Published in final edited form as:

Chemistry. 2017 September 12; 23(51): 12668–12675. doi:10.1002/chem.201702912.

Orotidine-Containing RNA: Implications for the Hierarchical Selection (Systems Chemistry Emergence) of RNA

Eun-Kyong Kim, Vincent Martin, and Ramanarayanan Krishnamurthy^[a]

^[a]Dr. E.-K. Kim, Dr. V. Martin, Prof. Dr. R. Krishnamurthy, Department of Chemistry, The Scripps Research Institutem, 10550, North Torrey Pines Rd, La Jolla, Ca 92037 (USA)

Abstract

The prebiotic synthesis of canonical nucleobases from HCN is a cornerstone for the RNA world hypothesis. However, their role in the primordial pathways to RNA is still debated. The very same process starting from HCN also gives rise to orotic acid, which (via orotidine) plays a crucial role in extant biology in the de novo synthesis of uridine and cytidine, the informational base-pairs in RNA. However, orotidine itself is absent in RNA. Given the prebiotic and biological relevance of orotic acid vis-à-vis uracil, we investigated orotidine-containing RNA oligonucleotides and show that they have severely compromised base-pairing properties. While not unexpected, these results suggest that the emergence of extant RNA cannot just be a consequence of the plausible prebiotic formation of its chemical constituents/building blocks. In combination with other investigations on alternative prebiotic nucleobases, sugars, and linkers, these findings imply that the selection of the components of extant RNA occurred at a higher hierarchical level of an oligomer/polymer based on its functional properties—pointing to a systems chemistry emergence of RNA from a library of precursors.

Keywords

base pairing; orotic acid; orotidine; prebiotic systems chemistry; RNA world

Introduction

The appearance of RNA—and in general the emergence of informational and functional oligomers—in the context of origin of life studies is still an enigma.^[1–3] In that context, it is widely accepted that the prebiotic origination of the four canonical nucleobases (adenine, uracil, guanine, and cytosine) and their extant biological function are directly intertwined.^[1–4] This is based on the vital functional role they play in extant biology and the ease of formation under potentially prebiotic conditions.^[4] Apart from these canonical nucleobases, there is a whole set of pyrimidines (and purines), with functional group substitutions at various positions, that are produced under the same prebiotic conditions.^[5–9] Particularly, pyrimidines with functional groups (NH₂, OH, or COOH) at the 5- or 6-position are formed,

Conflict of interest

The authors declare no conflict of interest.

Dedicated to Professor Alan W. Schwartz for his seminal contributions in origin of life research

sometimes with greater abundance in the HCN polymerization reactions.^[5–9] For example, Miller and co-workers have shown that when HCN oligomers are hydrolyzed around pH 8, the canonical nucleobases A, U, G, and C are formed.^[9] Intriguingly, another “biological nucleobase”, orotic acid, is also formed, and in much higher proportions (50 times more than uracil) among all of the pyrimidines.^[9] Orotic acid is a different kind of “biological nucleobase” that plays a crucial role in contemporary biology.^[10–15] Extant *de novo* biosynthetic pathways use orotic acid via orotidine (O), by a direct nucleosidation on the 5-phosphoribosyl α -diphosphate, to synthesize the canonical pyrimidine nucleotides (uridine and cytidine) in RNA, and (thymidine) in DNA, which function as the informational base-pairing partners (Figure 1).^[15] While paralleling the prebiotic formation potential coupled with biological function ascribed for the canonical nucleobases, orotidine however, differs from the rest of the canonical nucleobase in the sense that it is not part of the extant informational base-pairing alphabet set. Such a juxtaposition of the central role of orotic acid in pyrimidine biosynthesis along with its prebiotic relevance—versus its conspicuous absence in RNA—prompted our interest in investigating the efficacy of base-pairing of orotidine in RNA oligomers.

Base-pairing properties of orotidine, to date, have not been studied. The conformational preference of the β -anomer of orotidine nucleos(t)ides has been investigated by various groups indicating that the *syn*-conformation is preferred, with the COOH group essentially perpendicular to the pyrimidine ring.^[16–19] Such a *syn*-disposition of the nucleobase in the monomer unit has been proposed to be detrimental for stacking capability of orotic acid in a dimer.^[19] Hudson and co-workers were the first to investigate the base-pairing of orotic acid in an oligo-amide backbone (PNA), and showed that orotic acid, inserted in a PNA backbone-(T)₆ sequence, connected through the N(1)-position (Figure 2a), destabilized PNA₂:NA triplexes with poly(rA) or poly(dA).^[20] They concluded that the steric hindrance offered by the C(6)-substituent and the negative charge on the carboxylate were responsible for the destabilization.

