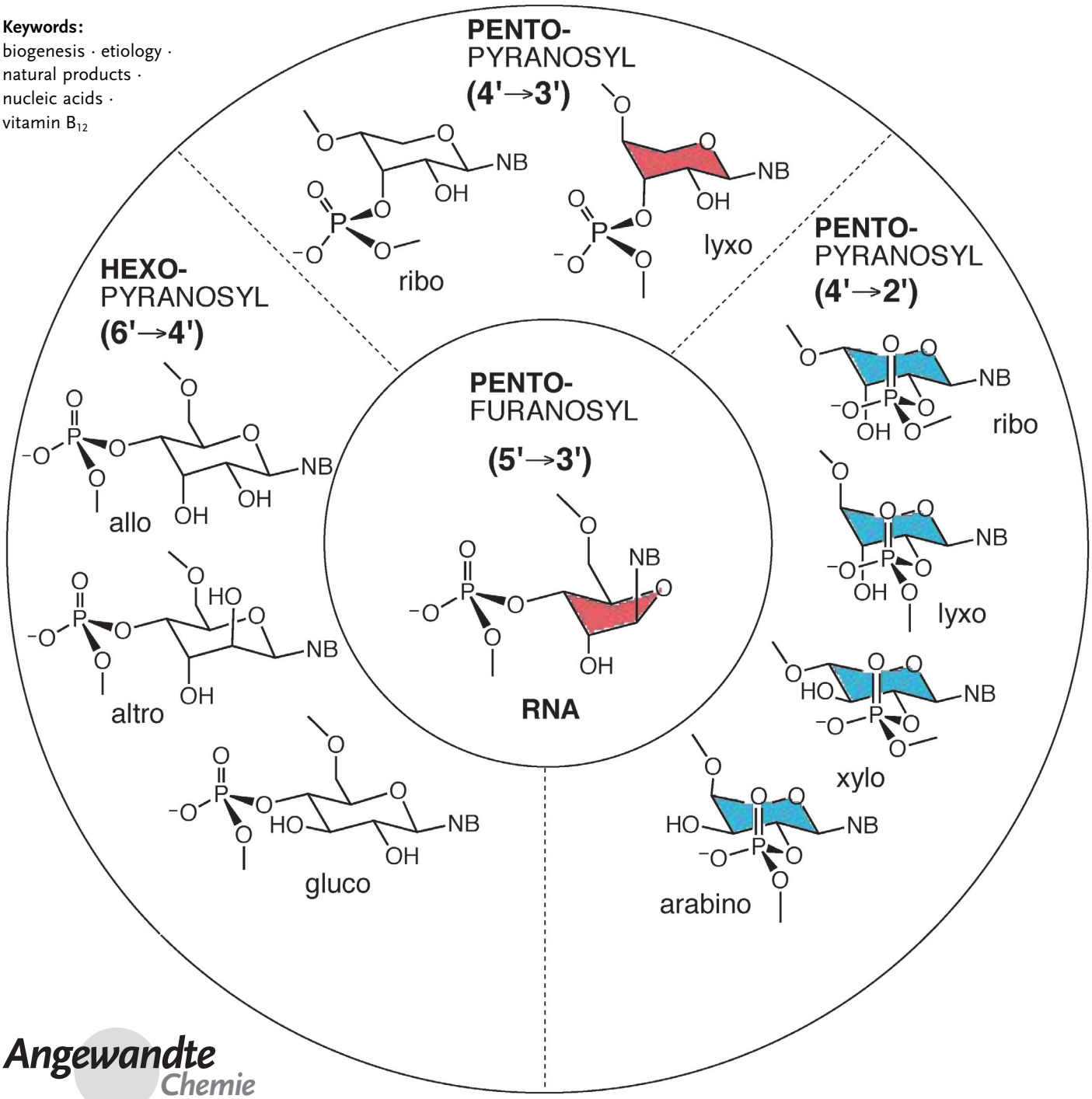
Biogenesis

DOI: 10.1002/anie.201103672

Etiology of Potentially Primordial Biomolecular Structures: From Vitamin B_{12} to the Nucleic Acids and an Inquiry into the Chemistry of Life's Origin: A Retrospective

*Albert Eschenmoser**

Keywords:
biogenesis · etiology · natural products · nucleic acids · vitamin B_{12}



Angewandte 50 International Edition Chemie

"We'll never be able to know" is a truism that leads to resignation with respect to any experimental effort to search for the chemistry of life's origin. But such resignation runs radically counter to the challenge imposed upon chemistry as a natural science. Notwithstanding the prognosis according to which the shortest path to understanding the metamorphosis of the chemical into the biological is by way of experimental modeling of "artificial chemical life", the scientific search for the route nature adopted in creating the life we know will arguably never truly end. It is, after all, part of the search for our own origin.

From the Contents

> "*Natural products are the result of three billion years of development of the living world, and they have survived the natural selection process over a long period of evolution. I am convinced they always carry a message, which it is our job to decipher*"
>
> Vlado Prelog, "Thoughts after 118 Semesters of Studying Chemistry"[1]

1. Introduction

According to Webster, *etiology* refers to "*the science of causes or origins*".[2] In the context of cultural history one speaks of etiological myths, sagas, and legends; of culture-specific traditions regarding the origin of a culture, of life, and of being. In the natural sciences, the term etiology has essentially been appropriated by medicine, where it means the evaluation of the causes and sources of diseases. But its potential significance for the natural sciences is much broader. For example, in natural product chemistry one might inquire not only into the isolation, structure, synthesis, biosynthesis, biological function, and medicinal significance of a substance, but also into its *origin*, raising questions about the roots of a natural product's chemical structure in the sense of the chemical–biological *origin* of its biosynthesis. In contrast to questions of the first type, answers to those about origin will always remain attended by uncertainty; indeed, precisely such uncertainty goes hand in hand with proper application of the term *etiology*.[3] But despite the intrinsic uncertainty in answers, there remains the importance and significance of the questions posed.

Ever since Charles Darwin at the latest, questions about origins have been among the central objectives in biology. In chemistry—an emphatically pragmatic and traditionally "non-historical" science—inquiry into the origin of a natural product would seem unusual. However, with the steadily progressing merger of parts of chemistry with biology, this will change. Today, molecular biology may make it possible to arrive at partial answers to questions of how a natural product's molecular structure has been spawned and selected by biological evolution at the molecular level. Moreover, in the field of natural product chemistry there are many examples where purely chemical insight enables organic chemists to predict overall pathways of biosynthesis for entire families of natural products, and to translate such conjectures into especially successful biomimetic designs for chemical syntheses of corresponding products. In such cases, it is quite often possible to rationalize both the origin and the very existence of these natural products and their biosyntheses in terms of purely chemical criteria. Such rationalizations can amount to insight into the etiology of the natural structures.

