REVIEW

Formamide Chemistry and the Origin of Informational Polymers

by **Raffaele Saladino**[a]), **Claudia Crestini**[b]), **Fabiana Ciciriello**[c]), **Giovanna Costanzo**[d]), and **Ernesto Di Mauro***[e])

[a]) Dipartimento A. B. A. C., Università della Tuscia, Via San Camillo De Lellis, I-01100 Viterbo
[b]) Dipartimento di Scienze e Tecnologie Chimiche, Università 'Tor Vergata', I-00133 Rome
[c]) Dipartimento di Genetica e Biologia Molecolare, University 'La Sapienza', P.le Aldo Moro, 5, I-00185 Rome
[d]) Istituto di Biologia e Patologia Molecolari, CNR, P.le Aldo Moro, 5, I-00185 Rome
[e]) Fondazione 'Istituto Pasteur-Fondazione Cenci-Bolognetti', c/o Dipartimento di Genetica e Biologia Molecolare, Università 'La Sapienza', P.le Aldo Moro, 5, I-00185 Rome
(phone: +390649912880; fax: +390649912500; e-mail: ernesto.dimauro@uniroma1.it)

Formamide ($HCONH_2$) provides a chemical frame potentially affording all the monomeric components necessary for the formation of nucleic polymers. In the presence of the appropriate catalysts, and by moderate heating, formamide yields a complete set of nucleic bases, acyclonucleosides, and favors both phosphorylations and transphosphorylations. Physico-chemical conditions exist in which formamide favors the stability of the phosphoester bonds in nucleic polymers more than that of the same bonds in monomers. This property establishes 'thermodynamic niches' in which the polymeric forms are favored. The hypothesis that these specific attributes of formamide allowed the onset of prebiotic chemical equilibria capable of *Darwinian* evolution is discussed.

Contents

1. The Interest in Formamide from a Prebiotic Perspective. – Self-organization and endurance of informational polymers entail the simultaneous presence of robust

chemical frames and of favorable thermodynamic conditions. This truism led us to investigate the role of formamide in the origin of genetic information, based on the following considerations. First, in order to form a polymer, all the involved monomeric components must be available, possibly at the same time and in the appropriate amounts. The simultaneous presence of the required compounds in the same reaction milieu would speed up the process, contributing to counterbalance instability problems. Second, the stability of both monomers and polymers was correctly recognized as a major problem in origin-of-life debates [1–6]. The presumably harsh early-Earth environment(s) [7–11], the absence of the protective biological structures and mechanisms that evolved at later times, and the very fact that polymers are, in principle, more unstable than their basic monomers, justify the worry on instability.

A unitary chemical frame affording all the required components would increase the frequency of productive encounters and would decrease the instability effects associated with waiting times on the assembly line. By increasing the chances of polymerization, a robust unitary chemical frame would, thus, be greatly favored as the pre-genetic chemical root over more fragmentary processes. The arguments presented below provide evidence that formamide has the right properties for such a role.

2. On the Elemental Complexity of First Precursors: Three or Four Elements? – The composition of early prebiotic Earth and of its atmosphere continues to be a debated matter. However, it is a shared opinion that the comets and asteroidal bodies still circulating in the Solar System represent the elemental and molecular composition of the cloud from which planets originated. The molecular composition of both comets/asteroids and of the interstellar clouds [12–14] reveals that the compounds made of the four biologically more relevant and (with the exception of the inert He) more common elements H, O, C, and N are isocyanate (HNCO) and formamide ($HCONH_2$). This latter C_1 molecule was in particular detected in the gas phase of interstellar medium [14], in the long-period comet *Hale–Bopp* [15], and, tentatively, also in the solid phase on grains around the young stellar object W33A [16]. Possible formamide production under *Europa*-like conditions was observed [17]. Thus, we have focused, as a first approach, on formamide rather than on isocyanate. Exploring the properties of HNCO in a biogenic perspective will be interesting.

Simpler small molecules composed of only three different types of atoms such as hydrogen cyanide (hydrocyanic acid; HCN) are, in principle, endowed with similarly high prebiotic potential. However, to afford a large panel of appropriate precursors for biogenic macromolecules, starting compounds comprising only three different elements must undergo additional reactions. The variegated requirements, achievements, and potentialities of HCN chemistry in a prebiotic perspective have been the focus of decades of studies [18], as highlighted by the seminal synthesis of adenine reported by *Oró* [19][20] as a product of cyanide polymerization. The results obtained based on HCN chemistry were reviewed [18]. The limits of HCN as the general actor of prebiotic processes are to be found in its extreme reactivity and in the lack of an efficient synthesis of pyrimidine nucleic bases from this compound.

Note that both HCN and H_2O are the most abundant three-atom organic and two-atom inorganic compounds, respectively, in the interstellar medium [12], their reaction affording formamide ($HCONH_2$), a still quite reactive molecule. Formamide chemistry

is showing to be a cornucopia, potentially providing the complete series of reactions needed for the increase of the molecular complexity from a C_1 precursor to informational nucleic polymers. As discussed below, our starting hypothesis that a compound made of four different atoms is a more focused and, thus, more efficient starting point than those made of three elements is empirically proving to be based on solid experimental ground.

Informational polymers with genetic functions, as conceived today with an *a posteriori* logic, are made of five different atoms. Great attention was devoted to the basics of the nucleic informational polymers and to the properties of their backbones [21–25]. Possible alternatives were extensively analyzed and reviewed [25]. There are two major conclusions of more than a decade of studies on this matter. *1*) The ribose and the 2′-deoxyribose sugar moieties present in extant nucleic acids may be substituted by several other compounds including pre-organized furanose in locked nucleic acid (LNA) [26][27], pyranosyl-RNA (pRNA) and its nucleo-δ-peptide analogue [28] (and refs. cit. therein), aminopropyl nucleic acid [29], 1′,5′-anhydrohexitol in hexitol nucleic acid (HNA) [30], D- and L-cyclohexenyl nucleic acid (CeNA) [31], threose in threose nucleic acid (TNA) [32][33], and 2′-fluoro N(3′)–P(5′) phosphoramidates [34]. *2*) Major alternatives to the phosphate-linked backbone chain also exist [21], notably those bearing bis(methylene)sulfone instead of phosphate as chain linker [25]. The properties of the methylene derivatives and those of their natural counterparts, *i.e.*, 5′-methylenephosphonate, bis(methylene)phosphinate, and bis(methylene)sulfone, have been determined [25]. These non-orthodox backbones suggest that a negative charge is not essential for *Watson–Crick* duplex formation.

Another alternative polymer endowed with properties of stability, the ability to replicate, and with a coding capacity comparable with that of its natural counterpart is 'peptide nucleic acid' (PNA) [35]. PNA is a DNA mimic with a pseudopeptide backbone composed of *N*-(2-aminoethyl)glycine units to which the nucleobases are linked through methylene-carbonyl linkers. The relevant property of PNA in this context is that it is built out of only four different atoms. The possible interest of PNA as a primordial alternative was explored [36], but its non-polyelectrolytic nature leads to a self-aggregation propensity [24] and limits its evolutionary potential. The relevance of serial negative charges in preventing the single strands from folding [24] and the convenient way in which this property is provided by phosphates were pointed out [21]. Nevertheless, the elemental parsimony of PNA establishes the principle that alternatives to five-element polymers, such as RNA and DNA, are possible in principle and that, from a molecular neo-*Darwinian* evolutionary perspective, the presence of phosphate is not a mandatory requisite. Relevant progress in non-phosphate backbones was recently reported by *Bean et al.* [37] who described glyoxylate as a backbone linkage for a prebiotic ancestor of RNA (*ga*RNA). The glyoxylate group is remarkably close in structure and electrostatic charge distribution to the phosphate group [37].

The inclusion of phosphate into the extant backbones of informational polymers has, *de facto*, winning *Darwinian* properties in the present terrestrial environment. Nevertheless, the *ga*RNA and PNA examples show efficient and radical alternatives based on building blocks made out of four different types of atoms. Reasonably, these latter compounds are starting points closer to the common prebiotic chemical roots than their more-complex congeners.

Summarizing these introductory remarks:

- Formamide ($HCONH_2$) is a molecule widespread in abiotic environments, both extraterrestrial and, by assimilation, presumably also on early Earth.
- Along with isocyanate (HNCO), formamide is the simplest molecule potentially able to undergo condensation and polymerization processes leading, within a unitary chemical frame, to a complete set of relevant pre-genetic macromolecules, as detailed under the next headline.
- The more simple hydrogen cyanide (HCN) is, in principle, endowed with the same potential as formamide. However, it requires, if not quenched with H_2O, a complex series of additional reactions to afford all the appropriate precursors for nucleic informational macromolecules.

3. The Bottom-Up Prebiotic Process to Informational Polymers. – The steps leading from any putative simple-molecule precursor to an extant-type nucleic informational polymer are: *1*) condensation into nucleic bases; *2*) formation of nucleosides thereof; *3*) phosphorylation of nucleosides; *4*) chain-wise linear polymerization; and *5*) survival of the formed polymer for a period long enough to allow replication. Formamide plays a positive role in most of these steps, both as a building block and, in defined instances, as a catalytic cofactor.

4. The Syntheses of Nucleic Bases from Formamide in the Presence of Catalysts. – The nucleic bases necessary for the synthesis of early informational polymers can be obtained from simple and largely diffused chemical precursors in several experimental models (for reviews on this topic, see [18][38–41]). In recent years, formamide received particular attention due to its unique property of being a synthon for both purine and pyrimidine nucleic bases, simply by heating the compound at 110–160° in the presence of metal oxides and minerals largely diffused in the Universe [42]. Formamide has a boiling point of 210° (at atmospheric pressure), with limited azeotropic effects; at difference from other prebiotic precursors, it can be easily concentrated from dilute aqueous solutions by simple evaporation of H_2O [43].

The efficacy of formamide chemistry in the origin of informational polymers strictly depends on the metal oxides and minerals used in its condensation. These compounds act both as heterogeneous catalysts, providing high selectivity in the products distribution [44], and as benign microenvironment for the protection of newly formed products from degradation processes [45][46]. In the absence of catalysts, the condensation of formamide affords only the simple purine scaffold in very low yield (for pioneering studies on thermal formamide condensation, see [47][48]).

Tables 1–3 report the extensive panel of nucleic acid components derived from formamide in the presence of various catalysts, giving rise to all the nucleobases of extant DNA and RNA, including uracil, thymine, cytosine, adenine, and hypoxanthine (bio-isoster of guanine); some intermediates of their present-day biosynthesis, *i.e.*, 5-amino-1*H*-imidazole-4-carboxamide (AICA) and 5-(formylamino)-1*H*-imidazole-4-carboxamide (fAICA); as well as purine acyclonucleosides bearing two or three C-atoms in their side chains. For a representative example of formamide degradation, see [43].

Table 1. *Formamide-Based Synthesis of Nucleic Acid Components Catalyzed by Different Inorganic Materials and Clays.* See also *Tables 2* and *3*.