We also had studied the base-pairing of orotic acid, but with a different connection: using the C(6)-carboxylate group to tether through an amide bond to a peptide backbone (Figure 2b).^[21] Our system did not base-pair, but for an entirely different reason, “the pK_a -pH effect”.^[22] The pK_a of N(3) orotamide moiety in our system is around 6.6, which leads to deprotonation at neutral pH and solvation of the deprotonated nucleobase, which interferes with base-pairing.^[22] This hypothesis was supported by making the corresponding 2,4-diamino derivative (pK_a 4.7), which was an efficient base-pairing partner (Figure 2c). In orotidine the pK_a of N(3) is 9.1 (like uracil), which based on the “ pK_a -pH argument”^[22] should be able to base pair. However, based on the results from the Hudson group,^[20] and the *syn*-conformational preference of the orotic acid^[16–19] in orotidine, weaker base-pairing properties could be expected. It is with this background we began our investigation of orotidine-containing RNA—inspired by the context of orotic acid’s availability by the same prebiotic routes that form uracil, and orotic acid’s current role in biosynthesis to give rise to the canonical uridine nucleotide, with a view to understand the pathways to the emergence of RNA.

Results and Discussion

In order to access the requisite orotidine–phosphoramidite (3) for automated oligomer synthesis, we developed and recently reported on a new route to the synthesis of orotidine by means of an intramolecular nucleosidation.^[23] With the methylester derivative of orotidine (1)^[23] as the starting point, we synthesized phosphoramidite (3) in five steps (Scheme 1) in good yields, suitable for the automated solid-support synthesis.

The synthesis of orotidine-containing RNA oligomers was straightforward. However, the deprotection of the methylester (of orotate) to liberate the carboxylate at the oligomer level, over a wide range of conditions (Table S1, Supporting Information), led to extensive strand scission and decomposition of the oligonucleotides. Consequently, homogenous orotidine sequences or oligomers with extensive incorporations of orotidine were not accessible. Therefore, we settled on strategic insertions of one or two orotidine residues within self-complementary and non-self-complementary RNA sequences (Table 1). We developed a three-step deprotection protocol:

1. Cyanoethyl deprotection was carried out first with Et₃N in pyridine.
2. Simultaneous deprotection of the methyl ester of orotidine and the canonical nucleobase protecting groups and detachment from the solid support were achieved, optimally, by use of 50 mM aq NaOH at room temperature over two days.
3. Cleavage of the 2'-*O*-TBDMS was achieved by use of Et₃N × 3HF.

This three-step protocol was developed and optimized by testing on a trimer and a tetramer of RNA, containing one and two orotidine inserts, respectively, and then was extended to other oligonucleotides. All HPLC-purified oligonucleotides were characterized by MALDI-TOF mass spectral data (Table S2, Supporting Information). The base-pairing properties of these orotidine (O)-containing RNA oligomers were investigated by temperature-dependent UV spectroscopy in standard phosphate buffer conditions with 1 M NaCl. We studied the effect of orotidine insertions in self-complementary, non-self-complementary, and the shortened Dickerson–Drew octamer duplexes and compared their thermal stability with the respective unmodified RNA duplexes (Table 1).

Self-complementary sequences bearing one orotidine (O) at 5'- or 3'-end resulted in lower thermal stability of the duplexes by about 4°C compared to the parent r(AU)₈ (Table 1, entry 2 and 3 vs. entry 1). Two juxtaposed O units at the both end of duplex reduced *T_m* value about 7°C (entry 4). The presence of one juxtaposed O in the middle of self-complementary duplex reduced the stability also by 7°C, indicating that the complementary AO pairs in the middle of duplex disturb the duplex stability to greater extent (entry 5, Figure 3A). Furthermore, when the O insert was offset from the center there was a (unexpected) dramatic drop in duplex stability (entry 6). The duplex containing two O inserts at the same position in AU alternate hexadecamer abolished duplex formation indicating that orotidine insertion is highly destabilizing (entry 7). The impact of orotidine incorporation in non-self-complementary duplexes was more drastic. Only one O residue reduced the duplex stability by ≈ 17°C (entry 16 vs. 15, Figure 3B) and two O inserts lowered *T_m* value by more than

double (entry 17, $T_m=38.3^\circ\text{C}$). The broadness of the UV T_m curves of orotidine-containing sequences (compared to unmodified sequences) may also indicate weakening of the duplexes by fraying of the ends or “bubbles” in the middle where orotidine residues are present.

In the case of the short Dickerson–Drew octamer, one interpolation of orotidine in the middle of sequence substantially reduced the stability of duplex (by ca. 37°C , entry 23 vs. entry 22, Figure 4) and there was no duplex formation when the short Dickerson–Drew octamer contained two orotidine residues in the middle of sequence (entry 24). These observations are in consonance with results of Hudson observed in the acyclic PNA systems. [20]

The unusual high thermal instability of the duplex from the insertion of even one orotidine unit suggests there is more than a simple weakening—by lack of hydrogen bonding or stacking—when compared to uridine. To understand the magnitude of the thermal instability of the AO pair we compared these results with duplex instability generated by mismatch base-pairs AA and UU.

