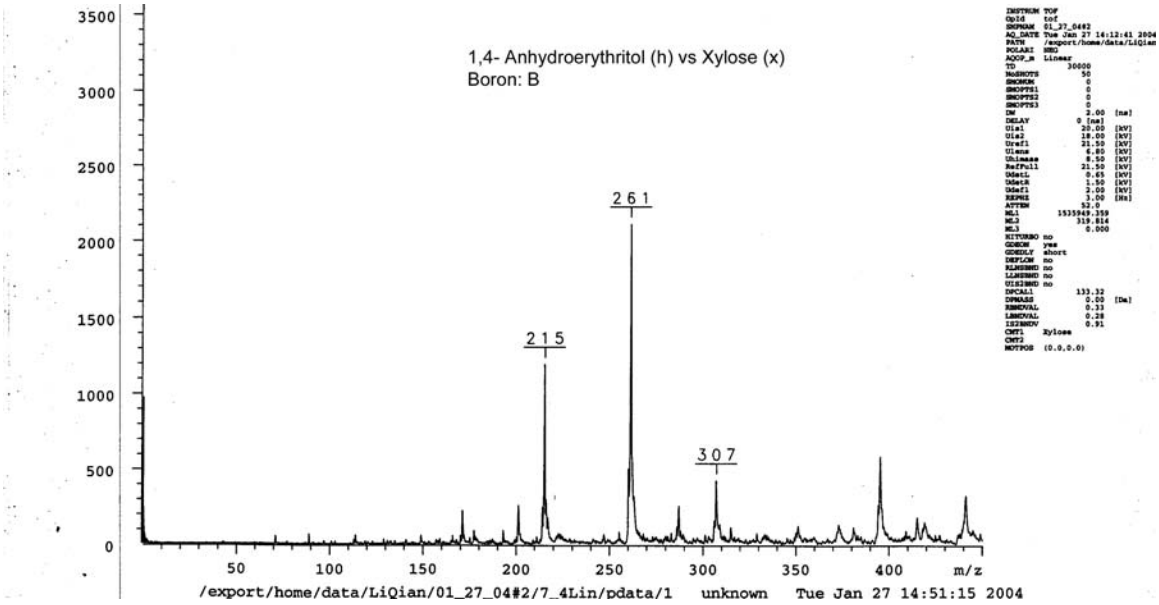


Figure A2.7. DIOS spectra of 1,4-Anhydroerythritol vs Xylose

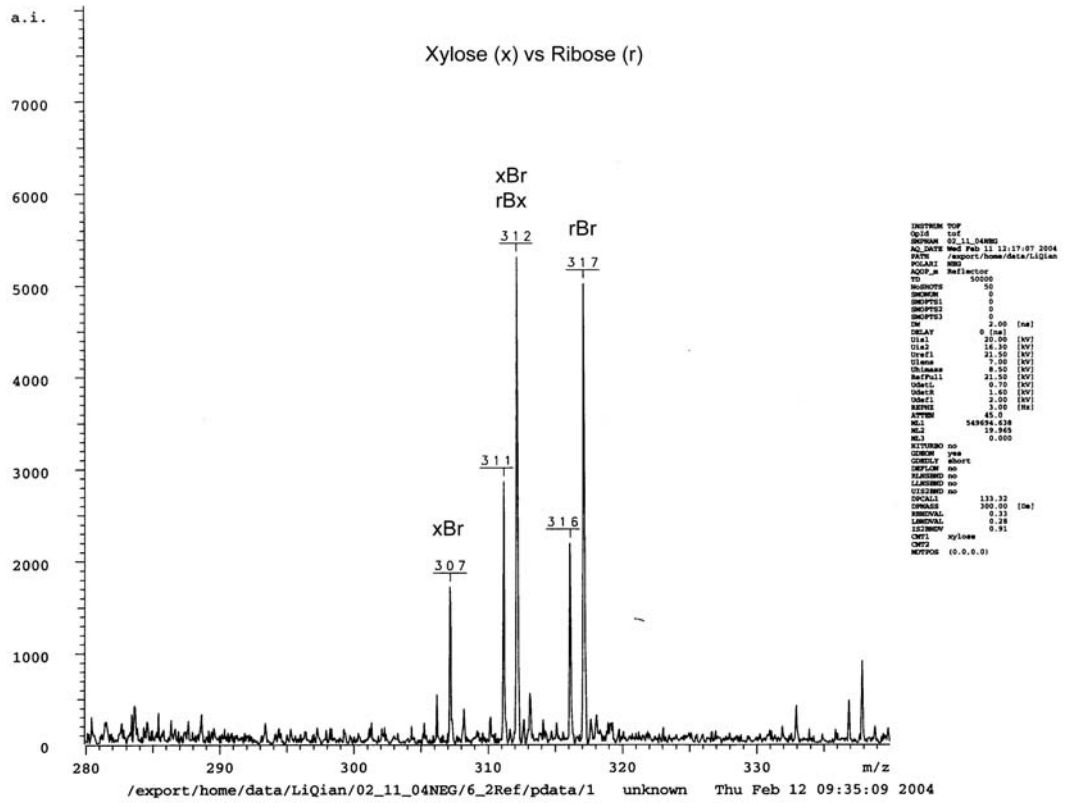


Figure A2.8. DIOS spectra of xylose vs ^{13}C-ribose

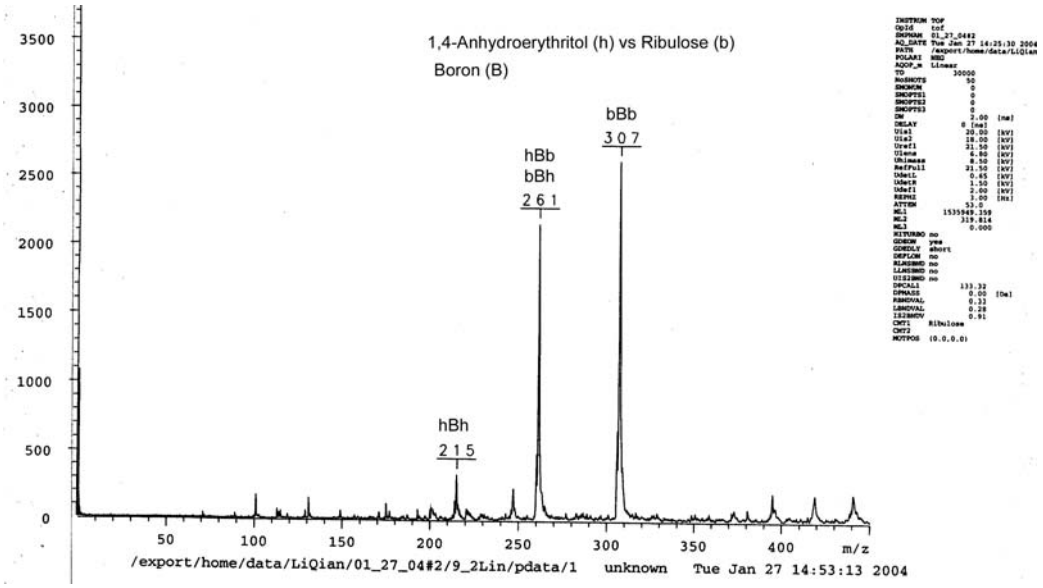


Figure A2.9. DIOS spectra of 1,4-anhyroerythritol vs ribulose

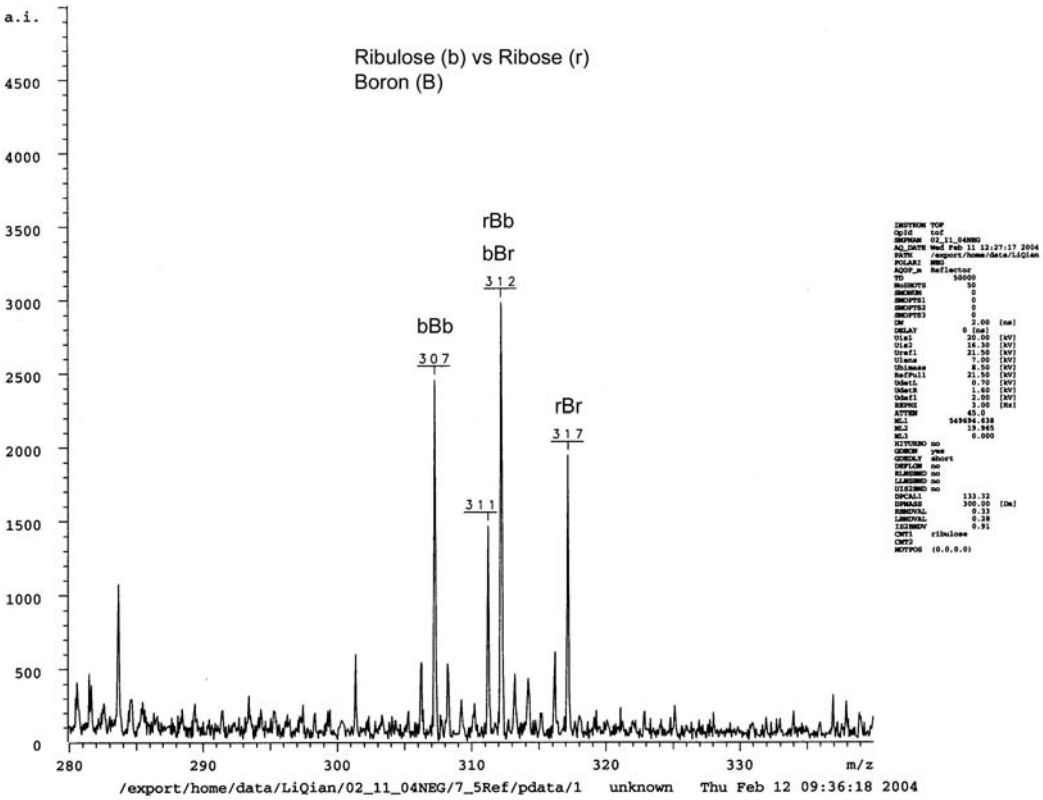


Figure A2.10. DIOS spectra of ribulose vs ^{13}C-ribose

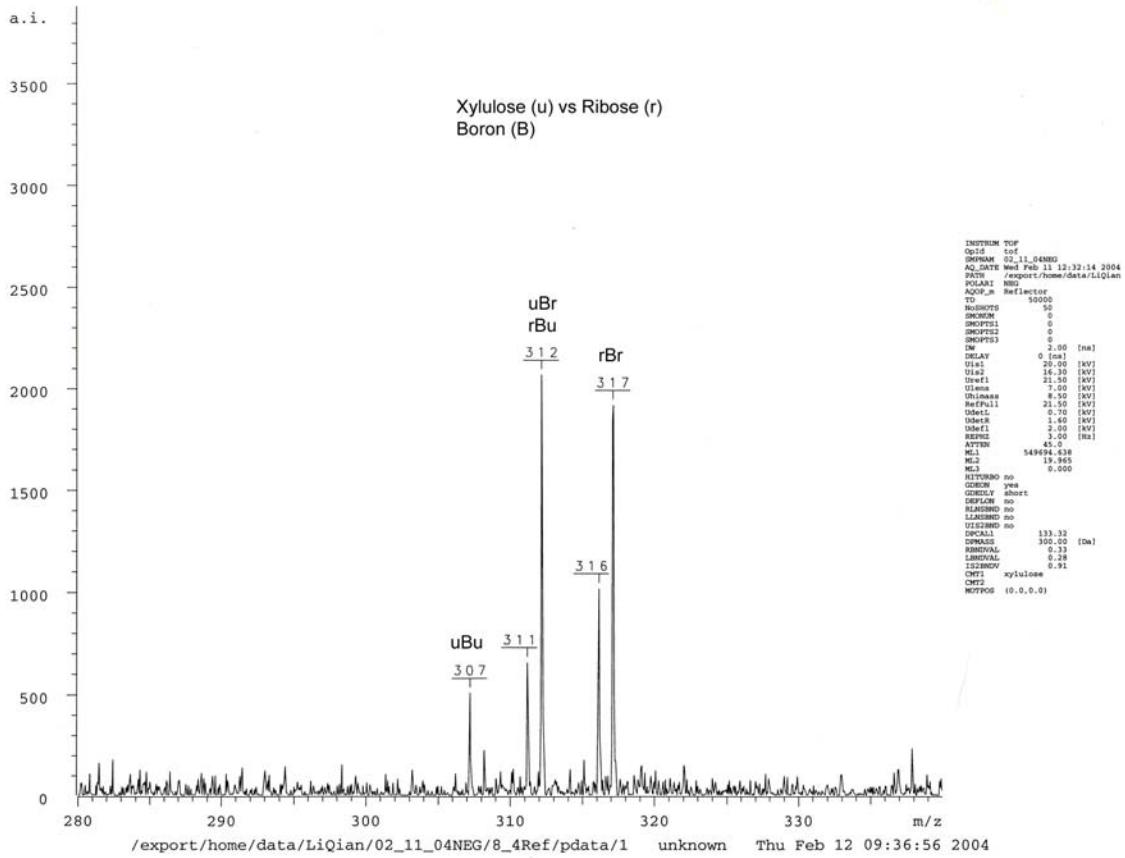


Figure A2.11. DIOS spectra of Xylulose vs ^{13}C-ribose.

LIST OF REFERENCES