Product[a])	Silica	Alumina	Kaolin	Zeolite	$CaCO_3$	Clays[b])			
						KP-10	K-30	KSF	Al-PILC
Purine	+	+++++	+++++	+++++	+++++	+++++	+++++	+++++	+++++
Adenine	+	+	0	+	0	++	++++	+++	+
Hypoxanthine	0	0	0	0	0	+	0	+	+
N^9-Formylpurine	0	0	0	0	0	+++	++++	+	+
N^9,N^6-Diformyl-adenine[1])	0	0	0	0	0	0	0	0	0
Acyclonucleosides	0	0	0	0	0	0	0	0	0
Cytosine	+	+	0	+	0	+++	+++	+++	+++
Hydroxy-pyrimidine	+	+	+	+	0	0	0	0	0
Pyrimidin-4(3*H*)-one	0	0	0	0	0	0	0	0	0
Uracil	0	0	0	0	0	+	+	+	+
Thymine	0	0	0	0	0	0	0	0	0
5-Hydroxymethyl-uracil	0	0	0	0	0	0	0	0	0
AICA[c])	0	0	0	0	0	0	0	+++	+++
fAICA[d])	0	0	0	0	0	++	++++	+++	+
Urea	0	0	0	0	0	0	0	0	0

[a]) The data refer to yield of product (in mg) per formamide (in g): + = 0.1–5.0 mg; ++ = 5–10 mg; +++ = 10–20 mg; ++++ = 20–40 mg; +++++ = > 40 mg. [b]) Activated Montmorillonites. [c]) 5-Aminoimidazole-4-carboxamide. [d]) 5-Formamidoimidazole-4-carboxamide.

The synthesis of purine acyclonucleosides is of particular prebiotic relevance because of the known difficulty to build up the glycosidic bond between nucleobases and sugars synthesized under primitive conditions [49–51]. It is worth to note that in these syntheses formamide acts as a multifunctional prebiotic precursor, being a primary reagent able to generate *in situ*, under simple environmental conditions, a complex mixture of reactive chemicals. As shown in *Scheme 1*, these include HCN, ammonia (NH_3), H_2O, formic acid (HCOOH), carbon monoxide (CO), and formaldehyde (H_2CO). As an example, formamide starts decomposing at 190° to afford HCN, NH_3, CO, and H_2O, all involved in the synthesis of a large panel of nucleic acid components [46]. This reaction is known to be catalyzed by different metal oxides [52]. Moreover, even though formamide is hydrolyzed slowly to ammonium formate ($K_w = 1.1 \times 10^{-10}\ s^{-1}$ at 25°), with a half life $t_{1/2}$ of nearly 200 years at ambient temperature, the reaction can be efficiently accelerated under acidic-clay conditions [53].

---

1) The trivial name 'N^9,N^6-diformyladenine', although commonly used, is not consistent. There is one CHO group at N(9), but no such group at position 6. Position 6 corresponds to atom C(6) of the purine nucleus, the second CHO group being at the N-atom (!) bonded to C(6). For convenience, the trivial name is, nevertheless, retained. The same applies to related compounds. Systematic names are given in *Sect. 6*.

Table 2. *Formamide-Based Synthesis of Nucleic Acid Components Catalyzed by TiO_2 or Cosmic-Dust Analogues*

Product[a])	TiO_2	Cosmic-dust analogue				
		$MgFeSiO_4$	$Mg_{1.5}Fe_{0.5}SiO_4$	$Mg_{0.5}Fe_{1.5}SiO_4$	Fe_2SiO_4	Mg_2SiO_4
Purine	+++	+	0	0	+	+
Adenine	++	0	0	0	0	0
Hypoxanthine	0	0	0	0	0	0
N^9-Formylpurine	+++	0	0	0	0	0
N^9,N^6-Diformyladenine	+	0	0	0	0	0
Acyclonucleosides	++	0	0	0	0	0
Cytosine	+	+++	++++	+++	+++++	0
Hydroxypyrimidine	0	0	0	0	0	0
Pyrimidin-4(3*H*)-one	0	++	++	+++++	+++++	0
Uracil	0	+	0	+	+	0
Thymine	+	0	0	0	0	0
5-Hydroxymethyluracil	+	0	0	0	0	0
AICA[b])	0	0	0	0	0	0
fAICA[c])	0	0	0	0	0	0
Urea	0	+	+	0	+	+

[a]) The data refer to yield of product (in mg) per formamide (in g): + = 0.1–5 mg; ++ = 5–10 mg; +++ = 10–20 mg; ++++ = 20–40 mg; +++++ = > 40 mg. [b]) 5-Aminoimidazole-4-carboxamide. [c]) 5-Formamidoimidazole-4-carboxamide.

Scheme 1. *Basic Formamide Chemistry*

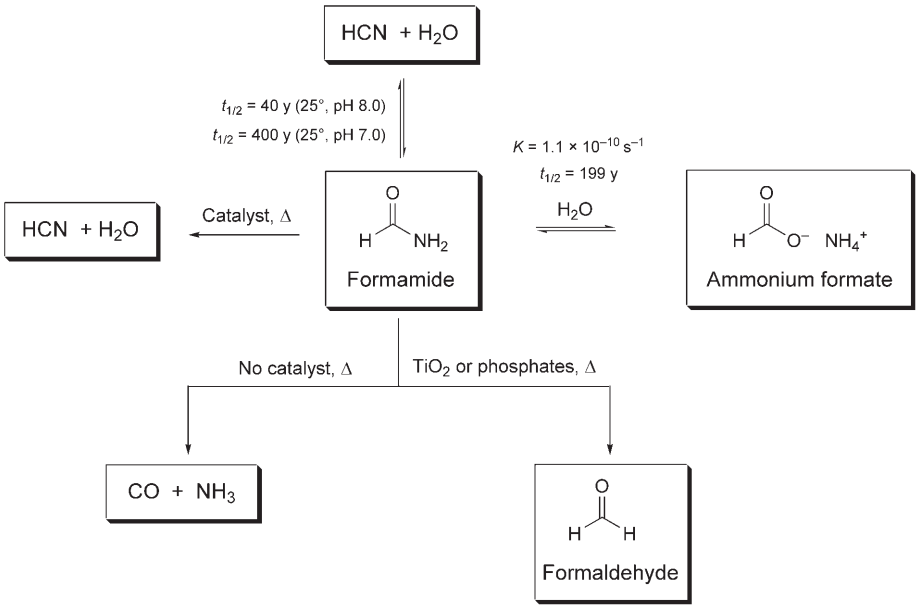


Table 3. *Formamide-Based Synthesis of Nucleic Acid Components Catalyzed by Various Minerals*. Data (modified) from [54].

Product[a])	Turquoise	Childrenite	Ludlamite	Vivianite	Vauxite	Lazulite	Hureaulite
Purine	++	0	+	+	+	+	+
Adenine	0	0	0	0	0	0	0
Hypoxanthine	0	0	0	0	0	0	0
Uracil	0	0	0	0	0	+	+
Cytosine	++	0	0	0	0	+	++
Dihydroxyuracil	0	0	0	0	0	0	0
Urea	0	+	0	+	0	0	0
Parabanic acid	0	0	0	0	0	+	0
N-Formylglycine	+	+++++	++++++	++++	++++	+++	0
Carbodiimide	0	0	0	0	0	0	+

Product[a])	Augelite	Wavellite	Libethenite	Pyromorphite	Na_3PO_4	$Na_4P_2O_7$	$Na_5P_3O_9$
Purine	+	+	0	+	+++	++	+
Adenine	0	0	0	0	0	+	0
Hypoxanthine	0	0	0	0	0	+	0
Uracil	0	0	+	0	+	+	0
Cytosine	+	+	+	+	++++	+++	+
Dihydroxyuracil	0	0	0	0	0	+	0
Urea	0	0	0	0	+	0	0
Parabanic acid	0	0	0	0	0	0	0
N-Formylglycine	+++	+	+	0	++	+	++
Carbodiimide	0	0	+	0	+	+	+

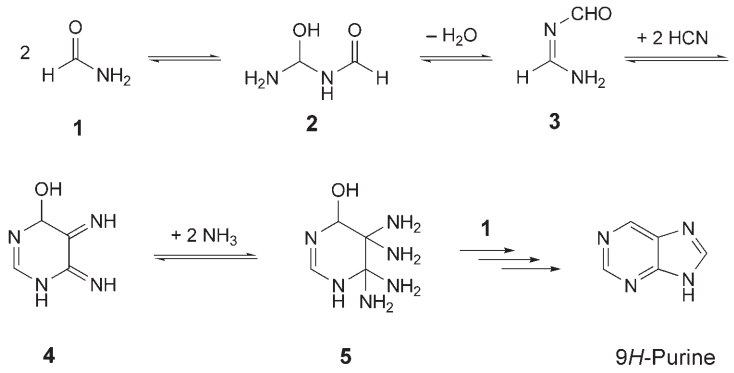
[a]) The data refer to yield of product (in mg) per formamide (in g): + = 0.1–5 mg; ++ = 5–10 mg; +++ = 10–20 mg; ++++ = 20–40 mg; +++++ = > 40 mg.

The multifunctional character of formamide as prebiotic precursor can be highlighted by analyzing the mechanisms suggested for its condensation to purine and pyrimidine nucleic bases. The mechanism for the formation of purine proper by thermal condensation of neat formamide (**1**) has been studied by ^{13}C,^{15}N-heteronuclear spin–spin-coupling NMR experiments (*Scheme 2*), both on products obtained from doubly enriched formamide and, alternatively, from a mixture of doubly enriched formamide and HCN [55].

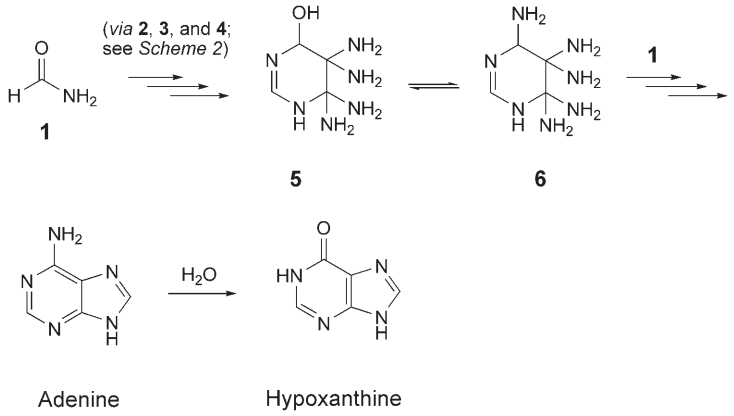
As reported in *Scheme 2*, the formamide dimer **2** is a key intermediate in the first step of the condensation process. This compound can lose H_2O to afford '*N*-formylformamidine' (= *N*-(iminomethyl)formamide; **3**), which, in turn, reacts with 2 equiv. of HCN to provide 1,4,5,6-tetrahydro-5,6-diiminopyrimidin-4-ol (**4**) and, successively, the corresponding ammonia adduct **5** [56]. Note that both HCN and NH_3 are generated *in situ* by thermal decomposition and hydrolysis of formamide (**1**), respectively. Successive formamide-addition and -elimination reactions then afford purine.