The quest for the etiology of a biomolecular structure, or a type of structure,[4] acquires special significance when the questions asked refer to biomolecules the existence and functioning of which are fundamental to life. This applies in particular to those biomolecules for which one may surmise, on the basis of convergent chemical and biological clues, that their origin is connected with the *origin* of life, with biogenesis. Questions and sources of answers referring to the provenance of such molecules lie in the realm of chemical rather than biological research, although the extent to which this is true actually constitutes part of the problem. The etiology of potentially primordial biomolecular structures relates to a period in chemical–biological evolution linking chemical systems and processes of a prebiotic era with the appearance of what biologists regard as "the last common precursor" of all biological organisms. It is this critical phase in the emergence of life about whose location, duration, and chemistry we in fact know nothing. Casting light by experimental chemical means into this darkness must ultimately be the goal of the research on the etiology of biomolecules. Use of the word "primordial" in this context is intended to signify affiliation with precisely the critical phase at issue.

The etiology of potentially primordial biomolecular structures does not of course describe facts, but rather ideas. These are conceptually and experimentally supported notions regarding both chemical and biological factors and conditions that were decisive in the genesis and biological selection of a particular structure (or structural type). Such a definition would need to refer also to that critical phase in biogenesis in which self-selection on a chemical level (through autocatalysis) may have preceded biological selection.

[*] Prof. A. Eschenmoser
Organisch-chemisches Laboratorium der ETH Zürich, Hönggerberg
Wolfgang-Pauli-Str. 10, CHI H309, CH-8093 Zürich (Switzerland)

Preconditions to any type of selection are a molecular structure's formation as well as its function; it is to these two aspects that a structure's etiology will relate. As objectives of such research, the two aspects differ not least in the fact that a search focusing on a structure's formation inevitably carries with it the handicap of "historical research", whereas a search aiming at a structure's function can be chosen to be, for example, an investigation into time-invariant relationships between chemical structure and biological function. In any inquiry into the etiology of a primordial biomolecular structure, the two routes to the target will ultimately have to merge, and a true answer will require synthesis of the results from both.

2. Regarding Prebiotic Chemistry

"*Par ma foi, il y a plus de quarante ans que je dis de la prose, sans que j'en susse rien*"[6]

Stanley Miller (1930–2007), the originator of prebiotic chemistry,[5] might well have cited this well-known dictum from Molière's "Le Bourgeois Gentilhomme" had he been asked about use of the term "etiology" in the context of the origin of potentially primordial natural products. He could have pointed out that it is "prebiotic chemistry" which, since the time of his and Juan Oro's historic experiments,[5a,b] has been concerned experimentally with questions regarding the source of life's basic building blocks and the search for the chemistry of life's origin. The results so far achieved in prebiotic chemistry of the Miller–Oro type have shaped the scientific view regarding the origin of life to the extent that it is now generally assumed that central chemical building blocks of life, α-amino acids, sugars, and nucleobases, were natural components of primordial geochemical environments on our earth (or elsewhere).[7] Representatives of these families of substances are found to be formed—albeit only in the slightest amounts, and always together with complex mixtures of abiological organic materials—under a wide variety of experimental conditions in which excess energy pummels basic carbon-, nitrogen- and hydrogen-containing material, with subsequent action of water. At least some of these conditions are assumed to model conditions of primordial environments. As to whether such experiments are relevant to our conceptions of the origin of life, today—in distinct contrast to the early days of prebiotic chemistry—there no longer is any general sense of agreement, and quite probably rightly so. The discord lies in the contradictory premises of rival postulates of a *heterotrophic*[8,9] versus an *autotrophic*[10] biogenesis. From the former viewpoint, organic material that had collected on earth over eons provided the substrate of a self-organization process which, assisted by energy from environmental sources, led ultimately to life. From the point of view of an autotrophic beginning, however, such organic material of terrestrial and extraterrestrial origin as may have accumulated on the earth was essentially irrelevant to the decisive process of biogenesis. Driven by the free energy of a small number of basic chemical starting materials and natural reducing agents from local sources, that decisive process synthesized from the outset its own organic building blocks and intermediates through selective chemical reactions assisted by autocatalysis.

No matter which of these pictures ultimately mimics reality more closely, the major observations of classical prebiotic chemistry will remain important beyond dispute. Since the basic biological building blocks are found to be generated under the most diverse and most elementary conditions (of supplying excess energy to organic materials), they must not require any specific instructions in order to be formed, and belong to the set of elementary organic compounds that matter is predetermined under such conditions to create. Carbon-containing meteorites are found to include mixtures of organic compounds of extraterrestrial origin strikingly similar to the mixtures formed in experiments in prebiotic chemistry.[11] This speaks for the ubiquity of the environmental conditions under which such determinacy is expressed. The increasingly rich palette of simple organic molecules whose existence in interstellar space has been observed spectroscopically points in the same direction.[12] The question whether determinacy extends beyond the realm of generating simple building blocks of life to the emergence of life itself has been widely affirmed.[13] Nevertheless, any answer to this important question must be regarded as premature so long as we have not shown how life arose or can arise.

Already at an early stage, physical chemists have proposed models, on a purely theoretical level, of how kinetic and thermodynamic principles could have governed a self-organization of matter into life.[14] The potential of matter to undergo self-organization is a consequence of the second law of thermodynamics and of the properties of matter at the molecular level. Prerequisite for this potential to become reality is a chemical environment held in a non-equilibrium state by kinetic barriers. Driven by its free energy, such an environment can develop the propensity to generate a dynamic chemical system of constituents and catalysts, one that circumvents those kinetic barriers and implicitly catalyzes the environment's approach to equilibrium, at the same time escaping its own thermodynamic instability through autocatalytic replication. The emergence of such a process may be said to correspond to a kinetic analogue of the Le Chatelier principle.[15] In one type of conceivable scenario, constituents and catalysts in a system of this sort are entangled with an autocatalytically self-replicating entity in such a way that replication of that entity indirectly implies the replication of constituents and catalysts. This defines a system's identity, autonomy, and self-preservation with reference to a specific environment. When such a system is capable of contingent diversification of its catalytic functions, it acquires the potential for adapting to changes in the environment, hence the potential to evolve.