Avdeef, A. (1993) pH-Metric Log-P 2. *J Pharm Sci* **82**, 183-190.

Balabanovich, A., Schnabel, W. (1998) *J. Photochem. Photobiol. A-Chem.* **113**, 145-153.

Balabanovich, A. I., Denizligil, S., Schnabel, W. (1997) *J. Vinyl Additive Technol.* **3**, 42-52.

Barbas, J.T., Sigman, M. E., Dabestani, R. (1996) *Environ. Sci. Technol.* **30,** 1776-1780.

Bartel, D. P., Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science,* 261, 1411-1418

Benner, S. A., Allemann, R. K., Ellington, A. D., Ge, L., Glasfeld, A., Leanz, G. F., Krauch, T., MacPherson, L. J., Moroney, S. E., Piccirilli, J. A., Weinhold, E. G. (1987) Natural selection, protein engineering and the last riboorganism. Rational model building in biochemistry Cold Spring Harbor Symp. *Quant. Biol.* 52, 53-63.

Biemann, K., Oró, J., Toulmin III, P., Orgel, L. E., Nier, A. O., Anderson, D. M., Simmonds, P. G., Flory, D., Diaz, A. V., Rushneck, D. R., Biller, J. E., Lafleur, A. L. (1977) *J. Geophys. Res.* **82**, 4641-4658.

Böeseken, J. (1949) The use of boric acid for the determination of the configuration of carbohydrates. *Adv. Carbohydr. Chem*. 4, 189-210.

Bone, W. A., Horton, L., Ward, S. G. (1930) *Proc. Roy. Soc.* **127A** 480-510.

Breaker, R.R. (2000) Molecular biology—Making catalytic DNAs. *Science*. 290, 2095-2096.

Breaker, R. R., Joyce, G. F. (1994) A DNA enzyme that cleaves RNA. *Chem. Biol*. 1, 223-229.

Breaker, R. R., Joyce, G. F. (1995) A DNA Enzyme with Mg2+-Dependent RNA Phosphoesterase Activity. *Chem. Biol.* 2, 655-660

Breslow, R. (1959). On the mechanism of the formose reaction. *Tetrahedron Lett*. 21, 22-26

Bunce, N. J., Liu, L., Zhu, J., Lane, D. A. (1997) *Environ. Sci. Technol.* **31,** 2252-2259.