The impact on duplex destabilization of AA and UU mismatches at the end of the sequences was insignificant (entry 9 and 10). Moreover, the two mismatch residues in the middle of sequences destabilized the duplexes by 9.9°C for AA mismatches and 7.2°C for UU mismatches, respectively (entries 11 and 12). The sequences containing the four AA and UU mismatches offset from the center of the duplex destabilized the duplexes significantly (entries 13 and 14). Especially, entry 13 with four AA mismatches hardly formed a duplex matching that of the AO duplex in entry 7. These mismatch studies suggest that the AO pair is comparable to the purine–purine mismatch rather than a pyrimidine–pyrimidine mismatch (entries 7 versus 13 and 14). The same trend was observed for the short Dickerson–Drew sequence (entries 23 versus 25 and 27, Figure 4). Such a “purine-mismatch”-like behavior of orotidine indicates that perhaps the C(6)-carboxylate may be inserting itself into the duplex, due to the *syn*-preference of orotidine and, thus, causing a steric clash mimicking a “purine”. We also tested a pyrimidine–orotidine mismatch and found it was still highly destabilizing (entry 29), lending further credence to the hypothesis that it is the orotidine insertion into the stacking structure, irrespective of its complementary partner, that may be responsible for the remarkable destabilization. One possibility is that the carboxylate moiety’s negative charge (in the *syn*-orientation) within the duplex could interfere with the hydrophobic stacking. To see if the negative charge of the carboxylate group was responsible, we prepared the neutral C(6)-carboxamide derivative of orotidine by deprotection of oligonucleotide using the standard deprotection method ($\text{NH}_4\text{OH}:\text{EtOH}$ 3:1, 55°C , 16 h), and the hybridization behavior of orotyl amide containing duplex was investigated. It was found that insertion of one orotidyl amide destabilized the duplex even more than orotidine (entry 8 vs. entry 5). This observation argues against the negatively charged carboxylate as a significant factor for such a dramatic destabilization caused by orotidine. Rather, it reinforces the previous observations^[17–20] that the *syn*-orientation of orotidine, which leads to the presence of the C(6) carboxylate(amide) group within the minor groove of the Watson–Crick base-paired duplex itself, is the major cause. The C(6)-carboxylate(amide) group, when perpendicular to the pyrimidine plane, would lead to unfavorable interactions with the base-pair stacks above

and below, causing further steric clashes (Figure 5). This leads to larger than usual destabilization observed for pyrimidine–purine mismatch,^[24,25] or 6-methyl substitution,^[26a] or pyrimidine–pyrimidine mismatch^[25] or abasic site^[25] within the duplex mimicking a purine–purine mismatch.

The preference for the *syn*-conformation of the β -anomer of orotidine nucleos(t)ides has been documented as pointed out earlier, with the COOH group essentially being perpendicular to the pyrimidine ring.^[16–19] A *syn*-disposition of the nucleobase seen in the monomer unit, if also maintained in the oligonucleotide, would be detrimental for base-pairing capability of orotic acid in the Watson–Crick mode, as our results seem to confirm. A circular dichroism (CD) study by Holy on a series of GpU dinucleotides with 6-methyl-U and orotic acid in place of U, showed no stacking at the dimer level, suggesting that the *syn*-conformation imposed by the C(6)-substituents may be responsible destabilization of stacked structures.^[20] Substituents at the 6-position of uridine (such as a 6-methyl or formyl group) are also known to decrease the stability of an oligomeric duplex.^[26] Another factor that could weaken the duplex is the possible presence of the enol tautomer of orotidine. However, spectroscopic studies of orotic acid derivatives show the predominance of the keto form with little or no enol tautomer present,^[21,27] discounting this possibility.

Implications for RNA emergence

That orotidine (in place of uridine) substantially weakens base-pairing in RNA, when combined with the results from studies of other potentially prebiotic nucleobases alternatives (see below), has implications for the proposed and generally accepted pathways^[1–4] for the appearance of RNA in the context of origin of life studies. The possible presence of alternative nucleobases as part of RNA or primordial informational systems has been proposed and studied by various groups.^[28] For example, “purine–purine base-pairing paradigm” as a forerunner of the extant “pyrimidine–purine” base-pairs has been proposed^[29] and investigated.^[30] Oligonucleotides containing 2,6-diaminopurine (which has been shown to be part of a suite of extra-terrestrial carbonaceous meteorites)^[31] have been considered as a pre-RNA candidate.^[30] Pyrimidine–pyrimidine base-pairing has also been studied within the context of pre-RNA systems as primitive informational oligomers.^[5,32] Other studies have considered even a wider variety of potentially prebiotic heterocycles that are able to self-assemble.^[6,33] An important question with all these alternative nucleobases is the following: if as hypothesized they were present in the primordial version of RNA, when (at what stage of evolution) did they stop being part of extant RNA?