Prebiotic chemistry may have been quite successful in demonstrating the formation of biomolecular building blocks under potentially primordial conditions, but it is still, so to speak, "light-years" away both conceptually and experimentally from the goal of being able to model a potentially primordial self-organization process starting from those natural building blocks. It is not least for this reason that a research field recently emerged from within chemistry that

calls itself "systems chemistry",[16] an expression of the insight that future chemical research on biogenesis must deal above all with function, not simply with the constitutional chemistry of potential molecular players.

Chemical research on biogenesis will certainly divide itself thematically into two branches, quite analogously to what happened to organic chemistry from its earliest days. Installed as the branch of chemistry intended to study chemical compounds of biological origin, it quickly went beyond that task to deal also with "artificial" carbon compounds created by organic chemists themselves. In an equivalent way, the search for the chemistry of life's origin will find ways to free itself from a limitation to natural biochemical building blocks and pursue the problem instead with pragmatically selected, artificial chemical systems. Experimental modeling of "artificial chemical life" will prove the shortest path to an understanding of what "chemical life" constitutes at its most elementary stage, and in what ways such life can emerge from inanimate matter. What constitutes life at the lowest level of complexity is of course a question posed also by biologists; creation of a so-called "minimal cell" has become one of the goals of a new field called synthetic biology. The distinction between the efforts of chemists as opposed to those of biologists in pursuing the common goal is reminiscent of the distinction organic chemists make when they call attention to the difference between a "total synthesis" and a "partial synthesis" of a natural product. To maintain this distinction only so long as not all the ingredients of the biologists' "minimal cell" would have been prepared in a totally synthetic way would be too superficial, however. John Desmond Bernal's pronouncement—"*If life once made itself, it must not be too difficult to make it again*"[9b]—provocative in its positivistic radicalism and crass exaggeration, but nevertheless unerring at its core, is directed ultimately toward chemistry, not biology.

3. Toward an Etiology of the Vitamin B_{12} Structure

About two decades ago, a paper published in this journal entitled "Vitamin B_{12}: Experiments Concerning the Origin of its Molecular Structure" described how work on the chemical synthesis of vitamin B_{12} (Figure 1) had eventually steered us toward an inquiry into the origin of the vitamin's molecular structure.[17] Coenzyme B_{12}, a cofactor derived from the vitamin (Figure 1), is still the most structurally complex cofactor known. Asking questions about the origin of the B_{12} structure may well have gone beyond standard practice in natural products chemistry at that time; in retrospect it appears worthwhile to note, however, that any such transgression can be seen as having justified itself by the results of experiments to which it led.[17] Since vitamin B_{12} is biosynthesized only by microorganisms, among them in particular the most primitive ones, it has long been suspected that it represents an "archaic" biomolecular structure. However, it was not this aspect, but rather a specific chemical experience encountered in the course of our work on the chemical synthesis of this vitamin that primarily inspired us to inquire into the structure's origin. That experience was an astonishing degree of success in the photochemical A/D precorrin-to-corrin cycloisomerization, the central step in the photochemical A/D variant of the B_{12} synthesis (Figure 2).[18–21] Here we witnessed a mechanistically complex reaction process, induced by visible light, in which a hydrogen atom from the methylene group in ring D of a precorrinoid cadmium(II) complex jumped to the methylidene carbon atom at ring A, concomitantly closing the ligand system to a macrocyclic corrin ring through bond formation between rings A and D. The methylidene carbon in question from the precorrinoid ligand is constitutionally no less than 16 atoms away from the ring D methylene group, yet the template effect of the metal ion places that carbon in the immediate vicinity of the ring-D

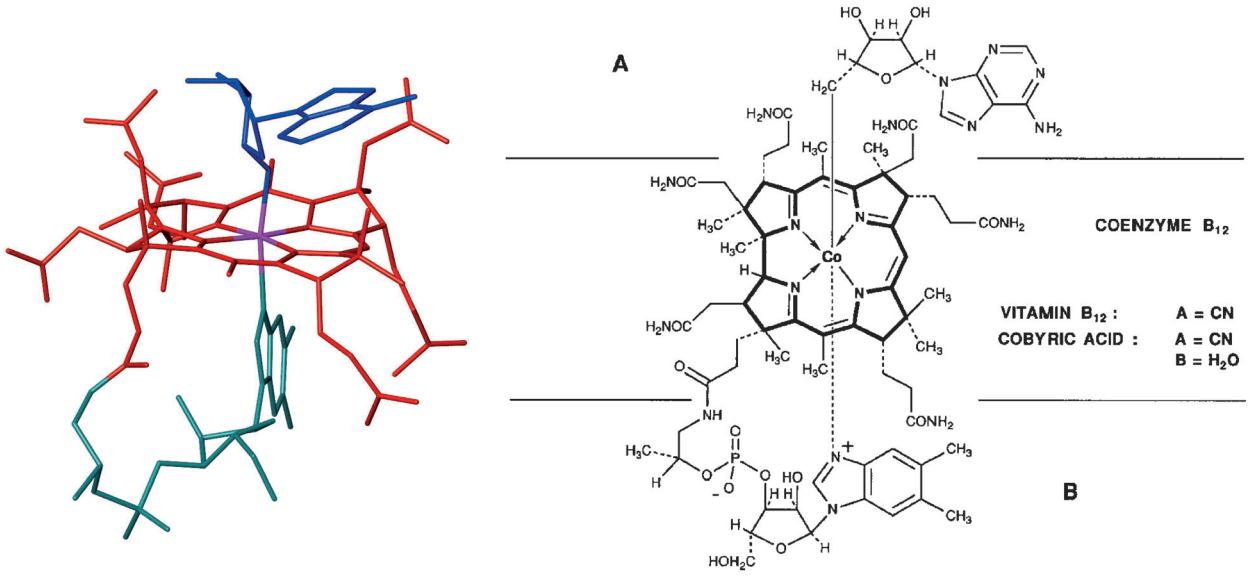


Figure 1. Chemical structure of coenzyme B_{12}: **A** = Ligand above, **B** = ligand below the corrin chromophore. Methylcobalamine: CH_3 as ligand in place of **A**. Vitamin B_{12}: CN as ligand in place of **A**. Cobyric acid: CN as ligand in place of **A**, H_2O as ligand in place of **B**. (X-ray structural image provided by Bernhard Kräutler, Innsbruck).