A similar mechanism, again requiring the formation of the dimer **2** as key intermediate (*Scheme 3*), was suggested for the synthesis of adenine. In this latter case, nucleophilic substitution of the OH group of **5** by NH_3 affords intermediate **6**, which further polymerizes to adenine [57]. In this reaction pathway, hypoxanthine is probably formed by hydrolytic deamination of adenine. Thus, the pyrimidine ring is the first

Scheme 2



Scheme 3

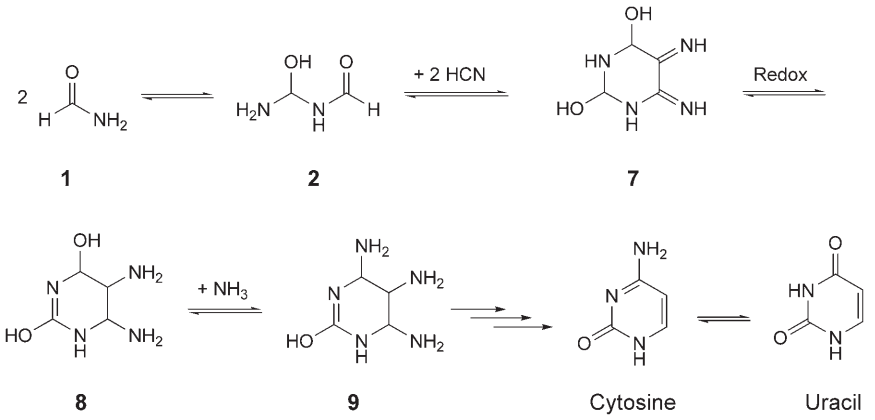


heterocyclic ring to be closed during the synthesis of purine nucleobases from formamide. For this reason, these syntheses can be dubbed in general terms as the 'routes of the pyrimidine ring'.

The formamide dimer **2** and a formamide trimer (compound **11** in *Scheme 8*) were recently detected by mass-spectrometric (MS) analysis under ultra-high-vacuum conditions upon UV irradiation of formamide on the surface of a TiO_2 (001) single crystal [58].

The multifunctional character of formamide is also a relevant prebiotic property in the formation of pyrimidine nucleic bases. The reaction pathway for the synthesis of cytosine and uracil is reported in *Scheme 4*. In agreement with previously described mechanisms for the synthesis of purines, the formamide dimer **2** reacts with 2 equiv. of HCN to hexahydro-5,6-diiminopyrimidine-2,4-diol (**7**). This intermediate may undergo an internal redox process (probably a disproportionation reaction) to afford **8**, which, in turn, after ammonia substitution, yields compound **9**. Successive elimination and

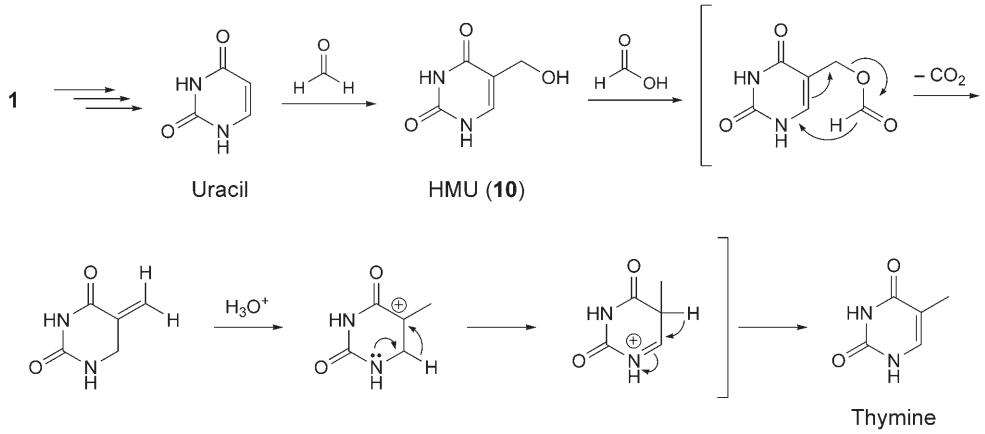
Scheme 4



rearrangement processes then afford cytosine. Once formed, cytosine can be hydrolyzed to uracil in a yield that depends on the experimental conditions (for a representative example of this transformation, see [59]).

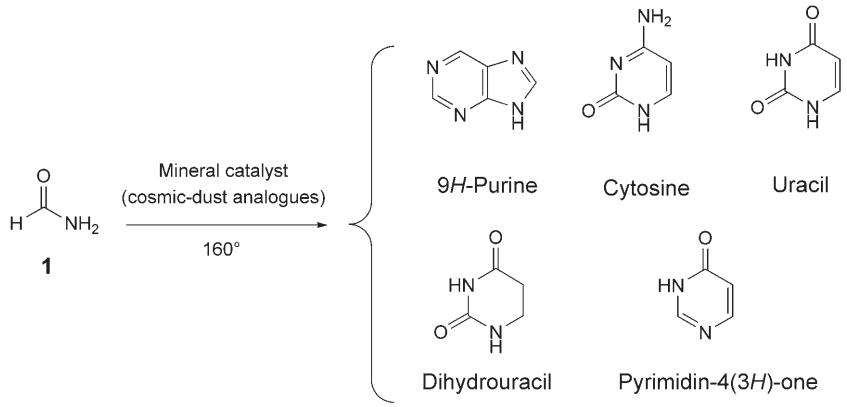
A more complex reaction pathway is operative in the synthesis of thymine, which requires the presence of TiO_2 as catalyst (*Scheme 5*) [60]. In this latter case, formamide first condenses to uracil, probably by the mechanism reported in *Scheme 4*. The newly formed uracil then reacts with formaldehyde (H_2CO), generated *in situ* by TiO_2-catalyzed degradation of formamide [61], by addition to the C(5)=C(6) bond to afford 5-hydroxymethyluracil (HMU; **10**), which was detected in the reaction mixture by GC/MS analysis [62]. Finally, thymine was obtained by the known hydride-shift mechanism involving formic acid as a product of formamide hydrolysis [63].

Scheme 5



Metals with different redox potentials tune the synthesis of pyrimidine nucleic bases by modifying the efficacy of the internal redox step, as shown in the case of cosmic-dust analogues (CDAs) of common terrestrial minerals such as fayalite, olivine, and fosterite. The CDAs can be selectively synthesized by laser-ablation techniques (Nd-YAG laser, with 10^8 W cm^{-2}). Oxide pellets of MgO, FeO_2, and SiO_2 are used as laser targets and vaporized in a 10-mbar O_2 atmosphere before irradiation. When formamide is treated at 160° in the presence of catalytic amounts of CDAs, a large panel of pyrimidine derivatives, including cytosine, uracil, pyrimidin-4(3*H*)-one, and 5,6-dihydrouracil, was obtained, purine being the only purine derivative to be recovered in the reaction mixture (*Scheme 6*) [64].

Scheme 6

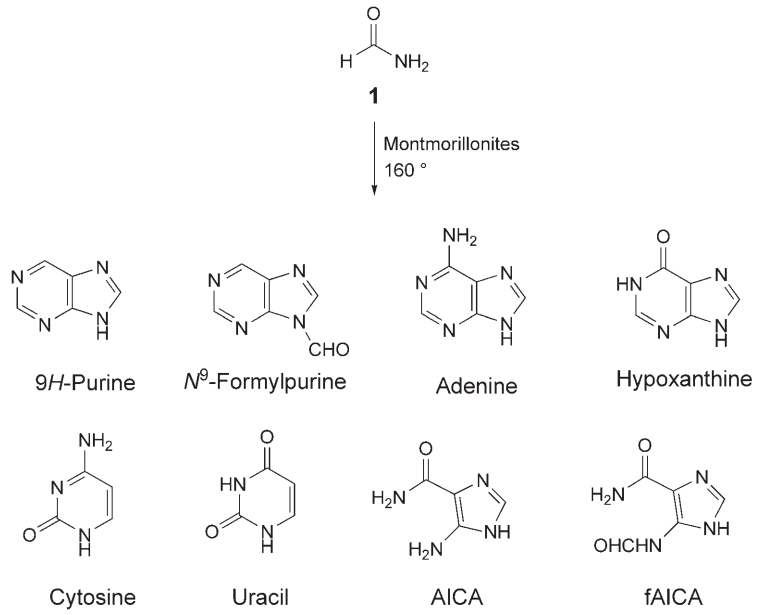


In these syntheses, iron (Fe)-based CDAs (such as fayalite and olivine) are more-efficient catalysts for the formation of pyrimidine nucleic bases than Mg-containing CDAs (*e.g.*, fosterite), the yield of each product increasing with the increase of the Fe content in the elemental composition of the mineral. Thus, Fe-containing minerals appear to be the best microenvironment for the synthesis of pyrimidine nucleic bases in space. This result is further confirmed by the absence of any correlation between the yield of purine and the elemental composition of CDAs, confirming the lack of an internal redox process in the prebiotic synthesis of purine.

5. Chemiomimesis in Formamide Chemistry. – The term *'chemiomimesis'* generally refers to a chemical reaction pathway that can be used as a template for the enzymatic processes that will appear later in evolution to yield the same final products.

The chemical correlations between *a*) the condensation of formamide and nucleic bases and *b*) present-day cellular biosynthetic pathways for nucleic acid components are remarkable. According to *Eschenmoser*'s concept, these correlations suggest the possibility that certain biosynthetic pathways can be considered as chemiomimetic of early prebiotic chemistry [65]. As an example, when formamide is heated in the presence of naturally occurring clays such as montmorillonites, two imidazole derivatives, 5-amino-1*H*-imidazole-4-carboxamide (AICA) and 5-(formylamino)-1*H*-

Scheme 7

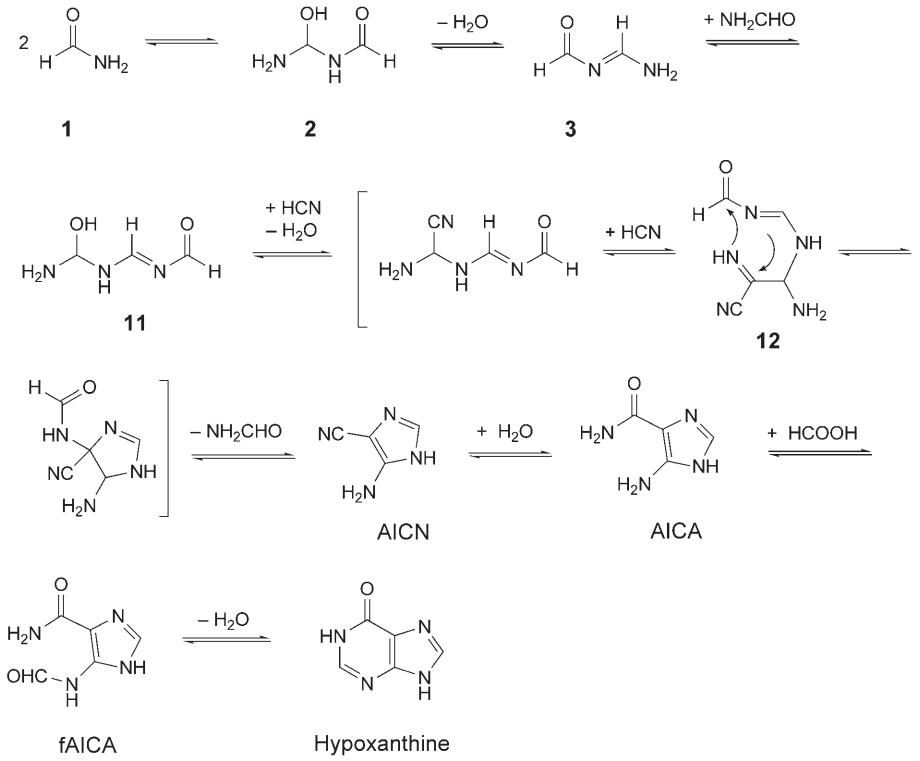


imidazole-4-carboxamide (fAICA), are obtained in high yield, along with purine, 'N^9-formylpurine' (=9*H*-purine-9-carbaldehyde), adenine, hypoxanthine, cytosine, and uracil (*Table 2* and *Scheme 7*) [66]. Given that AICA and fAICA are not expected intermediates in the formation of purine nucleic bases through the classical 'routes of the pyrimidine ring', another mechanism of condensation of formamide is reasonably operative in the presence of montmorillonites. However, AICA has been proposed as a key intermediate (in addition to 5-amino-1*H*-imidazole-4-carbonitrile (AICN)) in the prebiotic synthesis of purines by polymerization of HCN [67], suggesting that its conversion to fAICA is a fundamental step in the synthesis of hypoxanthine [68].