Bullock, M. A., Stoker, C. R., McKay, C. P., Zent, A. P. (1994) *Icarus* **107**, 142-154.

Butlerow, A. (1861). Bildung einer zuckerartigen Substanz durch Synthese. *Ann*. 120, 295-298

Carmi, N., Shultz, L. A., Breaker, R. R. (1996) In vitro selection of self-cleaving DNAs. *Chem. Biol*. 3, 1039-1046.

Chen, C. T., Tafuri, A. N., Rahman, M. & Foerst, M. B. (1998) *J. Environ. Sci. Health Part A-Toxic/Hazardous Substances & Environ. Eng.* **33**, 987-1008.

Chyba C., Sagan C. (1992) *Nature* **355,** 125-132.

Crick, F. H. C. (1968) The origin of the genetic code. *J. Mol. Biol*. 38. 367-379

Davis, R., Schultz, H. P. (1962) *J. Org. Chem.* **27**, 854-857.

Dean, R. T., Nicholson, P. (1994) *Free Radical Res.* **20**, 83-101.

Decker, P., Schweer, H., Pohlmann, R. (1982). X. Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography-mass spectrometry of n-butoxime trifluoroacetates on OV-225. *J. Chromatogr*. 244, 281-291

Faulhammer, D., Famulok, M. (1996) The Ca2+ ion as a cofactor for a novel RNA-cleaving deoxyribozyme. *Angew. Chem. Int. Edit*. 35, 2837-2841.

Faulhammer, D., Famulok, M. (1997) Characterization and divalent metal-ion dependence of in vitro selected deoxyribozymes which cleave DNA/RNA chimeric oligonucleotides. J. *Mol. Biol*. 269, 188-202.

Feigl, F. (1961) *Spot test in organic analysis*. Elsevier, Amsterdam, 256-258.

Flynn, G. J. (1996) *Earth, Moon, and Planets* **72**, 469-474.

Frick, L., MacNeela, J. P., Wolfenden, R. (1987) *Bioorg. Chem. 15*, 100.

Gabel, N. W., Ponnamperuma, C. (1967) *Nature*. 216, 453

Galwey, A. K. (1965) *J. Chem. Soc.* 4235-4239.

Geyer C. R., Sen, D. (1997) Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. *Chem. Biol*. 4, 579-593.

Giammar, D.E., Dzombak, D.A. (1998) *J. Solution Chem.* **27**, 89-105.

Gilbert, W. (1986) The RNA world. *Nature.* 319**,** 818

Granito, C., Schultz, H. P. (1963) *J. Org. Chem.* **28**, 879-881.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*.35. 849-857

Guillard, C., Delprat, H., Can, H. V., Pichat, P. (1993) *J. Atmos. Chem.* **16**, 47-59.

Hayatsu, R., Anders, E. (1981) *Topics Curr. Chem.* **99**, 1-37.

Hayatsu, R., Matsuoka, S., Scott, R. G., Studier, M. H., Anders, E. (1977) *Geochim. Cosmochim. Acta*. **41**, 1325-1339.

Hayatsu, R., Winans, R. E., Scott, R. G., McBeth, R. L., Moore, L. P., Studier, M. H. (1980) *Science* **207**, 1202-1204.

Hayes, J. M., Biemann, K. (1968) *Geochim. Cosmochim. Acta* **32**, 239-267.

Heberger, K., Nemeth, A., Cotarca, L., Delogu, P. (1994) *Applied Catalysis A-General* **119,** L7-L12.

Henry, L. (1895). *Compt. Rend*. 121, 210

Horowitz, N. H., Hobby, G. L. (1977) *J. Geophys. Res.* **82**, 4659.

Hubbard, J. S., Hardy, J., Voecks, G. E., Golub, E. E. (1973) *J. Molec. Evol.* **2**, 149-166.

Hubbard, J. S., Hardy, J. P., Horowitz, N. H. (1971) *Proc. Nat. Acad. Sci*. **3**, 574.

Hunten, D. M. (1974) *Rev. Geophys. Space Phys.* **12**, 529-535.