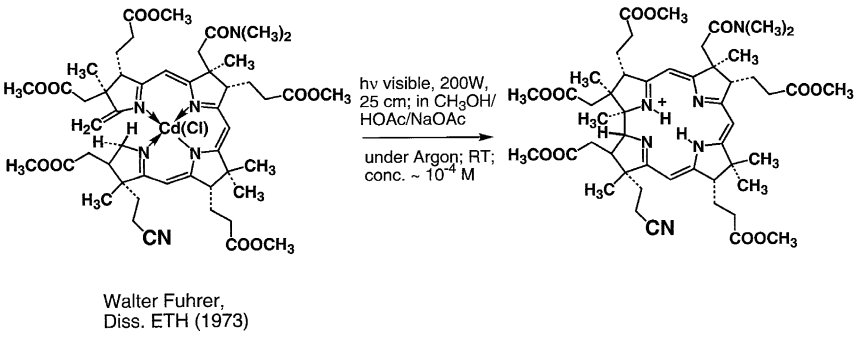


Figure 2. A/D precorrin–corrin cycloisomerization step in the photochemical variant of the vitamin B_{12} synthesis.[21] Designations for the peripheral 5-membered rings, proceeding clockwise: ring A (upper left), ring B, ring C, ring D (lower left).

methylene. The photochemical process amounts to a highly diastereoselective, antarafacially-proceeding, electrocyclic 1,15 macroring closure, establishing, through stereoelectronic as well as steric reaction control, the natural *trans* junction between rings A and D.[19–21]

It is worth pausing here for a moment to take note of one special aspect of this amazing reaction, one that, after completion of the synthetic work, was subjected to comprehensive scrutiny through model complexes (Figure 3).[22–24]

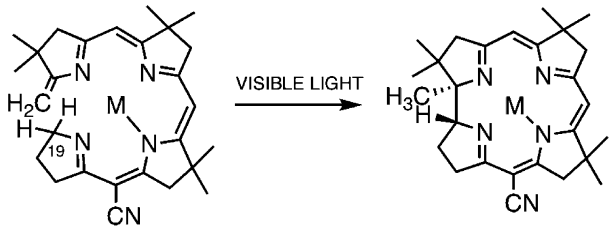


Figure 3. Characteristics of the photoinduced A/D precorrin–corrin cycloisomerization, as established through model complexes.[22a,b] 1) The reaction proceeds with M = Li, Mg(Cl), Zn(Cl), Cd(Cl), Pd^{+}, Pt^{+} but NOT with M = H, Cu^{+}, Ni^{+}, $Co(CN)_2$, Mn(Cl). 2) The 19,19-dideutero-Pd^{II} complex cyclizes slower than the protium analogue. One deuterium atom migrates to the methylidene carbon. 3) Oxygen and specific triplet quenchers thwart the cyclization of the zinc and cadmium complex. 4) The cyclization of the zinc and cadmium complex is strongly sensitized by the cyclization product (and by the synthetic precursor of the educt). [R. Neier, thesis ETH, No. 6178 (1978); *Chem. Soc. Rev.* **1976**, *5*, 377.]

Studies on a Cd^{II} complex showed unambiguously that the A/D-cycloisomerization proceeds autocatalytically. It turned out that the excited state of the educt chromophore required for initiating the process is attained efficiently only via sensitization, with the photoactivated product chromophore in its excited triplet state acting as the sensitizer. The first excited state of the (fluorescing) product chromophore apparently lives long enough to populate its triplet state, in contrast to the first excited state of the (non-fluorescing) educt chromophore (Figure 4).[22–24] In line with the direction of the sensitization, light absorption in the educt chromophore occurs at somewhat longer wavelength than that in the product chromophore due to the presence of the methylidene group in the former. This group is responsible for the inability of the educt chromophore to reach the reacting triplet state directly from its first photoexcited state, presumably due to a radiationless deactivation of the latter that is too rapid.[22a,23]

The remarkable constitutional efficiency of the A/D ring-closure reaction had led us to suspect that the (then still largely unknown) biosynthesis of vitamin B_{12} must also take advantage of this extraordinary formation potential of the critical A/D-region of the vitamin's structure.

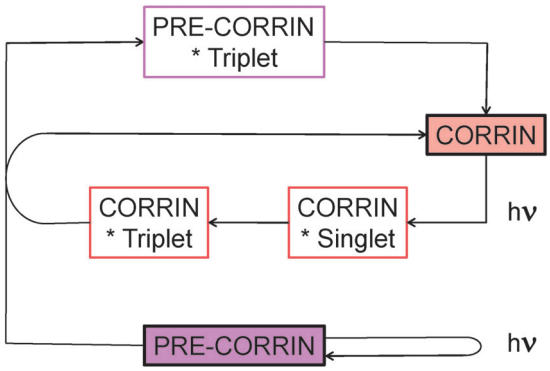


Figure 4. Autocatalytic course of the photoinduced A/D precorrin–corrin cycloisomerization in model complexes.[22a]

Such thoughts set in motion a comprehensive search, guided by mechanistic considerations, for potentially biocompatible (non-photochemical) variants of the A/D precorrin-to-corrin cycloisomerization. In fact, an entire palette of such *non*-photochemical modes of formation for the corrin macrocycle was found, each mechanistically related to the original photoinduced reaction (Figure 5).[25,26] What all this ultimately meant was that the prevailing notion at the outset of the B_{12} synthesis project—that the most difficult challenge to overcome in B_{12} synthesis would be the molecular region involving the junction between the peripheral rings A and D[19a,b,20b,c,27]—was transformed into quite the opposite. This new way of looking at the B_{12} structure marked the beginning of still another research project, one extending over many years and numerous dissertations.[17] In retrospect, it presents itself as a systematic attempt to gain insight by experimental chemical means into the *etiology of the vitamin B_{12} structure.* Therefore, the results of that project are summarized here once again, together with comments derived from a present-day perspective.