While the prebiotic formation of uracil, cytosine, guanine, and adenine from HCN has been deemed relevant to their presence in RNA,^[3,34] extensive prebiotic nucleosidation studies involving ribose and these canonical pyrimidine and purine nucleobases have been unsuccessful in demonstrating the formation of the corresponding nucleosides.^[35] Such a result raises the possibility that the extant metabolic (pyrimidine/purine salvage) pathways may not be a reflection of the prebiotic nucleosidation pathways involving prebiotically available ribose and the canonical pyrimidine nucleobases.^[36] Moreover, there is no known *de novo* biosynthetic pathway of any of the canonical (or other alternative) nucleobases. Only orotic acid has that “privilege”, and is synthesized from aspartic acid and converted to

orotidine by reaction with a ribose–phosphate derivative (Figure 1).^[15] Furthermore, the facile (abiotic) photochemical decarboxylative conversion of orotidine-5'-phosphate to uridine-5'-phosphate,^[37,38] is taken as a support for “transition from prebiology to biology”, pointing to the close similarity between the prebiotic and contemporary biotic pathways.^[39,40] In such cases, it has been proposed that slow non-enzymatic reactions may have been improved by proteins at a later stage.^[41]

The difficulties of pyrimidine (U/C) and purine (A/G) nucleoside/nucleotide formation through direct prebiotic nucleosidation have led to intensive search for, and successful demonstration of, indirect routes for their formation starting with simpler precursors by plausible prebiotic pathways.^[42] While direct nucleosidation with orotate has also not been achieved, the same considerations have not been extended to orotidine—probably since 1) it is not part of extant RNA, and 2) it is generally considered part of a (later evolutionary) biochemical pathway.^[36] Such interpretations are based on analysis and extrapolation of extant biological pathways backwards to prebiotic chemistry (top-down approach). However, they are not adequately justified from a prebiotic chemistry point of view (bottom-up approach). Recently it has been proposed that orotic acid may offer a reactive advantage by forming an ester at the primary 5-OH group of a ribofuranosyl moiety, and could react intramolecularly (instead of intermolecularly) to form the orotidine nucleoside.^[23]

Notwithstanding the above prebiotic-to-biotic-transition correlation and arguments, the fact that orotidine substantially weakens base-pairing at an oligomeric level suggests that orotidine-containing RNA sequences would be constrained in their ability to form structures capable of function (replication and catalysis). However, there is also the other side of the argument that weakened base-pairing of RNA may be advantageous overcoming template-strand inhibition.^[43] However, it is doubtful that orotic acid with its severely destabilizing effect could have played such or any “constructive” role, but it is possible that orotidine residue in a loop or bulge could have catalytic abilities (e.g., ribozymes), if the carboxylate moiety was accessible through its *syn*-conformation. Another point of view is that uracil is the better “performer” and was selected (or took) over orotic acid. That being the case, it is then reasonable to argue that the gain-in-base-pairing propensity by converting orotidine to uridine would provide an evolutionary advantage for RNA from a structural and functional point of view.

However, for all of these gain-of-function arguments (in a base-pairing perspective) to be valid there is a caveat: such a functional aspect/advantage (of any nucleobase) emerges and becomes apparent only in the presence of the complementary base-pairing partner, and only at the level of an oligomer.^[44] The selection of a given nucleobase (in RNA) is finally based on its function, which is at much higher hierarchical level. When a prebiotic reaction produces the canonical nucleobase or the nucleotide, there is no possibility (teleology) to assess how “useful” the nucleobase is going to be in the future (in the oligomer). Whether orotic acid/orotidine or uracil/uridine would be functionally useful would not be discernible at the rudimentary level of the prebiotic synthesis/formation of the nucleobase or the nucleos(t)ide. Thus, the presence/selection of the nucleobases in RNA should be considered—not just because they are prebiotically available (in this case orotidine and uridine) either through their meteoritic presence or prebiotic synthesis/formation—but also based on their

ability to function or give rise to a system capable of function in the context of chemical evolution. This argument would also extend to the suite of other potentially primordial nucleobases (Figure 6) that could have been part of an “expanded genetic alphabet”.^[45] For example, purine–purine pairings, which were proposed as forerunners that preceded the modern pyrimidine–purine paradigm,^[29] would result in stronger duplexes.^[30,46] While this would be desirable for enhancing template-mediated replication^[46] by forming stronger template–ligand complexes, it would also lead to possible template-strand inhibition problems.^[43,47] It has been proposed that the present purine–pyrimidine paradigm would have been selected only at the level of the oligomer, based on the ability of the oligomeric system to be sensitive to mismatch discrimination versus pyrimidine–pyrimidine or purine–purine base-pairs.^[44] That the emergence of canonical nucleobases could have been due to multiple selection pressures that operated across chemical and biological evolution at various levels has been systematically discussed by Rios and Tor.^[48] They include, predominantly, the stability of the nucleobases and the nucleosides at each level of transformation among many other factors, such as information storage, information transfer, and duplex-stability.^[48] Such analyses for nucleobases would be valid for all the other prebiotically plausible components (sugars and linkers) of RNA as has been discussed by Eschenmoser^[49] (for the hexose-versus-pentose and pyranose-versus-furanose sugars), by us^[50] (for the pentuloses-versus-pentoses), by Westheimer^[51] (for phosphates versus other possibilities), and by Usher^[52] (for selection of 3′,5′-linkages over 2′,5′-linkages).