Hunten, D. M. (1979) *J. Molec. Evol.* **14**, 57-64.

Ito, H., Kono, Y., Machida, A., Mitsumoto, Y., Omori, K., Nakamura, N., Kondo, Y., Ishihara, K. (2003) Kinetic study of the complex formation of boric and boronic acids with mono- and diprotonated ligands. *Inorg. Chim. Acta*. 344. 28-36.

Jeevarajan, A. S., Fessenden, R.W. (1992) *J. Am. Chem. Soc.* **114**, 10461-10470.

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994) *Continuous Univariate Distributions* Vol 1, Second ed., N.Y., Wiley, Chapter 21

Juettner, B. (1937) *J. Am. Chem. Soc.* **59**, 208-213.

Kieffer, S. W., Jakosky, B. M., Snyder, C. W., Matthews, M. S., editors (1992) *Mars* Tucson, University of Arizona Press.

Kirsten, W., Stenhagen, E. (1952) *Acta. Chem. Scand.* **6**, 682-689.

Klein, H. P. (1978) *Icarus*, **34**, 666-674.

Kore, A. R., Vaish, N. K., Morris J. A., Eckstein, F. (2000) In vitro evolution of the hammerhead ribozyme to a purine-specific ribozyme using mutagenic PCR with two nucleotide analogues. *J. Mol. Biol*. 301, 1113-1121.

Kruger, K., Grabowski, P. J., Zaug, A.J., Sands, J.; Gottschling, D. E.; Cech, T. R. (1982) Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*. 31. 147-157

Kustin, K., Pizer, R. (1968) Temperature-Jump study of the rate and mechanism of the boric acid-Tartaric acid complexation. *J. Am. Chem. Soc*. 91. 317-322

Laidler, K. J., Peterman, B. K. (1979) *Methods Enzymol*. 63, 234-257.

Lamrini, R., Lacan, P., Francina, A., Guilluy, R., Desage, M., Michon, J., Becchi, M., Brazier, J. L. (1998) *Free Radical Biol. Medi.* **24**, 280-289.

Lane, D. A., Fielder, S. S., Townsend, S. J., Bunce, N. J., Zhu, J., Liu, L., Wiens, B., Pond, P.(1996) *Polycyclic Aromatic Compounds* **9**, 53-59.

Larralde, R. Robertson, M. P., Miller, S. L. (1995) Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proc. Nat. Acad. Sci.* 92. 8158-8160.

Lee, D.Y., Martin, J. C. (1984). Compounds of pentacoordinate (10-B-5) and hexacoordinated (12-B-6) Hypervalent Boron. *J. Am. Chem. Soc*. **106**. 5745-5746

Levene, P. A., Jacobs, W. A. (1909). *Ber*. 42. 2102. 3247

Levene, P. A., London, E. S.(1929). *J. Biol. Chem*. 83. 793-802

Levin, G. V., Straat, P. A. (1979) *Science* **194**, 1322-1329.

Levin, G. V. (1997) *Instruments, Methods, and Missions for the Investigation of Extraterrestrial Microorganisms*, *Proc. Internat. Soc. Opt. Engnrg., Proc. Series*, **3111**, 146-161.

Li, Z. M., Comfort, S. D., Shea, P. J. (1997) *J. Environ. Quality* **26,** 480-487.

Lin, S. S., Gurol, M. D. (1998) *Environ. Sci. Technol.* **32**, 1417-1423.

Loew, O. (1886). *J. Prakt. Chem*. 33, 321

Lowenheim, F. A., Moran, M. K., eds. (1975) *Faith, Keyes and Clark's Industrial Chemicals*. 4th Ed. N.Y., Wiley Intersciences.

McDonald, G. D., de Vanssay, E., Buckley, J. R. (1998) *Icarus* **132**, 170-175.

McKay, D. S., Gibson, K.E., Thomas-Keprta, K.L., Vali, H., Romanek, C. S., Clemett, S. J., Chillier, X., Maechling, C. R., Zare, R. N. (1996) Search for past life on Mars: Possible Relic Biogenic Activity in Martian Meteorite ALH84001. *Science*. 273. 924-930.