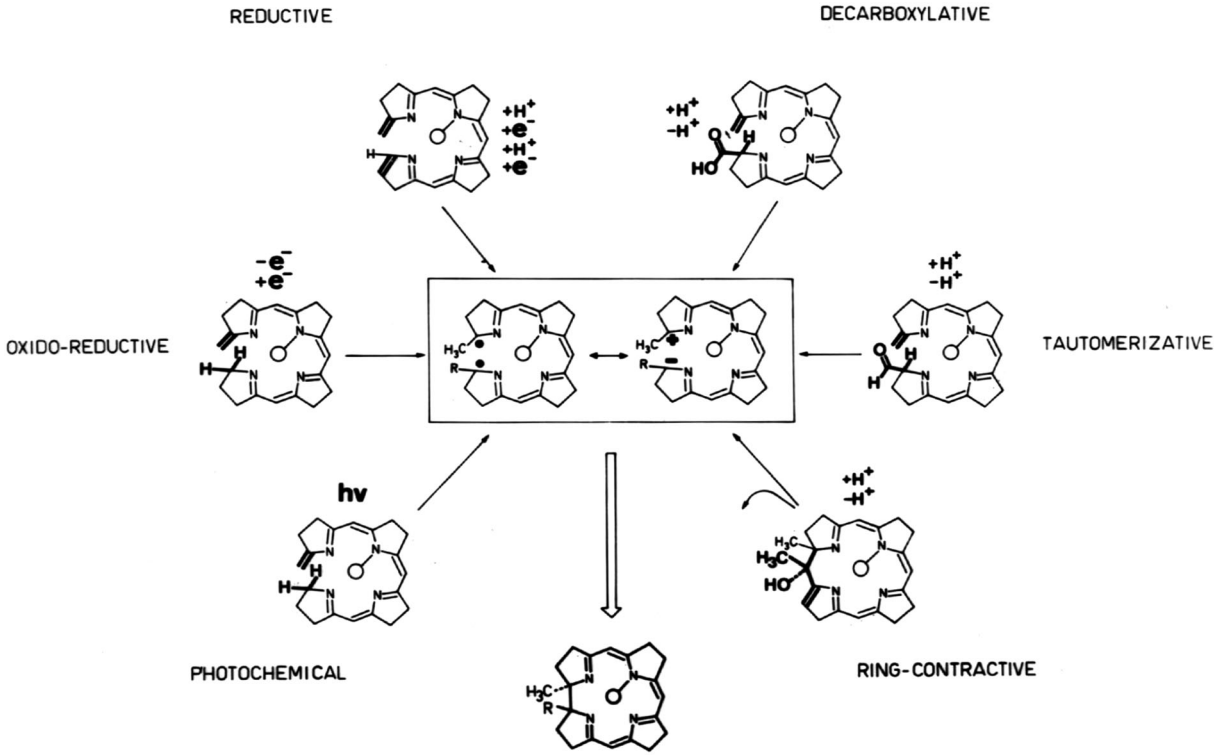
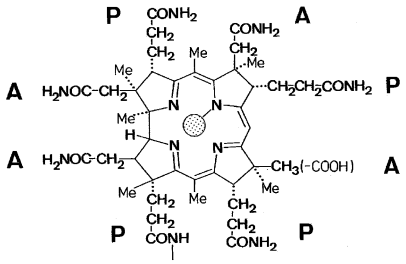


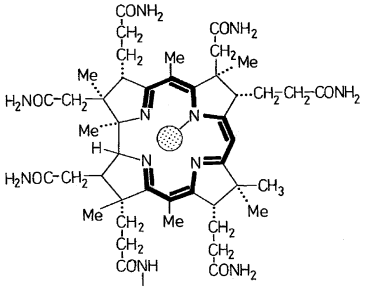
Figure 5. Palette of experimentally realized synthetic approaches to the ligand system of corrin.[17]

3.1. Thermodynamic Stability of Structural Elements of the Vitamin B_{12} Molecule as an Etiological Criterion



The Arrangement of Side Chains on the Peripheral Rings: The specific arrangement of the acetic acid and propionic acid amide chains on the periphery of the B_{12} molecular structure conforms to the arrangement of the corresponding side chains in the biosynthetic precursor of all natural porphinoids, uroporphyrinogen type III. It has long been recognized[28] that this particular type of constitution, in an equilibrium among the four constitutionally isomeric uroporphyrinogens types I–IV, is favored for entropic reasons (Figure 6). It is this isomer that in condensation reactions of monocyclic precursors (under thermodynamic control) is formed as the major product, and to the statistically predicted extent.[28,29]

The arrangement of side chains on the periphery of the vitamin B_{12} structure corresponds to the arrangement found in the thermodynamically most stable of the four uroporphyrinogen isomers.



The Corrin Ligand Chromophores: Subjecting model compounds of hexahydroporphyrin structures of the uroporphyrinogen type to complexation reactions with Mg^{II} or Zn^{II} salts in lipophilic medium leads to neutral metal complexes of *pyrrocorphin* ligand systems or, depending on the reaction conditions, to corresponding positively charged *corphin* complexes (Figure 7).[30] Both the pyrrocorphin and corphin ligands are tautomers of tetrapyrrolic porphyrinogens, but in contrast to the latter they contain highly conjugated chromophore systems that are constitutionally closely related to the chromophore system of the corrin ligand. If one demetallates such corphinoid complexes of Mg^{II} or Zn^{II}, and subjects the free pyrrocorphin ligands to tautomerizing conditions, the original porphyrinogen is formed once again (Figure 7).[30,31]

In metal complexes—but not as the free ligands—tautomers that contain a corrinoid chromophore lie in a thermodynamic trough within the equilibrium among hexahydroporphinoid tautomers.

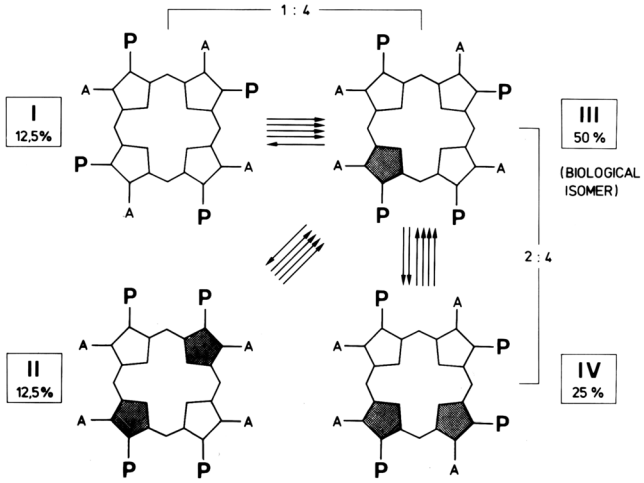


Figure 6. Uroporphyrinogens I–IV (A = **A**cetic acid side-chain, P = **P**ropionic acid side-chain). The direction and number of reaction arrows shown for possible isomerization steps at each of the four peripheral rings for each isomer is a statistical representation of positions of equilibrium among the four uroporphyrinogen isomers.