This line of thought leads to a paradigm of systems chemistry emergence of RNA that starts from the primordial formation of building blocks all the way to the supramolecular self-assembly and oligomer function.^[44] While the generation of the building blocks and their stability are essential towards the making of any informational/functional system, it is not to be expected that the mere formation or existence of the appropriate building blocks in a “prebiotic clutter”^[53] is enough to guarantee their presence in the final functioning system. There needs to be further selection pressures (such as physical availability, reactivity, stability, selectivity^[54] and functionality) that take place at each emerging level of complexity, and fine-tuning of the ensuing structures depending on the nature of the physicochemical environment (such as solvents, pH and temperature)—all these determine which building blocks get selected (at various levels of selection) and survive in the final system that emerges.^[44,48] Any change in one of these parameters will change what the final structure of the functional system is made of. Such a systems chemistry approach both at the molecular and the supramolecular level would lead to the optimized informational/functional system for a particular physicochemical environment as aptly summarized by Eschenmoser “... that optimization, not maximization, of base-pairing was a determinant of RNA’s selection”.^[49] We note with caution that the arguments presented here are based on the hypothesis of the emergence of a replicating polymer in heterotrophic scenario and, therefore, do not deal with questions relating to the more difficult, and complicated, protometabolic, or biosynthetic^[15] reasons that may also contribute to, or may well be, the justification for the choice of orotic acid and orotidine for a de novo route to the functionally superior uridine/cytidine.

It is important to note that these results do not debate against or for the RNA (or a pre-RNA) world. It just posits that the emergence of RNA (or any other informational/functional

oligomer) is not a given, simply based on the prebiotic availability of its canonical constituents, or solved by plausible prebiotic routes that give rise only to the canonical and extant form of RNA from the primordial beginning. To focus only on the components of extant RNA (ribose, phosphate, and the canonical nucleobases) from the very beginning, would imply that RNA is the only outcome. Rather, considering what has been argued by others^[6] and succinctly stated by Joyce,^[3] “The evolution of RNA is likely to have played an important role in the very early history of life on Earth but it is doubtful that life began with RNA. Consideration of what came before RNA must take into account relevant information from geochemistry, prebiotic chemistry and nucleic acid biochemistry”, and by Rios and Tor,^[48a] “The selection of the native bases did not occur in any one hypothetical period. It is more likely that a continuous process of refinement directed their selection throughout prebiotic and early biotic epochs”, it appears that RNA is a destination and not a destiny.

Conclusion

The results of the base-pairing properties of orotidine are to be viewed within the larger context of the emergence (and evolution) of an informational system at a level at which selections are based on a property (base-pairing) leading to function (information storage and catalytic structures). Based on the results that orotidine-based oligonucleotides have no, or weak, base-pairing properties, it would be reasonable to conclude that the selection/retention of orotic acid (in the extant metabolic pathways) could not have been based on its functional efficiency—in so far as the formation of informational base-pairing RNA is concerned. This, along with results from other prebiotically alternative nucleobases, would imply that the emergence and selection of RNA must have occurred at a higher order (level of function), and not based on its direct synthesis from its plausible prebiotic components and building blocks (formation).

Experimental Section

Description of materials and experimental methods, NMR spectra, and data of isolated compounds and oligonucleotides are given in the Supporting Information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the NASA Astrobiology: Exobiology and Evolutionary Biology Program (Grant NNX12AD62G and NNX15AL30G). We thank Professors Antonio Lazcano and Yitzhak Tor for helpful discussions and feedback on the manuscript. We thank one of the reviewers for suggesting the possibility of a catalytic role for orotidine in the context of a ribozyme.

References

- [1]. a) Robertson MP, Joyce GF, Cold Spring Harbor Perspect. Biol 2012, 4, a003608; b) Bernhardt HS, Biol. Direct 2012, 7, 23. [PubMed: 22793875]
- [2]. Joyce GF, Orgel LE, Cold Spring Harbor Monogr. Ser 2006, 43, 23.