McKay, C. P., Grunthaner, F. J., Lane, A. L., Herring, M., Bartman, R. K., Ksendzov, A., Manning, C. M., Lamb, J. L., Williams, R. M., Ricco, A. J., Butler, M. A., Murray, B. C., Quinn, R. C., Zent, A. P., Klein, H. P., Levin, G. V. (1998) *Planet. Space Sci.* **46**, 769-777.

Maniatis, T., Fritsch, E.F., Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Manion, J. A., McMillen, D. F. & Malhotra, R. (1996) *Energy & Fuels* **10**, 776-788.

Martell, A. E. & Smith, R. M. (1977) *Critical Stability Constants Vol 3. Other Organic Ligands.* New Yourk, Plenum.

Mazurek, M., Perlin, A. S. (1963) Borate complexing by five-membered-ring vic-diols. *Can. J. Chem*. 41. 2403.

Miller, S. L. (1953). Production of aminoacids under possible primitive earth conditions. *Science*., 117, 528.

Miller, S. L. (1984) *Advan. Chem. Phys*. LV, 85.

Mullie, F., Reisse, J. (1987) *Topics Curr. Chem.* **139**, 85-117.

Neuberg, C. (1902). *Ber*. 35, 2626

Orgel, L. E. (1968) Evolution of the genetic apparatus. *J. Mol. Biol*. 38. 381-393.

Oró, J., Holzer, G. (1979) *J. Molec. Evol.* **14**, 153-160.

Owen, R. W., Wimonwatwatee, T., Spiegelhalder, B., Bartsch, H. (1996) *European J. Cancer Prevent.* **5**, 233-240.

Oyama, V. I., Berdahl, B. J. (1977) *J. Geophys. Res.* **82**, 4669-4676.

Pering, K. L., Ponnamperuma, C. (1971) *Science* **173**, 237-239.

Perrin, D. M., Garestier, T., Helene, C. (2001) Bridging the gap between proteins and nucleic acids: A metal-independent RNAseA mimic with two protein-like functionalities. *J. Am. Chem. Soc*. 123, 1556-1563.

Pfeil, E, Ruckert, H. (1960) Die Bildung von zuckern aus formaldehyde unter der einwirkung von laugen. *Ann*. 641, 121.

Pignatello, J. J. & Baehr, K. (1994) *J. Environ. Quality* **23**, 365-370.

Pizer, R., Tihal, C. (1992) Equilibria and reaction mechanism of the complexation of methylboronic acid with polyols. *Inorg. Chem*. 31. 3243-3247

Pizer, R., Tihal, C. (1996) Mechanism of boron acid/polyol complex formation. Comments on the trigonal/tetrahedral interconversion on boron. *Polyhedron*. 15. 3411-3416

Ricardo, A., Carrigan, M. A., Olcot, A. N., Benner, S. A. (2004) Borate minerals stabilize ribose. *Science*. 303, 196.

Rich, A. (1962) On the problems of evolution and biochemical information transfer. In *Horizons in Biochemistry*, Kasha, M. and Pullman, B. editors, N.Y., Academic Press, 103-126

Rieder, R., Economou, T., Wänke, H., Turkevich, A., Crisp, J., Brückner, J., Dreibus, G., McSween, H. Y. Jr. (1997) The Chemical Composition of Martian Soil and Rocks Returned by the Mobile Alpha Proton X-ray Spectrometer. *Science*. 278. 1771

Roth, A., Breaker, R. R. (1998) An amino acid as a cofactor for a catalytic polynucleotide. *Proc. Nat. Acad. Sci. USA*. 95, 6027-6031

Ruckert, H., Pfeil, E., Scharf, G. (1965). *Chem. Ber.* 98. 2558.

Safarzadeh-Amiri, A., Bolton, J. R., Cater, S. R. (1997) *Water Res.* **31,** 787-798.