In agreement with these data, a reasonable reaction pathway for the condensation between formamide and purines in the presence of montmorillonites, including AICA and fAICA as key intermediates, called 'route of the imidazole ring', is reported in *Scheme 8*. In this latter case, the imidazole ring of the purine scaffold is formed before the pyrimidine nucleus. The reaction starts with the initial condensation of formamide (**1**) to the formamide dimer **2** and trimer **11** (*via* the formamidine **3**). The intermediate **12** (not isolated in this case) is possibly formed by elimination of H_2O from **11**, followed by addition of 2 equiv. of HCN to the newly formed imine and nitrile functions. Successive cyclization, formyl rearrangement, and elimination of formamide then affords AICN. Irrespective of the nature of the intermediates involved in these latter steps, the transformation of AICN to hypoxanthine through the successive synthesis of AICA and fAICA is a well-documented process [67][68].

Notably, AICA ribonucleotide-5′-monophosphate (AICA-monophosphate) and fAICA ribonucleotide-5-monophosphate (fAICA-monophosphate) are key intermedi-

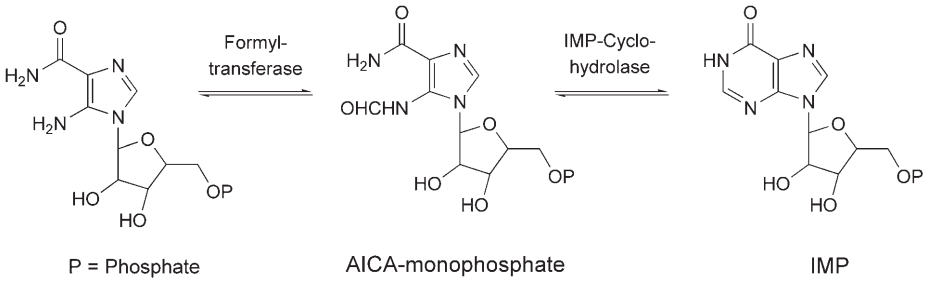
Scheme 8



ates in the last steps of the extant biosynthesis of inosine-5′-monophosphate (IMP), the main route to purine nucleotides in the cell. Briefly, AICA-monophosphate is selectively formylated by a formyltransferase to fAICA-monophosphate, which, in turn, is cyclized by IMP cyclohydrolase to IMP (*Scheme 9*).

The high similarity of the last steps of the hypoxanthine synthesis from formamide with its present biological synthesis suggests the possibility that the enzymatic pathway

Scheme 9



can be, at least in part, a chemiomimesis of the corresponding primitive chemical template. The possibility that early chemical events played the role of templates for the development of more-complex (but also more-efficient and -selective) enzymatic pathways is a fascinating concept to be further evaluated in the study of the molecular evolution of informational polymers.

6. The Formose Reaction and the Formation of Acyclonucleosides: A One-Pot Formamide-Based Road to Nucleosides. – It would seem natural that the prebiotic formation of nucleosides proceeded by linking preformed ribose or 2′-deoxyribose and nucleic bases. As reviewed by *Joyce* [69], and re-examined by *Zubay* and *Mui* [70], this is not the case. The question of how ribose and purines were linked together has only partially been resolved for inosine [71], but not for the other nucleosides. The synthesis of ribose itself is crammed with difficulties [72][73]. Even though improved conditions for the synthesis of potentially natural alternatives derived from sterically less-bulky sugars were reported [74] (and refs. cit. therein), synthetic pathways different from the formaldehyde-based formose condensation were seldom considered to be plausible.

Starting from the consideration that direct attachment of cytosine to ribose does not occur in H_2O, *Sutherland* and co-workers [75] reported the stepwise assembly of the base on arabinose-3-phosphate affording cytidine nucleoside. These reactions readily occur in aqueous solutions at near-neutral pH. Their plausible contribution to the prebiotic formation of ribonucleotides was suggested [75].

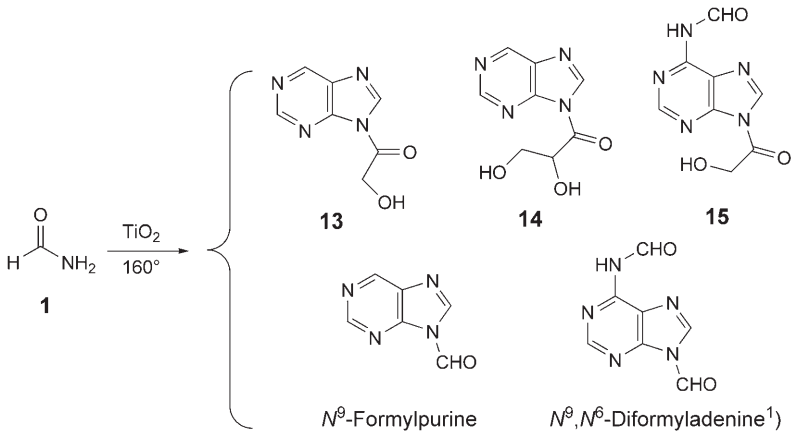
Concerns on this matter were previously reported, related to the chemical potential of the different components [71]. Due to their greater stability, the nitrogen bases are likely to have been made first, while ribose must be protected from decomposition immediately after its synthesis. Phosphorylation was suggested as a possible stabilizing factor [71]. In agreement, we observed that the β-glycosidic bond in adenosine is stabilized tenfold by the binding of a phosphate group in 5′-position [76]. The problem set by the difficulty of connecting a sugar, whether phosphorylated or not, to a base has not been solved yet.

While it was suggested that alternative sugars may have preceded ribose as an ancestor of RNA [72][74], it was also proposed [75] that RNA may have evolved from acyclonucleoside-containing polymers [77–79]. This suggestion aimed in part at alleviating the problems associated with the synthesis of RNA from a racemic mixture of ribonucleotides [75].

The recent proposal of glyoxylate as an ancestor of phosphate bonds [37] is compatible with a primitive acyclonucleoside-based polymer scenario, requiring at least two OH groups for the formation of acetal linkages. This chemical scheme finds experimental support in the products observed in the TiO_2-catalyzed condensation of formamide (*Table 2*).

Noteworthy, purine acyclonucleosides bearing a C_2 or C_3 sugar side chain, *i.e.*, compounds **13**–**15**, can be directly synthesized during the condensation of formamide in the presence of TiO_2 [60], in addition to 'N^9-formylpurine' (= 9*H*-purine-9-carbaldehyde) and 'N^9,N^6-diformyladenine' (= *N*-(9-formyl-9*H*-purin-6-yl)formamide)[1]) derivatives (*Scheme 10*).

Scheme 10



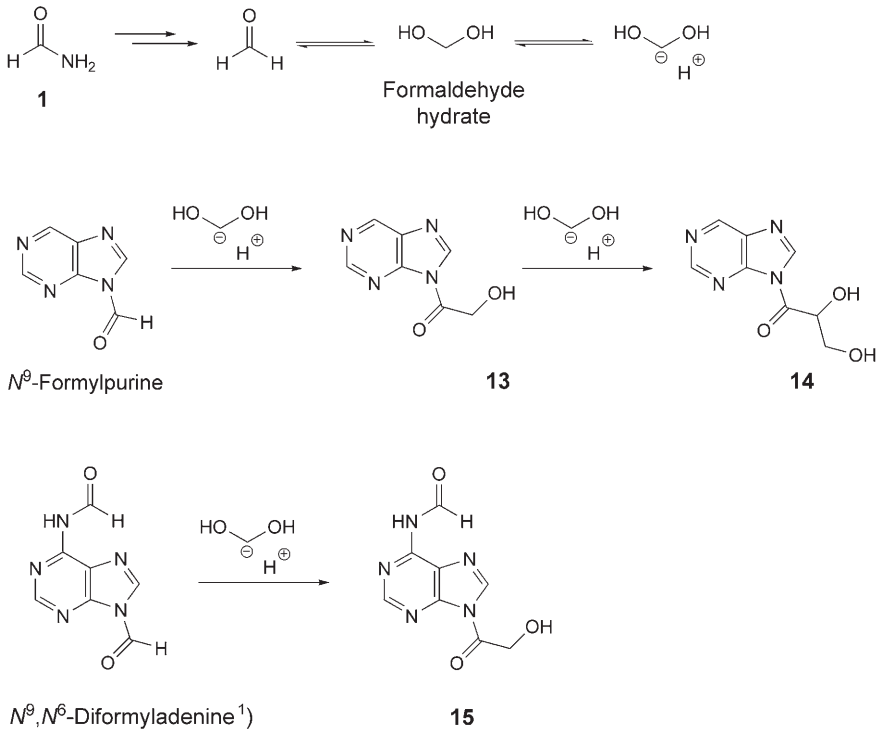
N-Formyl purines play a key role in the formation of the acyclonucleosides **13**–**15** because of their preformed glycosidic bond masked in the reactive endocyclic formyl (CHO) moiety. As reported in *Scheme 11*, formaldehyde, formed *in situ* by TiO_2-catalyzed degradation of formamide, can add to the electrophilic CHO moiety by a 'formose-like' condensation [80][81], increasing step by step the complexity of the alicyclic sugar side chain. Probably, the role of TiO_2 is both to make easier the formation of the formaldehyde-hydrate anion, and to activate the CHO moiety for nucleophilic attack. Acyclonucleosides are bioactive compounds able to pair with natural nucleic bases by *Watson–Crick* interactions [82]. In the context of prebiotic chemistry, glycerol-derived acyclonucleosides, which are characterized by a vicinal diol moiety (as in **14**), have received particular attention, mainly because of the possibility to polymerize after activation as phosphates [83].

In conclusion, side chains with up to three C-atoms were observed [60], supporting the possibility of a glyoxylate-like polymerization mechanism. Alternatively, more-favorable and as yet unexplored experimental conditions (*i.e.*, co-presence of higher concentrations of formaldehyde in a formamide-rich condensation milieu) could afford longer side chains and provide the conditions for their cyclization. In this case, threose, ribose, or deoxyribose moieties directly grown onto the nucleic bases would provide the much sought solution to nucleoside formation.

The one-pot prebiotic synthesis of purine acyclonucleosides from formamide, based on the construction of the sugar moiety directly on the nucleic base, opens novel scenarios for the origin of the first informational polymers on primitive Earth. Formamide chemistry does not require independent synthetic pathways to sugars and to nucleic bases, nor requires the stringent conditions needed for the formation of the glycosidic bond by condensation between nucleic bases and sugars.