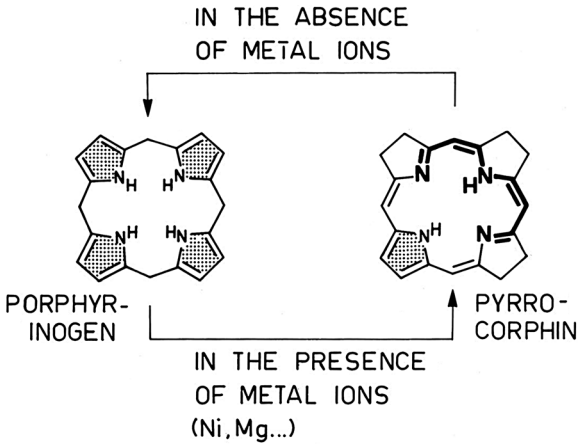
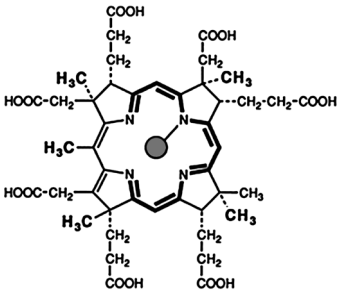


Figure 7. The porphyrinogen forms of hexahydroporphyrins isomerize to pyrrocorphin forms under conditions of complexation with metal ions and inert reaction conditions, whereas free pyrrocorphin ligands are converted under tautomerization conditions into the corresponding porphyrinogens.[30]



Size of the Macro Ring in Corrin Ligands: The corrin ligand system differs from a corphin ligand system in its oxidation level as well as in constitutional details within the chromophore, yet above all in the size—15-membered versus 16-membered—of the internal macro ring.[32] Comparative X-ray structure determinations of synthetic square-planar nickel(II) complexes of corphinoid and corresponding corrinoid ligands reveal a conformational deformation of the ligand system in the case of the former, but not the latter ligand (Figure 8). The type of deformation observed suggests that the corphin complex is strained, i.e., that the coordination space of the corphin ligand is too large for the spatial coordination demands of a Ni^{II} ion, whereas the coordination space of the corrin ligand seems just to fit.[33] This notion is strongly supported by the observation that both nickel(II) and cobalt(III) complexes of a synthetic 20-methyl-20-hydroxydihydrocorphin undergo irreversible thermal rearrangement to corresponding 19-acetylcorrin complexes (Figure 9).[26]

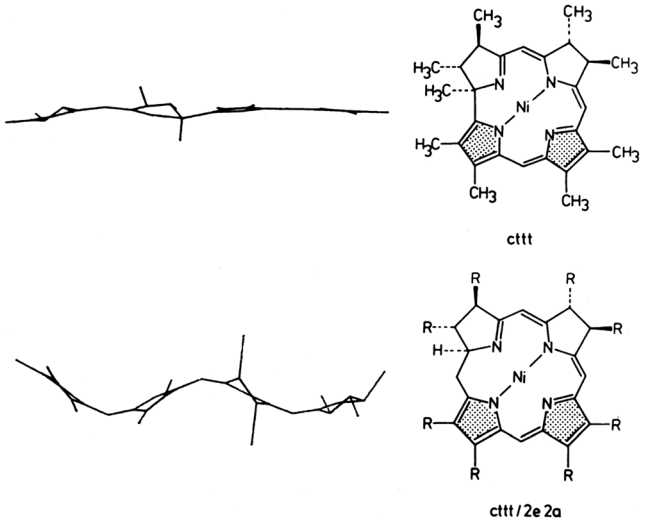
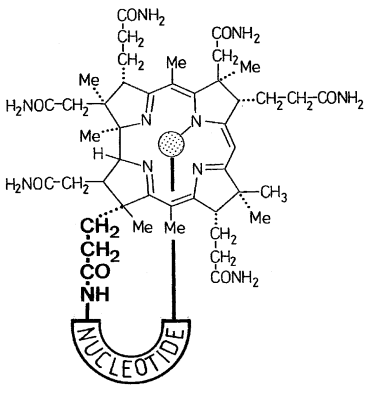


Figure 8. X-ray structures of Ni^{II} complexes of a corrinoid and a corphinoid ligand.[33] With a metal ion like Ni^{II} the dimensions of the coordination space for a corrin ligand appear to be just right, whereas those of a corphinoid ligand are too large.

Ni^{II} and Co^{III} corrin complexes are thermodynamically more stable than corresponding corphin complexes.



Attachment of the 19-Membered Nucleotide Loop to the Corrin Ligand: Attachment of the nucleotide chain on the "underside" of the vitamin B_{12} molecule to the propionic acid side chain of ring D and to cobalt via a 19-membered ring is one of four constitutionally analogous possibilities. In principle, on the "underside" of the molecule the nucleotide chain

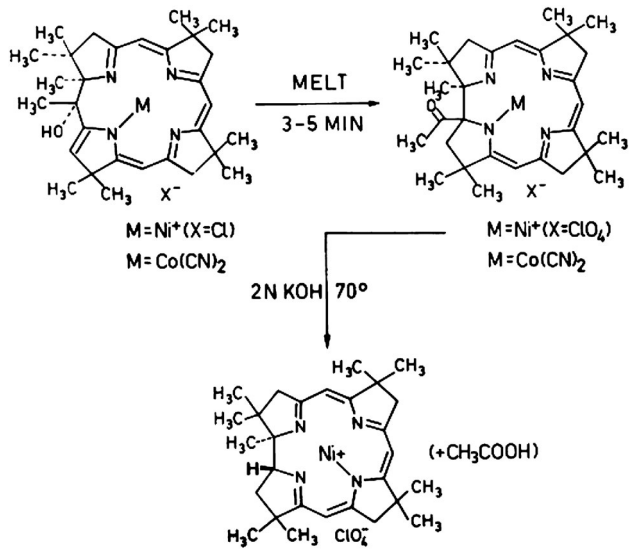


Figure 9. Thermal conversion of Ni^{II} and Co^{III} complexes of a 20-methyl-20-hydroxydihydrocorphin into the corresponding 19-acetylcorrin complexes, and hydrolytic deacetylation of the Ni^{II}-19-acetylcorrin complex to the Ni^{II}–corrin complex.[26]

might also be affixed to cobalt and to one of the propionic acid chains at rings A, B, or C. In a crucial experiment, the following was observed: when all seven carboxyl functions of cobyrinic acid (the heptacarboxylic acid corresponding to vitamin B_{12}) were activated in identical fashion, with the activation judiciously chosen such that it is mild enough to allow under suitably mild conditions intra- but not intermolecular amidation reactions to occur, and when the complete nucleotide chain as a free reaction partner was then allowed to act under very mild reaction conditions upon the seven-fold activated cyanoacetoxy Co^{III} complex, the reaction product isolated (after amidation with ammonia) was almost exclusively vitamin B_{12}, along with the hepta-amide of the starting material (Figure 10).[17,34] In a cobalt(III) complex thought to be reversibly formed as the reactive intermediate, the terminal amino group of the nucleotide chain, in an intramolecular attack, selectively chooses the activated carboxyl function of ring D to form the natural 19-membered nucleotide loop. The reaction is irreversible, and must, therefore, proceed under kinetic control; its remarkable regioselectivity is seen as the result of conformational control exerted by the relative thermodynamic stabilities of the four conformations of the nucleotide chain, which correspond to the four possible 19-membered nucleotide loops.[35]