- [3]. Joyce GF, Nature 1989, 338, 217. [PubMed: 2466202]
- [4]. Ferris JP, Joshi PC, Edelson EH, Lawless JG, J. Mol. Evol 1978, 11, 293. [PubMed: 31491]
- [5]. Eschenmoser A, Tetrahedron 2007, 63, 12821.
- [6]. Cafferty BJ, Hud NV, Isr. J. Chem 2015, 55, 891.
- [7]. Voet AB, Schwartz AW, Origins Life 1982, 12, 45.
- [8]. Yamagata Y, Sasaki K, Takaoka O, Sano S, Inomata K, Kanemitsu K, Inoue Y, Matsumoto I, Origins Life Evol. Biospheres 1990, 20, 389.
- [9]. Miyakawa S, Cleaves HJ, Miller SL, Origins Life Evol. Biospheres 2002, 32, 209 and references cited therein.
- [10]. Michelson AM, Drell W, Mitchell HK, Proc. Natl. Acad. Sci. USA 1951, 37, 396. [PubMed: 14853953]
- [11]. Radzicka A, Wolfenden R, Science 1995, 267, 90. [PubMed: 7809611]
- [12]. Hu H, Boone A, Yang W, J. Am. Chem. Soc 2008, 130, 14493. [PubMed: 18839943]
- [13]. Singleton DA, Merrigan SR, Kim BJ, Beak P, Phillips LM, Lee JK, J. Am. Chem. Soc 2000, 122, 3296.
- [14]. Appleby TC, Kinsland C, Begley TP, Ealick SE, Proc. Natl. Acad. Sci. USA 2000, 97, 2005. [PubMed: 10681442]
- [15]. McMurry J, Begley T, The Organic Chemistry of Biological Pathways Roberts, Englewood, Colorado (USA), 2005.
- [16]. Davies DB, Rajani P, Husaini S, J. Chem. Soc. Perkin Trans 2 1985, 279.
- [17]. Hruska FE, J. Am. Chem. Soc 1971, 93, 1795. [PubMed: 5550253]
- [18]. Follmann H, Pfeil R, Witzel H, Eur. J. Biochem 1977, 77, 451. [PubMed: 196855]
- [19]. Guschlbauer W, Fric I, Holy A, Eur. J. Biochem 1972, 31, 1. [PubMed: 4640464]
- [20]. a) Hudson RHE, Wojciechowski F, Can. J. Chem 2005, 83, 1731; b) Hudson RHE, Wojciechowski F, Nucleosides Nucleotides Nucleic Acids 2005, 24, 1123. [PubMed: 16248105]
- [21]. Zhang X, Krishnamurthy R, Angew. Chem. Int. Ed 2009, 48, 8124; Angew. Chem. 2009, 121, 8268.
- [22]. Krishnamurthy R, Acc. Chem. Res 2012, 45, 2035. [PubMed: 22533519]
- [23]. Kim E-K, Krishnamurthy R, Chem. Commun 2015, 51, 5618.
- [24]. Ke SH, Wartell RM, Nucl. Acids Res 1993, 21, 5137. [PubMed: 8255768]
- [25]. Aboul-ela F, Koh D, Tinoco I, Martin FH, Nucl. Acids Res 1985, 13, 4811. [PubMed: 4022774]
- [26]. a) Sanghvi YS, Hoke GD, Freier SM, Zounes MC, Gonzalez C, Cummins L, Sasmor H, Cook PD, Nucl. Acids Res 1993, 21, 3197; [PubMed: 8393563] b) Sochacka E, Czerwinska G, Guenther R, Cain R, Agris PF, Malkiewicz A, Nucleosides Nucleotides Nucleic Acids 2000, 19, 515; [PubMed: 10843489] c) Ziomek K, Kierzek E, Biala E, Kierzek R, Biophys. Chem 2002, 97, 233; [PubMed: 12050012] d) Kittaka A, Kuze T, Sarai A, Takayama H, Tanaka H, Miyasaka T, Inoue J.-i., Ishii S, Heterocycles 2003, 59, 207.
- [27]. a) Fox JJ, Yun N, Wempen I, Biochim. Biophys. Acta 1957, 23, 295; [PubMed: 13412724] b) Hilal R, Zaky ZM, Elroby SAK, J. Mol. Struct 2004, 685, 35; c) Kubica D, Gryff-Keller A, J. Phys. Chem. A 2015, 119, 5832.
- [28]. Engelhart A, Hud NV, Cold Spring Harbor Perspect. Biol 2010, 2, a002196 and references cited in this review.
- [29]. Crick FH, J. Mol. Biol 1968, 38, 367. [PubMed: 4887876]
- [30]. Heuberger BD, Switzer C, ChemBioChem 2008, 9, 2779. [PubMed: 18985646]
- [31]. Callahan MP, Smith KE, Cleaves JH, Ruzicka J, Stern JC, Glavin DP, House CH, Dworkin JP, Proc. Natl. Acad. Sci. USA 2011, 108, 13995. [PubMed: 21836052]
- [32]. a) Mittapalli GK, Reddy KR, Xiong H, Munoz O, Han B, Riccardis FD, Krishnamurthy R, Eschenmoser A, Angew. Chem. Int. Ed 2007, 46, 2470; Angew. Chem. 2007, 119, 2522; b) Mittapalli GK, Osornio YM, Guerrero MA, Reddy KR, Krishnamurthy R, Eschenmoser A, Angew. Chem. Int. Ed 2007, 46, 2478; Angew. Chem. 2007, 119, 2530.
- [33]. Cafferty BJ, Fialho D, Khanam J, Krishnamurthy R, Hud N, Nat. Commun 2016, 7, 11328. [PubMed: 27108699]