7. Phosphorylation and Transphosphorylation Processes in Formamide Medium. – Pioneering studies of the phosphorylation of nucleosides by inorganic phosphates under prebiotic conditions were reported by *Lohrmann* and *Orgel* [84][85]. However,

Scheme 11



phosphorylation reactions in aqueous solution are usually inefficient because of the competition of H_2O for the activated phosphate intermediate [18][85]. Studies were, thus, performed in the solid state [86]. By heating dry films containing either uridine or cytidine as well as inorganic phosphate, ammonium salts, bicarbonate, and urea, 2′,3′-cyclic phosphate derivatives were obtained, along with nucleoside 5′-monophosphates (5′-NMPs). However, as pointed out [18][87], these studies did not provide an efficient procedure for the synthesis of activated nucleosides. Modifications of this approach aiming at improving the yields of 5′-NMP derivatives were reported [88], along with stability analyses showing that there is a 3:1 preference for removal of phosphates attached to the 3′-OH group over the 5′-OH group. These more-recent studies suggest an approach to the synthesis of NMPs involving recycling, thus allowing a more efficient utilization of precursors.

Following this suggestion, and moving from the seminal observations by *Schoffstall* and co-workers [89–92], we have studied the effect of formamide on the phosphorylation of nucleosides by using as phosphate donors both inorganic phosphates or NMPs (manuscript in preparation). As originally indicated by *Schoffstall* [89], formamide is an efficient promoter of the reaction. We observed that phosphorylated nucleosides may amount, at the tested temperature of 90°, for up to *ca.* 20% of the nucleoside input. Phosphorylation occured in pure formamide on adenosine or cytidine

(the two analyzed nucleosides), with similar efficiency, regardless of the source of phosphate, be it soluble KH_2PO_4 or an MNP. In this latter case, transphosphorylation showed an initial lag, corresponding to the time needed to release the phosphate moiety from the donor-phosphorylated nucleoside. The acceptor nucleoside was alternately phosphorylated in 5′-, 3′-, 2′-, or 2′,3′-cyclic positions; the 3′,5′-cyclic form was not detected due to limitation in assay resolution. The form appearing first, and amounting to higher levels, was adenosine 5′-monophosphate (5′-AMP), followed by the 3′-, the 2′-, and the 2′,3′-cyclic forms. Due to the different half-lives of the phosphorylated nucleosides (2′,3′-cyclic > 5′ > 2′ > 3′) and due to the efficiently occurring internucleotide transphosphorylation, the pool of the phosphorylated nucleosides progressively enriched with the 2′,3′-cyclic form.

Remarkably, formamide allows the reaction in the liquid state, circumventing the limits set by the reaction product H_2O, and ensuing unfavorable thermodynamics [93], which provides obvious advantages in terms of yield and kinetics. An alternative system was reported [94], describing the phosphorylation of adenosine with trimetaphosphate ($P_3O_9^{3-}$) under simulated prebiotic conditions in a cycle of wetting and drying reactions.

Other different pathways of potential interest in the production of phosphorylated nucleic precursors were reported: *1*) diiminosuccinonitrile (DISN), formed by the oxidation of diaminomaleonitrile (DAMN), effects the cyclization of 3′-AMP to adenosine 2′,3′-cyclic phosphate in high yield. However, the DISN-mediated phosphorylation of uridine to uridine monophosphate does not occur efficiently in H_2O [95]. *2*) Condensation of glycolaldhehyde phosphate in the presence of formaldehyde under alkaline conditions affords pentose and hexose di- and triphosphates [96]. *3*) The products derived from phosphono-acetaldehyde in the presence of formaldehyde were characterized [97]. Among these, a five-carbon dialdehyde – when converted to a nucleotide analogue by condensation with a purine or pyrimidine, and polymerized to form pyrophosphonic acid linkages – could form duplex structures similar to those reported for pyranose-based RNA [98].

Taken together, these observations show that the phosphorylation of preformed nucleosides or of the sugar moieties to become part of the nucleosides is not a bottleneck on the route from simple precursors to informational polymers. The question remains to be solved of the phosphate source that was actually used. The promptness of the phosphorylation reactions occurring with soluble phosphates tells that the problem is now that of a better knowledge of early-Earth geochemistry.

8. Polymerization. – Non-enzymatic autocatalytic self-replication systems based on template-directed synthesis of oligonucleotides were reported [99–109]. These studies established the principle. However, since the formation of a phosphodiester bond is thermodynamically uphill, protein-free template-directed syntheses of phosphodiester-linked oligonucleotides required the use of chemically activated nucleotides. Pioneering instances were reports, describing how 1-ethyl-3-[3-(dimethylamino)propyl]carbodiimide hydrochloride, *i.e.*, $EtN{=}C{=}N(CH_2)_3NMe_2 \cdot HCl$, brings about formation of oligoadenylic acids from adenylic acid on a polyuridylic template [104], the template-directed syntheses with adenosine-5′-phosphorimidazolide [105], and with 5′-phospho-2-methylimidazolide (2-MeImpN) nucleotides [106]. *Ferris* and co-workers [107][108] have importantly shown that drying the chemically activated nucleoside 5′-phosphor-

imidazolide-adenosine (ImpA) on the surface of a montmorillonite clay leads to the elongation of a pre-adsorbed decanucleotide by up to 30 additional nucleotides, whereas the same decanucleotide is not elongated by ImpA in solution, nor when dried in the absence of the clay [107][108].

Although activated mononucleotides spontaneously form oligonucleotides on existing nucleic acid templates, or when dried on certain mineral surfaces [107–110], chemically activated nucleotides are of questionable relevance to the origin of life, as their synthesis by a plausible prebiotic chemistry was not corroborated [18][111]. The limits of template-directed synthesis with nucleoside-5′-phosphoro-(2-methyl)-imidazolides has been pointed out [111], leading to the conclusion that an efficient catalyst must have been necessarily involved in the origin of polynucleotide replication.

9. 2′,3′-Cyclic Phosphate Ribonucleotides as Potential Source of Activated Precursors for Spontaneous Polymerization. – A plausible solution may be offered by the 2′,3′-cyclic phosphate ribonucleotide system. This system is sufficiently simple and spontaneous to be worth considering with a certain detail in the prebiotic perspective.

In 1971, *Orgel* and co-workers found that dinucleoside diphosphates can be formed from adenosine 2′,3′-cyclic phosphate [112]. The reaction is catalyzed by a number of polybasic amine and glycine derivatives, which were considered as simple models for prebiotic catalysts. The same compounds appreciably stabilize the triple helix formed by the adenosine 2′,3′-cyclic phosphate with poly(U). The authors remarked that the use of cyclic phosphates alone does not provide the complete solution to the problem of extensive polymerization, since the equilibrium mixture would not contain many highly polymerized molecules. The *Gibbs* free-energy problem, critically evaluated by *van Holde* [93], and the intrinsic instability of nucleic polymers set strict limits to the possibility of formation and endurance of long polymers in aqueous environments. Specific catalysts and a favorable physico-chemical setting (*i.e.*, reactions in organic non-aqueous solvents and conditions favoring the stability of the bonds in the polymer relative to those of the same bonds in the monomers; see below) were likely to be needed in prebiotic environments to provide a way out from otherwise prohibitive thermodynamic conditions.

The fact that *1*) the hydrolysis of preformed oligonucleotides firstly yields 2′,3′-cyclic extremities (which are successively hydrolyzed, affording 2′- and 3′- nucleoside phosphates [113]), and *2*) the facility of formation of 2′,3′-cyclic nucleosides in any of the several protein-free phosphorylation systems (as reported under the previous headline) both point to the potential function of 2′,3′-cyclic forms as precursors for polymerization.

Following the seminal observation by *Orgel* and co-workers in 1971, it was shown that 3′,5′-linked hexaadenylic acid with a 2′,3′-cyclic phosphate terminus couples on a polyuridylic acid template in the presence of ethylenediamine (=ethane-1,2-diamine) to form the corresponding dodeca- and octadecamers [114]. The bond created was largely that of the 2′,5′-isomers, but *ca.* 5% of 3′,5′-bonds were also formed. The same authors observed [115] that, upon annealing with a 3′,5′-linked complementary poly(U) strand, the stability of the 2′,5′-bond becomes *ca.* 900-fold lower than that of the 3′,5′-

bond. This helical-conformation-induced selective instability rapidly leads to the majority presence of the 'natural' 3′,5′- phosphoester bonds.

The large negative standard enthalpy of hydrolysis [116] and high reactivity [117] of 2′,3′-cyclic monophosphate nucleosides were pointed out. The fact that 2′,3′-cyclic phosphate extremities were actually shown to join nucleic polymers [116], and that double-strand formation with the complementary sequence selectively causes the steady state presence of the 'natural' 3′,5′-bonding [117], as well as their prebiotically plausible formation from nucleoside, phosphates, and formamide, all this points to a possibly important role of 2′,3′-cyclic nucleotides in initial-polymerization events. Detailed *ad hoc* studies are still lacking, though. It was noted that 2′,3′-cyclic phosphate nucleotides are unlikely to oligomerize efficiently, because the equilibrium constant for dimer formation is only of the order of 1.0 l mol^{-1} [118][119]. Identification of the appropriate catalyst and, possibly, polymerization in an organic solvent as formamide might overcome this difficulty.

10. Stability of Informational Polymers in Formamide Medium. – 10.1. *General Considerations.* Stability is a major concern in understanding the origin of informational polymers. The degradation of precursor nucleic bases is relatively rapid [1–6], and de-polymerization of nucleic acids in H_2O is a common observation. Initial attention to this problem in the 1960's and early 1970's [120] established interesting specificities: *1*) the rate of cleavage of β-glycosidic bonds of free deoxynucleosides [121][122] is 10–50 times higher relative to that in single-stranded DNA [123]; *2*) hydrolysis of glycosidic bonds in deoxynucleosides is higher than in deoxynucleotides, which, in turn, is higher than in DNA [124–126]; *3*) the depurination is fourfold in single- *vs.* double-stranded DNA (rate constant for SS DNA: 4×10^{-9} s^{-1} at pH 7.4 and 70°) [127].

At apurinic sites in DNA, the deoxyribose occurs both in the free-aldehyde and the furanose form, which leads to the cleavage of the phosphodiester bonds both at 5′- and at 3′-positions [128] (and refs. cit. therein). However, the study of relative stabilities has not been a frequently visited topic, and a comprehensive and comparative analysis of the stabilities of the β-glycosidic bonds, and of the 3′- and 5′-phosphoester bonds in monomers *vs.* polymers, in both the ribosyl and deoxyribosyl polymers, have long been lacking. We have made such an analysis for the deoxyribose [76] and the ribose [129] systems, which revealed that extensive differences exist depending on the molecular context and the physico-chemical environment. The influence of the mineral catalysts described above on bond stabilities was also determined [44][60][64][66]. The overall indication provided by these comparative measurements is that phosphoester bonds in monomers are not necessarily more stable than in polymers, and that conditions exist in which the opposite is true. Under these conditions, *the polymeric form is favored over the monomeric one, potentially leading to accumulation and survival of macromolecular information*. In the following sections, this will be outlined in more detail.

10.2. *DNA*. The stabilities of the phosphoester bonds as a function of the molecular context were compared as follows: *1*) the stability of the 3′- and the 5′-phosphoester bonds in a model deoxyribo-oligonucleotide was defined as a function of temperature and of H_2O/formamide ratio [76]. *2*) The stability of the 3′- and the 5′-phosphoester

bonds in monophosphate deoxynucleosides in a corresponding set of conditions was also determined [76]. The comparison of these two sets of data provided information relevant to the bias of the stability of monomer *vs.* polymer.