Specific attachment of the vitamin B_{12} molecule's nucleotide chain to the propionic acid function of ring D corresponds to the thermodynamically most stable option among four constitutionally analogous possibilities.

The upshot of the experimental studies summarized above is the following: All four structural elements in the vitamin B_{12} molecule that were targeted in these experiments turn out to be thermodynamically privileged—at least under the experimental conditions employed—relative to structural alternatives that might be envisaged from a chemical point of view. Can such experimental evidence of thermodynamic benefit of portions of a biomolecular structure be etiologically relevant? Or, in more specific terms: was the thermodynamic benefit of the four structural elements decisive for the natural processes of *formation* and biological *selection* of the B_{12} structure?

There should be little doubt that, in general, relative thermodynamic stability can act as a decisive factor in biological selection of a biomolecule. Biochemical "tinkering" as a prerequisite for a biological evolutionary step will follow thermodynamic product stability so long as the differences involved are not masked by diametrically opposed selectivity of a kinetic nature. The chemical nucleotidation

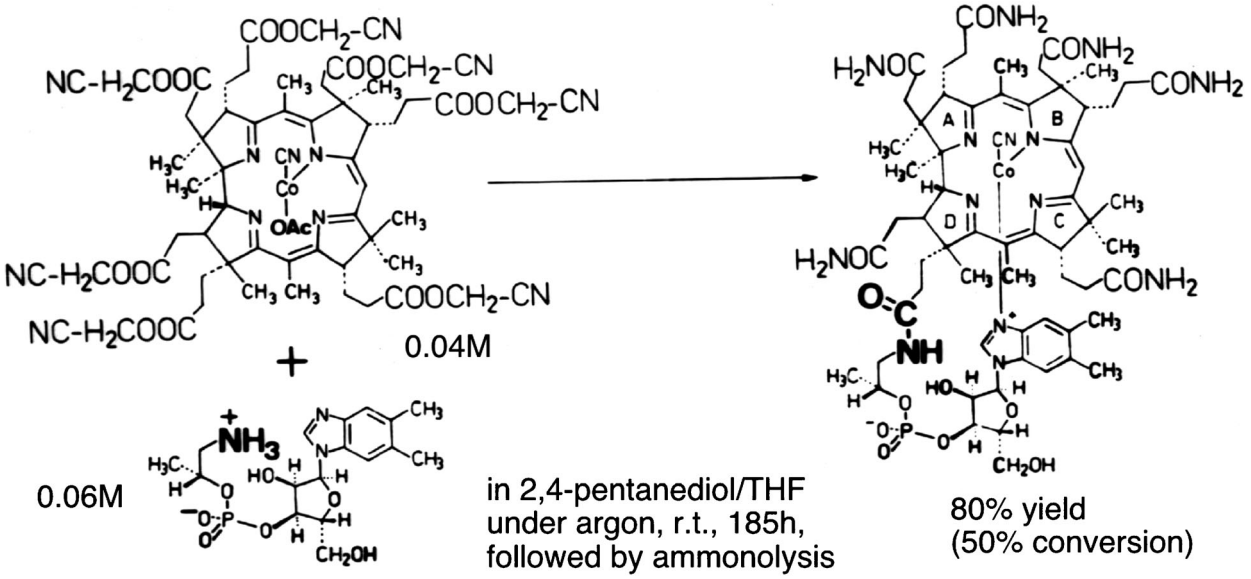


Figure 10. Kinetic preference for coupling of the vitamin B_{12} nucleotide chain with the propionic acid side-chain of ring D, caused by the relative thermodynamic stability of the product.[17,34] Hepta-cyanomethyl-cobyrinate reacts with the free nucleotide component regioselectively at the propionic carboxyl function of ring D without prior differentiation of the reacting ester group. [F. Kreppelt, thesis ETH, No. 9458 (1991); *Angew. Chem.* **1988**, *100*, 5; *Angew. Chem. Int. Ed.* 1988, 27, 5.]

reaction leading selectively to vitamin B_{12} is illustrative of how thermodynamic stability can, in this case by way of "conformational tinkering", override kinetic control and irreversibility as a selectivity factor. Moreover, this particular experiment can even lay claim to a role as a specific *model* for biological selection of the nucleotide chain's position in the B_{12} structure. This is under the assumption that, among the four loop variants, either the most stable one proved to be functionally the best, or else that function adapted itself to this specific isomer because in being the most stable it was also the most abundant. From present knowledge regarding the mechanism of B_{12}-catalyzed reactions, one would expect the position of the "base-on/base-off" equilibrium to have been one of the molecular characteristics related to function that controlled selection of the nucleotide loop from among the four variants.[35]

Another characteristic structural element of the vitamin B_{12} molecule is the crown of methyl groups bound to carbon along the periphery of the ligand system. The stepwise enzymatic incorporation of such methyl groups into the uroporphyrinogen molecule, with adenosylmethionine as methylating agent, makes it look as though the B_{12} biosynthesis[36,37] had to "force" a transition from the tetrapyrrolic form of the macrocycle into the corphinic form, and from there into the corrinic form, without a need to call upon metal ions for help. One of these methyl groups, the one at the meso position between rings A and D, has a special function: in the B_{12} biosynthesis this methyl group, by its very presence, assists in the process of ring contraction from a corphinoid to a corrinoid macro ring, and may even be prerequisite to that process. Moreover, one observation made in experiments with model systems[26b,38,39] can be interpreted to point toward still another function of this methyl group, namely blocking a side reaction likely to occur in its absence: a simple and irreversible process of the type illustrated in Figure 11, a ketonization that is very favorable both kinetically and thermodynamically.[40]

In the course of our systematic studies of the chemistry of model ligand systems with a corphinoid chromophore structure, we also examined C-methylations. What we observed was a remarkable regioselectivity that gives these reactions the appearance of being biomimetic (Figure 12).[41] The question of whether such C-methylations might class among prebiological events must most probably be answered in the negative, however. As simple as C-methylations may appear on the surface, based at least on the current state of knowledge they are not thought to have been available in the reaction palette of prebiotic chemistry.