Safarzadeh-Amiri, A., Bolton, J. R., Cater, S. R. (1996) *Solar Energy* **56**, 439-443.

Sandstrom, B. E., Svoboda, P., Granstrom, M., Harms-Ringdahl M., Candeias, L. P. (1997) *Free Radical Biol. Medi.* **23**, 744-753.

Santoro, S. W., Joyce G. F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. USA*. 94, 4262-4266

Santoro, S. W., Joyce G. F. (1998) Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochem.* 37, 13330-13342.

Sephton, M. A., Pillinger, C. T., Gilmour, I. (1998) *Geochim*. *Cosmochim*. *Acta* **62**, 1821-1828.

Shapiro, R. (1988) Prebiotic ribose synthesis. A critical analysis. *Origins Life Evol. Biosphere* 18, 71-85

Stage-Zimmermann, T. K., Uhlenbeck, O. C. (1998) Hammerhead ribozyme kinetics. *RNA*, 4, 875-899.

Sun, Y. F., Pignatello, J. J.(1993) *J. Agricul. Food Chem.* **41**, 308-312.

Theurich, J., Bahnemann, D. W., Vogel, R., Ehamed, F. E., Alhakimi, G., Rajab, I. (1997) *Res. Chem. Intermediates* **23,** 247-274.

Usher, D. A., McHale, A. H. (1976) Hydrolytic stability of helical RNA : A selective advantage for the natural 3',5'-bond. *Proc. Natl. Acad. Sci. USA*. 73. 1149-1153.

Van Duin, M., Peters, J.A., Kieboom, A. P. G., Van Bekkum, H. (1984) Studies on borate esters I. *Tetrahedron*. 40. 2901-2911

Van Duin, M., Peters, J.A., Kieboom, A. P. G., Van Bekkum, H. (1985) Studies on borate esters II. *Tetrahedron*. 41. 3411-3421

Walling, C. (1997) Fenton's reagent revisited. *Acc. Chem. Res.* **8**, 125-131.

Walton, J. H., Graham, D. P. (1928) *J. Am. Chem. Soc.* **50**, 1641-1648.

Watts, R. J., Jones, A. P., Chen, P. H., Kenny, A. (1997) *Water Environment Res.* **69**, 269-275.

Wieland, H., Franke, W. (1927) *Ann*. 457, 1-70.

Weiss, A. H., LaPierre, R. B., Shapira, J. (1970). Homogeneously catalyzed formaldehyde condensation to carbohydrates. *J. Catal*. 16. 332

Woese, C. R. (1967) *Origins of the Genetic Code*. Harper & Row, N. Y.

Wu, L.P., Munakata, M., Kuroda-Sowa, T., Maekawa, M., Suenaga, Y. (1996) *Inorg. Chim. Acta* **249**, 183-189.

Yamashita, M., Yamamoto, Y., Akiba, K., Nagase, S. (2000). Synthesis of a versatile tridentate anthracene ligand and its application for the synthesis of hypervalent pentacoordinate boron compounds. *Angew .Chem. Int. Edit*. **39**. 4055-4058

Yomo, T. (2003) Molecular evolution in static and dynamical landscapes. *J. Biol. Phys*. 28, 471-482.

Zakharov, I. V., Kumpan, Y. V. (1996) *Kinet. Catal.* **37**, 174-178.

Zhai, M., Shaw, D. M. (1994). Boron cosmochemistry. Part I: Boron in meteorites. *Meteoritics* 29, 607-615.

Zubay, G. (1998) Studies on the lead-catalyzed of synthesis of aldopentoses. *Orig. Life Evol. Biosphere* 28, 13-26

Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 31, 1-10

BIOGRAPHICAL SKETCH

Alonso Ricardo was born in Santiago de Cali, Colombia, in 1977. He received a B.S in chemistry in 2000 from the Universidad del Valle while working with Dr. Luz Marina Jaramillo. After completing his B.S he joined the chemistry graduate program at the University of Florida working under the supervision of Dr. Steven A. Benner to pursue a doctorate.