10.2.1. *Monomers.* Numerous effects exerted by the H_2O/formamide ratio and temperature (90°) on the stability of the β-glycosidic bond, and on those of phosphate bonds in the nucleoside and in its 3′- and 5′-phosphate forms were found. Briefly, for a temperature of 90°, the following was observed [76] for β-glycosidic bonds. *1*) The bond stability in H_2O is intermediate ($t_{1/2} \approx 1 \times 10^4$ min) for the nucleoside, with a marked protection by H_2O/formamide (*w*/*w*) ratios higher than 33% in formamide. *2*) Enhanced stability by the presence of a phosphate group. Thus, the presence of *formamide in solution and a bound phosphate group both stabilize the nucleoside structure.*

The following observations were made for phosphoester bonds [76]. *1*) There is a marked stability difference for the 5′- *vs.* the 3′-phosphoester bonds in H_2O, with a longer half-live ($t_{1/2} > 2.5 \times 10^4$ min) for the 5′-phosphoester bond, which is one order of magnitude above that of the 3′-form ($t_{1/2} \approx 2.5 \times 10^3$ min). *2*) Both the 5′- and the 3′-phosphoester bonds are destabilized by the presence of formamide.

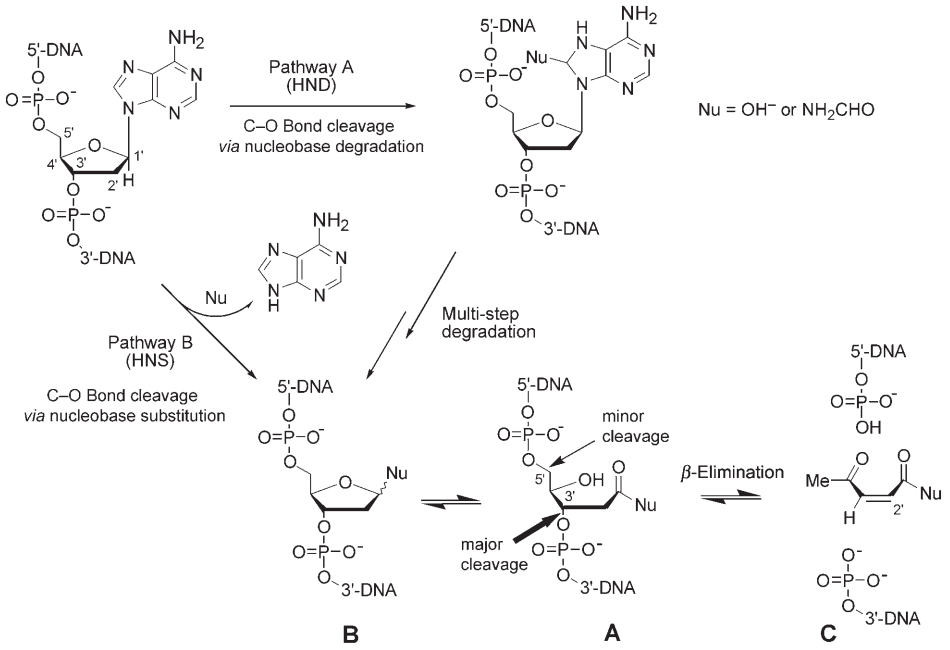
10.2.2. *Polymers.* The degradation of the same bonds, but now in polymers, was analyzed in both H_2O and formamide. In either aqueous solution or in the presence of formamide (a weak base), the DNA chain undergoes degradation according to one of the two mechanisms shown in *Scheme 12*. At temperatures of *ca.* 100° or above, and in the absence of H_2O, formamide reacts with both purine and pyrimidine nucleobases by nucleophilic addition. This reaction pathway (called 'hydrolysis following nucleobase degradation'; HND) was described in detail [76]. For purines, degradation occurs by nucleophilic attack at C(8), leading to degradative ring opening of the imidazole ring [128]. For pyrimidines, degradation occurs by nucleophilic attack at C(6) and/or at C(4) [130]. Following degradation and removal of the heterocyclic purine or pyrimidine bases, the two reactive H-atoms, H−C(2′) and H−C(4′), of the sugar moiety (**A**) are available for β-elimination, leading to 3′- and 5′-phosphoester bond cleavage, respectively. The strength of the bond H−C(2′) is higher than that of H−C(4′) [131] (and refs. cit. therein), so that, in the presence of a weak base such as formamide, the cleavage of the 3′-phosphoester bond occurs preferentially relative to that of the 5′-bond [130][131].

The hydrolytic degradation of the oligonucleotide in H_2O starts with the cleavage of the β-glycosidic bond, leading to the removal of a non-degraded base, followed by phosphodiester bond cleavage. This second pathway, termed 'hydrolysis following nucleobase substitution' (HNS; *Scheme 12*) has been described in detail [76]. Intermediate conditions exist in which the two degradation pathways are not mutually exclusive.

10.2.3. *Summary.* Both pathways *A* (HND) and *B* (HNS) in *Scheme 12* lead to the loss of the base (**B**) and to β-elimination at 3′- or 5′-position (**A** and **C**). The HNS pathway is preferential in H_2O, cleavage occurring at C(3′)−O. The HND pathway, in turn, takes place in formamide, giving rise to cleavage both at positions 3′ and 5′. The stability of both the 3′- and 5′-phosphoester bonds in oligonucleotides was determined as a function of temperature and of varying H_2O/formamide ratios [76].

Formamide exerts differential effects on the stability of the 3′- and of the 5′-phosphoester bonds in monomers *vs.* polymers. The data obtained by means of the

Scheme 12. *Mechanisms of DNA Degradation. Pathway A:* 'hydrolysis by nucleophilic degradation' (HND), consisting of nucleophilic attack of formamide or OH^- at C(8) of the purine (adenine) ring, followed by nucleobase degradation/elimination (**B**) and two β-eliminations (**A**) under C(3′)–O and C(5′)–O bond cleavage (**C**). *Pathway B:* 'hydrolysis by nucleophilic substitution' (HNS), consisting of direct, OH^--promoted nucleophilic substitution of adenine, followed by two β-eliminations under C(3′)–O and C(5′)–O bond cleavage (modified from [76]).



analysis described above regarding the stability of the β-glycosidic bond and of the phosphoester bonds in monomers and polymers reveal strong and differential sensitivity. These bonds are highly sensitive to variations of the environmental conditions. Rationalization for a large part of these differential stabilities were provided [76], based on state-of-the-art chemical knowledge and on previous observations.

Independently on the detailed chemical mechanisms involved, the plot of the half-lives of the 5′-phosphoester bonds in 5′-dAMP monomer and in its oligomeric form as a function of formamide concentration reveals a relevant property: formamide *decreases* the half-life of this bond in the monomer, but *increases* it in the polymer. The effect is strong: $t_{1/2}$ values of 28.5×10^3 *vs.* 0.23×10^3 min were found for monomeric dAMP *vs.* polymer, which corresponds to a factor of 125. The situation, however, is reversed when starting at *ca.* 75% formamide content in H_2O. In neat formamide, the 5′-phosphoester bond becomes *more stable* in DNA than in the precursor monomer [76]. For the 3′-phosphoester bond, the phenomenon and general trend are similar, although somewhat less marked: $t_{1/2} = 2.55 \times 10^3$ *vs.* 0.14×10^3 min, which corresponds to a factor of 18.3.

In summary, the stabilities of the phosphoester bonds in H_2O for monomers relative to those in DNA differ by approximately two orders of magnitude for the 5′-bond, and

by more than one order of magnitude for the 3′-bond. Given similar environmental conditions, these two types of bonds would be broken, respectively, *ca.* 125- and 18-times faster in DNA than in the precursor monomer. In this physico-chemical frame, polymerization would obviously be forbidden. In other words: *formamide reverses these stability parameters, potentially allowing the survival of the polymeric form.*

These measurements and findings provide an assay for the evaluation of the conditions, the catalysts, the shielding surfaces, and the physico-chemical environment that could sufficiently further enhance the stability of the polymers relative to that of the monomers.

10.3. *RNA*. The stabilities of the bonds critical for the half-life of ribonucleotides (namely the β-glycosidic and the 3′- and 5′-phosphoester bonds) were also measured under a wide range of temperatures and H_2O/formamide ratios [129]. The stability of the phosphoester bonds in oligonucleotides was determined under the same conditions as described above for DNA. The comparison of bond stabilities in the monomers with that of the polymers revealed that also for the ribose system physico-chemical conditions exist in which *the polymeric state of phosphorylated nucleosides is thermodynamically favored over the monomeric one* (*Figure*). The comparison is

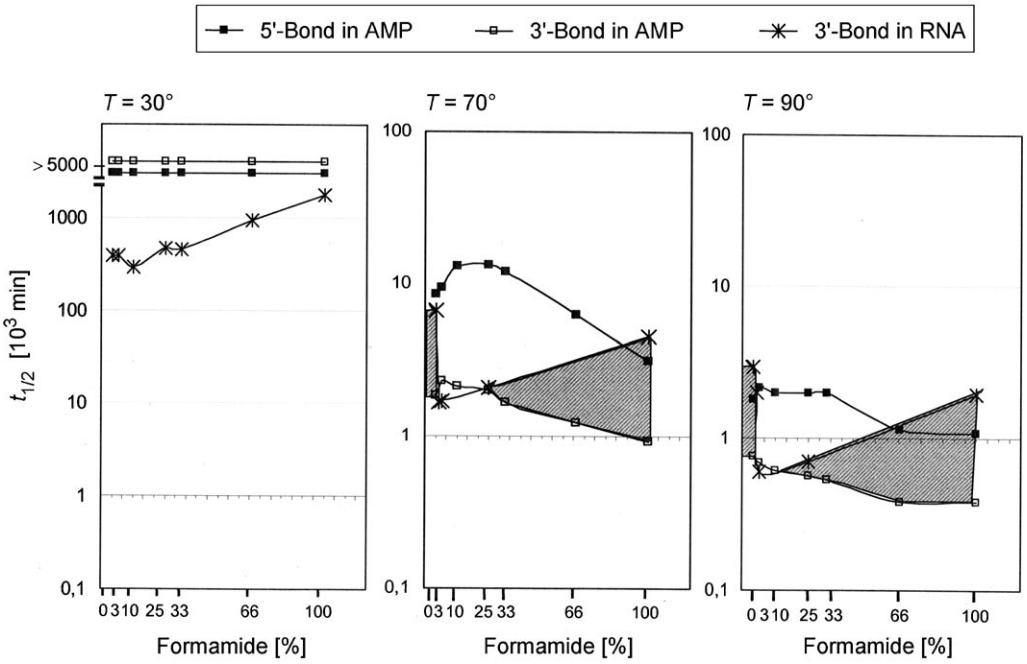


Figure. *Stability of phosphoester bonds in monomer* (AMP) *and polymer* (RNA) *in aqueous formamide solution*. The half-life of the 3′-phosphoester bond in RNA is compared with that in 3′-AMP, and of the phosphoester bond in 5′-AMP as a function of formamide concentration and temperature. Data were taken from [129]. Grey shades indicate areas (conditions) in which bond stability is higher in the polymer than in the monomer.

limited to the 3′-phosphoester bond, the only bond that in RNA is appreciably cleaved. Interestingly, the set of conditions in which the polymer is favored is wider for RNA than for DNA: the 'thermodynamic niche' in which RNA is favored as a polymer over its monomers (and over DNA) extends from pure H_2O to pure formamide, markedly so at temperatures of *ca.* 70 and 80° [129].