- [34]. Orgel LE, Crit. Rev. Biochem. Mol. Biol 2004, 39, 99. [PubMed: 15217990]
- [35]. a) Sanchez RA, Orgel LE, J. Mol. Biol 1970, 47, 531; [PubMed: 5418170] b) Zubay G, Mui T, Origins Life Evol. Biospheres 2001, 31, 87;c) Saewan N, Michael A. Crowe, Helliwell M, Raftery J, Chantrapromma K, John D. Sutherland, Chem. Biodiversity 2005, 2, 66;d) Bean HD, Sheng Y, Collins JP, Anet FAL, Leszczynski J, Hud NV, J. Am. Chem. Soc 2007, 129, 9556. [PubMed: 17629276]
- [36]. Armenta-Medina D, Segovia L, Perez-Rueda E, BMC Genomics 2014, 15, 800. [PubMed: 25230797]
- [37]. Ferris JP, Joshi PC, J. Org. Chem 1979, 44, 2133.
- [38]. Ferris JP, Joshi PC, Science 1978, 201, 361. [PubMed: 208152]
- [39]. Becerra A, Lazcano A, Origins Life Evol. Biospheres 1998, 28, 539.
- [40]. Lazcano A, Miller SL, J. Mol. Evol 1999, 49, 424. [PubMed: 10486000]
- [41]. Lazcano A, Miller SL, Cell 1996, 85, 793. [PubMed: 8681375]
- [42]. a) Powner M, Gerland B, Sutherland JD, Nature 2009, 459, 239; [PubMed: 19444213] b) Becker S, Thoams I, Deutsch A, Gehrke T, Meyer P, Zipse H, Carell T, Science 2016, 352, 833. [PubMed: 27174989]
- [43]. Trevino SG, Zhang N, Elenko MP, Luptak A, Szostak JW, Proc. Natl. Acad. Sci. USA 2011, 108, 13492. [PubMed: 21825162]
- [44]. Krishnamurthy R, Isr. J. Chem 2015, 55, 837.
- [45]. "Did the RNA world Exploit an Expanded Genetic Alphabet": Benner SA, Burgstaller P, Battersby TR, Jurczyk S, The RNA World, 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (USA), 1999.
- [46]. Kozlov IA, Orgel LE, Helv. Chim. Acta 1999, 82, 1799. [PubMed: 11543571]
- [47]. a) Grossmann TN, Strohbach A, Steitz O, ChemBioChem 2008, 9, 2185, [PubMed: 18752239] b) Fernando C, Von Kiedrowski G, Szathmary E, J. Mol. Evol 2007, 64, 572. [PubMed: 17437149]
- [48]. a) Rios AC, Tor Y, Astrobiology 2012, 12, 884, [PubMed: 22984873] b) Rios AC, Tor Y, Isr. J. Chem 2013, 53, 469. [PubMed: 25284884]
- [49]. Eschenmoser A, Science 1999, 284, 2118. [PubMed: 10381870]
- [50]. Stoop MA, Meher G, Karri P, Krishnamurthy R, Chem. Eur. J 2013, 19, 15336. [PubMed: 24150882]
- [51]. Westheimer F, Science 1987, 235, 1173. [PubMed: 2434996]
- [52]. Usher D, Science 1977, 196, 311. [PubMed: 583625]
- [53]. Joyce GF, Nature 2002, 418, 214. [PubMed: 12110897]
- [54]. a) Islam S, Bucar D-K, Powner MW, Nat. Chem 2017, 9, 584;b) Islam S, Powner MW, Chem 2017, 2, 470.

Figure 1.

De novo biosynthetic pathway for the synthesis of pyrimidine nucleotides through the direct nucleosidation of orotic acid with 5-phosphoribosyl-diphosphate.

Figure 2. Previous studies dealing with base-pairing properties of orotic acid derivatives attached to oligoamide and oligopeptide backbones.^[20,21]

Figure 3.

Thermal melts documenting the destabilizing effect of single orotidine (O) insert into a self-complementary-RNA 5'-AUAUAUAUAUAUAUAU-3' duplex, equaling that of the purine-purine 5'-AUAUAUAAAUAUAUAU-3' mismatch (top) and destabilizing effect of single orotidine (O) insert into a non-self-complementary-RNA 5'-AAAAUUUAUAUAUAUA-3' + 3'-UUUUAAAUAUAUAUAU-5' duplex, equaling that of the purine-purine 5'-AAAAUUUAUAUAUAUA-3' + 3'-UUUUAAAUAUAUAUAU-5' mismatch (bottom). For conditions see footnote of Table 1.

Figure 4.

Thermal melts documenting the destabilizing effect of single orotidine (O) insert into a self-complementary-RNA 5'-CGAAUUCG-3' duplex, equaling that of the purine-purine 5'-CGAAAUCG-3' mismatch. For conditions see footnote of Table 1.

Hydrogen bonds similar to Watson-Crick pairing mode but leading to steric clash of carboxylate (amide) with phosphate backbone

anti-conformation

Figure 5.