11. Evolving Equilibria: A Cyclic Scenario for Formamide Prebiotic Chemistry. – The very fact that formamide yields just by heating in the presence of catalysts the full panel of nucleic bases shows the possibility of their accumulation at elevated temperature, and reveals that conditions exist allowing a positive balance between syntheses and degradations. In terms of the ability to form informational macromolecules, the intrinsic instability of the bases [1–6] and their prompt synthesis [44][60][64][66], as reviewed in [39], are not only disadvantages. At the contrary, a synthesis/degradation/re-synthesis cycle offers the possibility to form a dynamically equilibrated pool of precursors, whose composition depends on the synthesis/degradation rate of each molecular species, and on the catalyst(s) present. The advantage provided by a flexible and adjustable pool of precursors is an important evolutionary property. This so-called 'equilibration-of-the-pool' scheme [39][46] is based on two simple considerations:

- The mechanism of degradation of purines [128] and pyrimidines [130] by formamide shows that formamide acts initially by nucleophilic addition, rapidly leading to base degradation. This process eventually re-affords formamide, and is markedly more efficient for purines than for pyrimidines. The order of base degradation is $G \geq A > C \gg T$ [131]. In the absence of catalysts, all the bases are eventually degraded back to formamide.
- On the other hand, in the presence of the appropriate catalysts, formamide condenses into purines and pyrimidines [44][60][64][66], the yield of each base depending on the catalyst present. Each catalyst affords a complex and specific panel of nucleic bases.

Without a way out towards polymers, and/or in the absence of protective mechanisms, conditions, or compounds, the cycle of syntheses/degradations is bound to remain pre-biologically futile. The final composition of the pool depends on the equilibrium between syntheses and degradations of each base under the local, specific physico-chemical conditions. Let us dub this phenomenon 'base-equilibration pool'.

Another equilibration pool relates to phosphorylations. In a previous section, we reported that nucleosides are promptly phosphorylated in the presence of formamide and of inorganic phosphates, as originally reported by *Schoffstall*'s and co-workers [89–92]. We have observed that comparable levels of nucleoside phosphorylation are obtained in the presence of formamide using a nucleoside monophosphate as phosphate donor. This phenomenon is caused by the formamide-mediated cleavage of the nucleotide phosphoester bond, followed by the transfer of the phosphate group to an acceptor site on the sugar moiety of an acceptor nucleoside. The acceptor may be the same starting nucleoside or a different one, phosphorylation occurring at any of the possible 2′-, 3′-, 5′-, 2′,3′-cyclic, and, presumably, also 3′,5′-cyclic positions. In a composite pool of nucleosides, the final composition of the nucleotide pool will depend

on the on/off rates of the phosphoester bond of each different nucleotide at each different position. The major factors affecting the final balance are *1*) the higher stability of the cyclic phosphates relative to open forms, and *2*) the preferential phosphorylation at the 5′-position, a phenomenon called 'phosphorylation-equilibrated pool'.

Although the same type of pool-equilibration behavior has not been yet studied for the formation of nucleosides, there is no *a priori* reason to exclude its occurrence.

The way out from futile syntheses/degradation cycles of nucleobases and of their phosphorylations could have been provided by polymerization itself. Thus, the comparison of the stability of the key bonds in the polymer relative to the stability of the same bonds in the monomers is a central point in understanding the basics of the thermodynamics and kinetics of polymerization.

If conditions were eventually met in the pristine warm little pond imagined by *Darwin* [132], conditions that favored polymerization, the sequence composition of the ensuing polymer necessarily reflects the composition of the 'base-equilibration pool' or the 'phosphorylation-equilibrated pool'. Given that the formation of a polymer subtracts from the pool, thus changing its composition and affecting the equilibria involved, polymerization itself could have been a relevant evolutionary factor.

12. *Occam*'s Razor. – In addition to its positive effects on surfactants leading to the formation/stability of micelles, as reported by *Akhter* and *Alawi* [133], and by *Shirota* and *Segawa* [134], we have described how formamide acts on the syntheses of the bases and on their degradation, on the formation of nucleosides, on their phosphorylation/dephosphorylation/transphosphorylation (notably on the formation of 2′,3′-cyclic nucleotides), and on the stability of polymers. The *Occam*'s razor logic principle states that among different alternatives the simplest one provides the solution. Considering that of all the steps from a one-carbon molecule to nucleic polymers the only (fundamental) remaining gap is that of the formation of the 3′,5′-phosphodiester bonds, and keeping in mind the versatility of formamide in allowing phosphorylations of nucleosides, *Occam*'s logic would state that formamide chemistry is the solution to polymerization. This is currently under experimental analysis.

Formamide has the potential to act as a general determinant of the pre-genetic equilibria in the not-yet-living *Darwinian* warm little pond – at least as long as the conditions were met in which the passage to H_2O would chemically and thermodynamically be permitted.

This work was supported by the *Italian Space Agency* (ASI Project MoMa).

REFERENCES

[1] M. Levy, S. C. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7933.
[2] J. L. Bada, A. Lazcano, *Science* **2002**, *296*, 1982.
[3] S. L. Miller, A. Lazcano, *J. Mol. Evol.* **1995**, *41*, 689.
[4] M. Levy, S. L. Miller, K. Brinton, J. L. Bada, *Icarus* **2000**, *145*, 609.
[5] S. Miyakawa, H. J. Cleaves, S. L. Miller, *Origins Life Evol. Biosphere* **2002**, *32*, 195.
[6] S. Miyakawa, H. J. Cleaves, S. L. Miller, *Origins Life Evol. Biosphere* **2002**, *32*, 209.
[7] F. Westall, 'Astrobiology: Future Perspectives', Kluwer/Springer, Dordrecht, 2004, Chapt. 12, p. 287.

[8] L. P. Knauth, D. R. Lowe, *Geol. Soc. Am. Bull.* **2003**, *115*, 566.
[9] D. R. Lowe, G. R. Byerly, F. T. Kyte, A. Shukulyukov, F. Asaro, A. Krull, *Astrobiology* **2003**, *3*, 7.
[10] F. T. Kyte, A. Shukolyukov, G. W. Lugmaor, D. R. Lowe, G. R. Byerly, *Geology* **2003**, *31*, 283.
[11] N. T. Arndt, in 'Archean Crustal Evolution', Ed. K. C. Condie, Elsevier, Amsterdam, 1994, p. 11.
[12] T. J. Millar, 'Astrobiology: Future Perspectives', Kluwer/Springer, Dordrecht, 2004, Chapt. 2, p. 17, and refs. cit. therein.
[13] A. J. Markwick, S. B. Charnley, 'Astrobiology: Future Perspectives', Kluwer/Springer, Dordrecht, 2004, Chapt. 3, pp. 33–66, and refs cit. therein.
[14] J. Crovisier, 'Astrobiology: Future Perspectives', Kluwer/Springer, Dordrecht, 2004, Chapt. 8, p. 179–203, and refs. cit. therein.
[15] D. Bockelee-Morvan, D. C. Lis, J. E. Wink, *Astron. Astrophys.* **2000**, *353*, 1101.
[16] W. A. Schutte, A. C. A. Boogert, A. G. G. M. Tielens, D. C. B. Whittet, P. A. Gerakines, J. E. Chiar, P. Ehrenfreund, J. M. Greenberg, E. F. van Dishoeck, T. de Graauw, *Astron. Astrophys.* **1999**, *343*, 966.
[17] Personal communication, K. Hand, R. W. Carlson, Department of Geological and Environmental Sciences, Stanford University, September 2006.
[18] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99.
[19] J. Oró, *Biochim. Biophys. Res. Commun.* **1960**, *2*, 407.
[20] J. Oró, *Nature* **1961**, *191*, 1193.
[21] F. H. Westheimer, *Science* **1987**, *235*, 1173.
[22] M. Egholm, O. Buchardt, P. E. Nielsen, R. H. Berg, *J. Am. Chem. Soc.* **1992**, *114*, 1895.
[23] J. G. Schmidt, L. Christensen, P. E. Nielsen, L. E. Orgel, *Nucleic Acids Res.* **1997**, *25*, 4792.
[24] S. A. Benner, D. Hutter, *Bioorg. Chem.* **2002**, *30*, 62.
[25] M. O. Beatler, D. Hutter, S. Benner, *Helv. Chim. Acta* **2002**, *85*, 2777.
[26] S. K. Singh, P. E. Nielsen, A. A. Koshkin, J. Wengel, *Chem. Commun.* **1998**, 455.
[27] J. Wengel, *Acc. Chem. Res.* **1999**, *32*, 301.
[28] S. Ilin, I. Schlönvogt, M. O. Ebert, B. Jaun, H. Schwalbe, *ChemBioChem* **2002**, *3*, 93.
[29] D. Zhou, I. M. Lagoja, J. Rozenski, R. Busson, A. Van Aerschot, P. Herdewijn, *ChemBioChem* **2005**, *6*, 2298.
[30] R. Declercq, A. Van Aerschot, R. J. Read, P. Herdewijn, L. Van Meervelt, *J. Am. Chem. Soc.* **2002**, *124*, 928.
[31] P. Gu, G. Schepers, C. Griebel, J. Rozenski, H.-J. Gais, P. Herdewijn, A. Van Aerschot, *Nucleosides Nucleotides Nucleic Acids* **2005**, *24*, 993.
[32] K. Schoning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347.
[33] C. J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser, M. Egli, *J. Am. Chem. Soc.* **2002**, *124*, 13716.
[34] D. G. Schultz, S. M. Gryaznov, *Nucleic Acids Res.* **1996**, *24*, 2966.
[35] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497.
[36] J. G. Schmidt, P. E. Nielsen, L. E. Orgel, *Nucleic Acids Res.* **1997**, *25*, 4797.
[37] H. D. Bean, F. A. L. Anet, I. R. Gould, N. V. Hud, *Origins Life Evol. Biosphere* **2006**, *36*, 39.
[38] R. Pascal, L. Boiteau, A. Commeyras, *Top. Curr. Chem.* **2005**, *259*, 69.
[39] R. Saladino, C. Crestini, G. Costanzo, E. Di Mauro, *Top. Curr. Chem.* **2005**, *259*, 29.
[40] J. P. Ferris, *Rev. Mineral. Geochem.* **2005**, *59*, 187.
[41] Q. W. Chen, C. L. Chen, *Curr. Org. Chem.* **2005**, *9*, 989.
[42] R. Saladino, C. Crestini, G. Costanzo, E. Di Mauro, *Curr. Org. Chem.* **2004**, *8*, 1425.
[43] E. C. T. Kirk-Othmer, in 'Encyclopedia of Chemical Tecnology, Formic Acid and Derivatives (Formamide)', Wiley Interscience, 1978, Vol. 11, p. 258.
[44] R. Saladino, C. Crestini, G. Costanzo, R. Negri, E. Di Mauro, *Bioorg. Med. Chem.* **2001**, *9*, 1249.
[45] R. Saladino, C. Crestini, F. Ciciriello, G. Costanzo, R. Negri, R. E. Di Mauro, ESA, [Special Publication] SP, 2004, SP-545(Proceedings of the III European Workshop on Exo-Astrobiology, 2003, 287).

[46] R. Saladino, C. Crestini, F. Ciciriello, G. Costanzo, R. Negri, E. Di Mauro, 'Astrobiology: Future Perspectives', Kluwer/Springer, Dordrecht, 2004, Chapt. 16, p. 393.
[47] H. Yamada, T. Okamoto, *Chem. Pharm. Bull.* **1972**, *20*, 623.
[48] H. Yamada, T. Okamoto, Z. Yakugaku, *Chem. Pharm. Bull.* **1975**, *95*, 493.
[49] C. Reid, L. E. Orgel, C. Ponnamperuma, *Nature* **1967**, *216*, 936.
[50] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25.
[51] E. A. Kuzicheva, N. V. Tsupkina, *Zhurnal Evolyutsionnoi Biokhimii i Fiziologii* **1978**, *14*, 213.
[52] E. Kirkpatrick, to *Dupont de Nemours & Co. Inc.*, U.S. Pat. 134,765, 1944.
[53] H. Slebocka-Tilk, F. Sauriol, M. Monette, R. S. Brown, *Can. J. Chem*. **2002**, *80*, 1343.
[54] R. Saladino, C. Crestini, V. Neri, F. Ciciriello, G. Costanzo, E. Di Mauro, *ChemBioChem* **2006**, *7*, 1707.
[55] K. T. Suzuki, H. Yamada, M. Hirobe, *J. Chem. Soc., Chem. Commun*. **1978**, 485.
[56] H. Yamada, M. Hirobe, K. Higashiyama, H. Takahashi, K. T. Suzuki, *Tetrahedron Lett*. **1978**, *19*, 4039.
[57] H. Yamada, M. Hirobe, K. Higashiyama, H. Takahashi, K. T. Suzuki, *J. Am. Chem. Soc.* **1978**, *100*, 4617.
[58] S. D. Senanayake, H. Idriss, *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1194.
[59] M. P. Robertson, S. L. Miller, *Nature* **1995**, *375*, 772.
[60] R. Saladino, U. Ciambecchini, C. Crestini, G. Costanzo, R. Negri, E. Di Mauro, *ChemBioChem* **2003**, *4*, 541.
[61] D. A. Friesen, J. V. Haedely, C. H. Langford, *Environm. Sci. Technol*. **1999**, *33*, 3193.
[62] A. S. U. Choughuley, M. S. Chada, A. S. Subbaraman, *Biosystems* **1977**, *9*, 73.
[63] H. W. Gibson, *Chem. Rev.* **1969**, *69*, 673.
[64] R. Saladino, C. Crestini, V. Neri, J. R. Brucato, L. Colangeli, F. Ciciriello, E. Di Mauro, G. Costanzo, *ChemBioChem* **2005**, *6*, 1368.
[65] A. Eschenmoser, E. Loewenthal, *Chem. Soc. Rev.* **1992**, 1.
[66] R. Saladino, C. Crestini, U. Ciambecchini, F. Ciciriello, G. Costanzo, E. Di Mauro *ChemBioChem* **2004**, *5*, 1558.
[67] J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **1961**, *94*, 217; A. Eschenmoser, E. Loewenthal, *Chem. Soc. Rev.* **1992**, *21*, 1; L. E. Orgel, *Trends Biochem. Sci.* **1998**, *23*, 491.
[68] J. D. Sutherland, J. N. Whitfield, *Tetrahedron* **1997**, *34*, 11493.
[69] G. F. Joyce, *Nature* **1989**, *338*, 217.
[70] G. Zubay, T. Mui, *Origins Life Evol. Biosphere* **2001**, *31*, 87.
[71] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol.* **1972**, *1*, 249.
[72] A. Eschenmoser, *Origins Life Evol. Biosphere* **1997**, *27*, 535.
[73] A. W. Schwartz, P. M. De Graaf, *J. Mol. Evol*. **1993**, *36*, 101.
[74] K. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347.
[75] A. A. Ingar, R. W. Luke, B. R. Hayter, J. D. Sutherland, *ChemBioChem* **2003**, *4*, 504.
[76] R. Saladino, C. Crestini, V. Busiello, F. Ciciriello, G. Costanzo, E. Di Mauro, *J. Biol. Chem.* **2005**, *280*, 35658.
[77] I. A. Kozlov, B. De Bouvere, A. Van Aerschot, P. Herdewijn, L. E. Orgel, *J. Am. Chem. Soc*. **1999**, *121*, 5856.
[78] M. Tohidi, L. E. Orgel, *J. Mol. Evol.* **1989**, *28*, 367.
[79] I. A. Kozlov, P. K. Politis, S. Pitsch, P. Herdewijn, L. E. Orgel, *J. Am. Chem. Soc*. **1999**, *121*, 1108.
[80] A. Butlerow, *Liebigs Ann. Chem*. **1861**, *120*, 295.
[81] R. Breslow, *Tetrahedron Lett.* **1959**, *21*, 22.
[82] J. E. Rekoske, M. A. Barteau, *Langmuir* **1999**, *15*, 2061.
[83] K. E. Nelson, M. Levy, S. L. Miller, *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3868.
[84] A. Beck, R. Lohrmann, L. E. Orgel, *Science* **1967**, *157*, 958.
[85] R. Lohrmann, L. E. Orgel, *Science* **1968**, *161*, 64.
[86] R. Lohrmann, L. E. Orgel, *Science* **1971**, *171*, 490.
[87] R. Lohrmann, *J. Mol. Evol.* **1977**, *10*, 137.

[88] R. Reimann, G. Zubay, *Origins Life Evol. Biosphere* **1999**, *29*, 229.
[89] A. M. Schoffstall, *Origins of Life* **1976**, *7*, 399.
[90] A. M. Schoffstall, R. J. Barto, D. L. Ramos, *Origins of Life* **1982**, *12*, 143.
[91] A. M. Schoffstall, E. M. Laing, *Origins of Life* **1985**, *15*, 141.
[92] A. M. Schoffstall, S. M. Mahone, *Origins Life Evol. Biosphere* **1988**, *18*, 389.
[93] K. Van Holde, 'The Origins of Life and Evolution', Eds. H. O. Halvorson and K. E. van Holde, Alan R. Liss, Inc., New York, 1980, p. 31.
[94] C. Cheng, C. Fan, R. Wan, C. Tong, Z. Miao, J. Chen, Y. Zhao, *Origins Life Evol. Biosphere* **2002**, *32*, 219.
[95] J. P. Ferris, H. Yanagawa, W. J. Hagan Jr., *Origins of Life* **1984**, *14*, 99.
[96] D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Witner, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1410.
[97] R. M. De Graaf, J. Visscher, A. W. Schwartz, *Origins Life Evol. Biosphere* **1998**, *28*, 271.
[98] A. W. Schwartz, *J. Theor. Biol.* **1997**, *187*, 523.
[99] G. von Kiedroski, *Angew. Chem., Int. Ed.* **1986**, *25*, 932.
[100] G. von Kiedroski, *Biorg. Chem. Front.* **1993**, *3*, 113.
[101] L. E. Orgel, *Nature* **1992**, *358*, 203.
[102] D. Sievers, G. von Kiedrowski, *Nature* **1994**, *369*, 221.
[103] K. C. Li, T. Nicolai, *Nature* **1994**, *369*, 218.
[104] J. Sulston, R. Lohrmann, L. E. Orgel, H. T. Miles, *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *59*, 726.
[105] B. J. Wieman, R. Lohrmann, L. E. Orgel, H. Schneider-Bernloehr, J. E. Sulston, *Science* **1968**, *161*, 387.
[106] G. F. Joyce, T. Inoue, L. E. Orgel, *J. Mol. Biol.* **1984**, *176*, 279.
[107] K. J. Prabahar, J. P. Ferris, *J. Am. Chem. Soc.* **1997**, *119*, 4330.
[108] J. P. Ferris, A. R. Hill Jr., R. Liu, L. E. Orgel, *Nature* **1996**, *381*, 59.
[109] G. Ertem, J. P. Ferris, *Nature* **1996**, *379*, 238.
[110] G. Ertem, J. P. Ferris, *Origins Life Evol. Biosphere* **1998**, *28*, 485.
[111] A. R. Hill, L. E. Orgel, T. Wu, *Origins Life Evol. Biosphere* **1993**, *23*, 285.
[112] M. Renz, R. Lohrmann, L. E. Orgel, *Biochim. Biophys. Acta* **1971**, *240*, 463.
[113] G. A. Soukup, R. R. Breaker, *RNA* **1999**, *5*, 1308.
[114] D. A. Usher, A. H. McHale, *Science* **1976**, *192*, 53.
[115] D. A. Usher, A. H. McHale, *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 1149.
[116] J. A. Gerlt, F. H. Westheimer, J. M. Sturtevant, *J. Biol. Chem.* **1975**, *250*, 463.
[117] F. H. Westheimer, *Acc. Chem. Res.* **1968**, *1*, 70.
[118] S. C. Mohr, R. E. Thach, *J. Biol. Chem.* **1969**, *244*, 6566.
[119] J. E. Ermen, G. G. Hammes, *J. Am. Chem. Soc.* **1996**, *88*, 5607.
[120] N. K. Kochetkov, E. I. Budovskii, E. D. Sverdlov, N. A. Simukova, M. F. Turchinskii, V. N. Shibaev, in 'Organic Chemistry of Nucleic Acids', Eds. N. K. Kochetkov, E. I. Budovskii, Plenum Press, London, New York, 1972, Part B, Chapt. 8, p. 425, and Chapt. 10, p. 472.
[121] R. Shapiro, S. Kang, *Biochemistry* **1969**, *8*, 1806.
[122] E. R. Garrett, P. J. Mehta, *J. Am. Chem. Soc.* **1972**, *94*, 8542.
[123] T. Lindahl, O. Karlström, *Biochemistry* **1973**, *12*, 5151.
[124] H. S. Shapiro, E. Chargaff, *Biochim. Biophys. Acta* **1957**, *26*, 596.
[125] H. Venner, *Hoppe Seylers Z. Physiol. Chem.* **1966**, *344*, 189.
[126] R. Shapiro, M. Danzig, *Biochim. Biophys. Acta* **1973**, *319*, 5.
[127] T. Lindahl, B. Nyberg, *Biochemistry* **1972**, *11*, 3610.
[128] R. Saladino, E. Mincione, C. Crestini, R. Negri, E. Di Mauro, G. Costanzo, *J. Am. Chem. Soc.* **1996**, *118*, 5615.
[129] R. Saladino, C. Crestini, F. Ciciriello, E. Di Mauro, G. Costanzo, *J. Biol. Chem.* **2006**, *281*, 5790.
[130] R. Saladino, C. Crestini, E. Mincione, G. Costanzo, E. Di Mauro, R. Negri, *Bioorg. Med. Chem.* **1997**, *5*, 2041.
[131] R. Negri, G. Costanzo, R. Saladino, E. Di Mauro, *BioTechniques* **1996**, *21*, 910.

[132] F. Darwin, 'The Life and Letters of Charles Darwin', John Murray, London, 1888, Vol. 3, p. 18 (letter to Joseph Hooker).
[133] M. S. Akhter, S. M. Alawi, *Colloids Surf.* **2003**, *219*, 281.
[134] H. Shirota, H. Segawa, *Langmuir* **2004**, *20*, 329.

Received September 18, 2006