The *syn*-conformation of orotic acid (over the *anti*-conformation) within a duplex would destabilize not only by absence of base-pairing, but also by disturbing the base-stacks above and below. Circles represent 3',5'-phosphodiester links.

Figure 6.

Extant nature uses a small subset of canonical/biological nucleobases (within the circle) out of a library of prebiotically plausible nucleobases.

Scheme 1.

Synthesis of orotidine methyl ester phosphoramidite 3; a) di-*tert*-butylsilyl ditriflate, DMF, 0°C, 1 h; b) *tert*-butyldimethylsilyl chloride, imidazole, 60°C, overnight; c) HF-pyridine, DCM, 0°C, 3 h; d) DMTrCl, pyridine, 4°C, overnight; e) 2-cyanoethyl-diisopropylchlorophosphoramidite, DIPEA, DCM, 0°C to RT, overnight.

Table 1.

Comparison of thermal stability of RNA duplexes containing orotidine with unmodified RNA duplexes (as determined by UV- T_m values).

Entry	Sequences ^[a]	T_m [°C]	[b]	Entry	Sequences ^[a]	T_m [°C]	[b]
1	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	51.9 53.0 ^[c] (51.4) ^[d]	0	15	5'-AAAAUUUUUUUUUU-3' 3'-UUUUUUUUUUUUUU-5'	54.0	0
2	5'-AAUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	47.7 (45.0) ^[d]	-4.2	16	5'-AAAAUUUUUUUUUU-3' 3'-UUUUUUUUUUUUUU-5'	36.7	-17.3
3	5'-AUAAUAAUAAUAAU-3' 3'-AAUAAUAAUAAUAA-5'	47.8 (47.3) ^[d]	-4.1	17	5'-AAAAUUUUUUUUUU-3' 3'-UUUUUUUUUUUUUU-5'	15.7	-38.3
4	5'-AAUAAUAAUAAUAAU-3' 3'-AAUAAUAAUAAUAA-5'	45.0	-6.9	18	5'-AAAAUUUUUUUUUU-3' 3'-UUUUUUUUUUUUUU-5'	38.6	-15.4
5	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	44.7	-7.2	19	5'-AAAAUUUUUUUUUU-3' 3'-UUUUUUUUUUUUUU-5'	23.1	-30.9
6	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	38.0	-13.9	20	5'-AAAAUUUUUUUUUU-3' 3'-UUUUUUUUUUUUUU-5'	40.0	-14
7	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	no T_m (no T_m) ^[d]	-	21	5'-AAAAUUUUUUUUUU-3' 3'-UUUUUUUUUUUUUU-5'	27.0	-27
8	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	39.3	-12.6	22	5'-CGAAUUCG-3' 3'-GCUUUAAGC-5'	41.1	0
9	5'-AUAAUAAUAAUAAU-3' 3'-AAUAAUAAUAAUAA-5'	50.7	-1.2	23	5'-CGAAUUCG-3' 3'-GCUUUAAGC-5'	4.3 7.7 ^[c] (4.0) ^[d]	-36.8
10	5'-UUAAUAAUAAUAAU-3' 3'-UAAUAAUAAUAAU-5'	50.7	-1.2	24	5'-CGAAUUCG-3' 3'-GCUUUAAGC-5'	no T_m	-
11	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	42.0	-9.9	25	5'-CGAAUUCG-3' 3'-GCUUUAAGC-5'	2.9	-38.2
12	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	44.7	-7.2	26	5'-CGAAUUCG-3' 3'-GCUUUAAGC-5'	no T_m	
13	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	6.3	-45.6	27	5'-CGAAUUCG-3' 3'-GCUUUAAGC-5'	25.7	-15.4
14	5'-AUAAUAAUAAUAAU-3' 3'-UAUAAUAAUAAUAA-5'	26.0	-25.9	28	5'-CGAAUUCG-3' 3'-GCUUUAAGC-5'	no T_m	
				29	5'-CGAAUUCG-3' 3'-GCUUUAAGC-5'	no T_m	

[a] $4 \mu\text{M}$ RNA duplex melting temperatures measured in 10 mM phosphate buffer ($\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$, pH 7.0), 1M NaCl, 0.1 mM Na_2EDTA (pH 8.0).

[b] Difference in T_m relative to that of the corresponding natural RNA duplex.

[c] $8 \mu\text{M}$ RNA duplex melting temperature measured in 10 mM phosphate buffer ($\text{Na}_2\text{HPO}_4/\text{NaH}_2\text{PO}_4$, pH 7.0), 1 M NaCl, 0.1 mM Na_2EDTA (pH 8.0).

[d] $4 \mu\text{M}$ RNA duplex melting temperatures measured in 100 mM Tris-HCl buffer (pH 7.2), 0.5 M NaCl, 0.5 M KCl, 0.05 M MgCl_2 . **O**=orotyl amide, **Q**=orotyl acid, **A**, **U** signify a mismatch site.