Cofactors, at least those that in the eyes of chemists have "archaic looking" structures, have often been referred to as representing molecular fossils of primordial metabolic processes, where in the absence of enzymes they would have fulfilled catalytic functions.[17,40,42] With regard to any of these cofactors, the question can be raised as to whether structurally simpler but still functional precursors might have existed. This is a question to be asked most pointedly with respect to that cofactor which is structurally the most complex of all, vitamin B_{12}. One might consider corrinoid complexes, still derived from uroporphyrinogen type III, but with hydrogen atoms at all those carbons on the ligand periphery where the B_{12} molecule carries methyl groups. Chemical reasons for considering such complexes are the direct accessibility of corphinoid metal complexes from porphyrinogens, the remarkable formation of uroporphyrinogens from the monopyrrolic precursor porphobilinogen and, above all, the structure of the natural uroporphyrinogen which corresponds to the statistically most probable among the four possible uroporphyrinogen isomers. This curious fact points with some emphasis toward a pre-enzymatic origin for the entire family of porphinoid natural products.[28]

The acid-catalyzed chemical tetramer(oid)ization of porphobilinogen to a statistically controlled mixture of four uroporphyrinogens[28,29,43] can be counted among the most appealing results observed to date of chemical "natural product synthesis under potentially *prebiotic* conditions". Unfortunately, all attempts to achieve the formally equally attractive step of dimer(oid)ization of aminolevulinic acid to porphobilinogen under comparable conditions have proved disappointing.[44] Even aminolevulinic acid, as simple as it appears from both constitutional and biosynthetic perspectives, remains a puzzle for prebiotic natural product synthesis.[45] We have attempted to conceive, as well as to pursue experimentally, pathways from uroporphinogen derivatives to corresponding corphin complexes, and from there to corrin complexes bearing hydrogens along their periphery rather than methyl groups.[46,47] The first step, porphinogen-to-pyrrocorphinate conversion, proved achievable, albeit with significantly more difficulty than in previous work with uroporphyrinogen models substituted on their periphery with simple alkyl groups. Despite extensive efforts, however, we were unsuccessful in this series at transforming a corphin complex into a corrin complex, irrespective of whether the model did or did not have a methyl group at the meso position between rings A and D.[48]

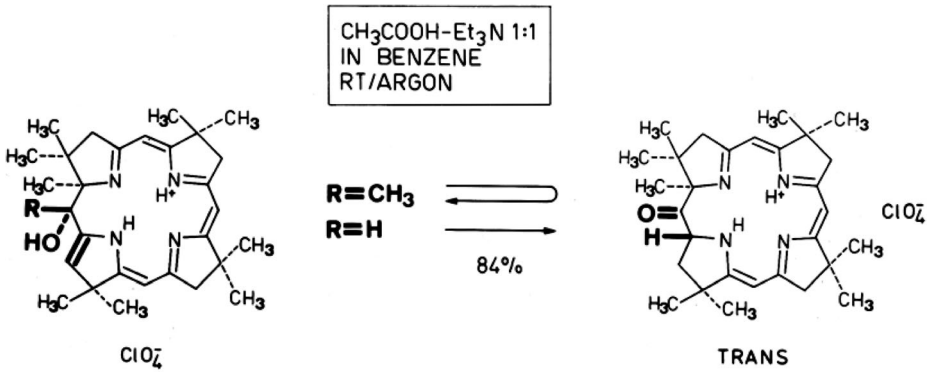


Figure 11. Irreversible transformation into a 20-hydroxydihydrocorphin, which is blocked for a corresponding 20-hydroxy-20-methyldihydrocorphin.[26b]

In research like this, much more than in work on purely

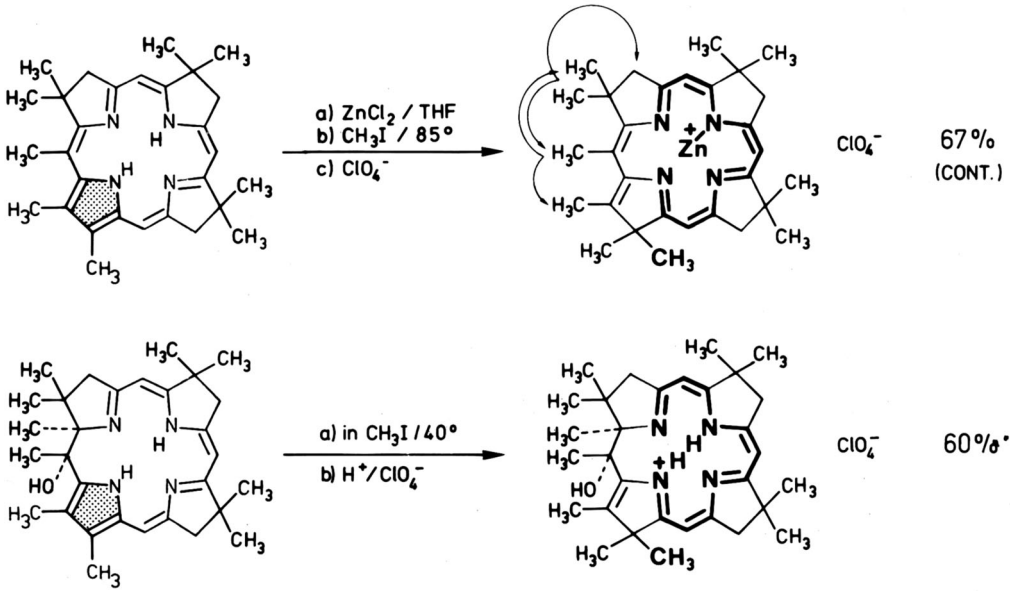


Figure 12. Biomimetic C-methylation at the periphery of pyrrocorphin model ligands.[41]

chemical synthetic projects, a lack of success may mean little, if anything. Nevertheless, our failure appears to correspond to the degree of chemical complexity of the reaction paths in which the biosynthesis of vitamin B_{12} must invest in order to convert its corphinoid intermediates into the natural corrin ligand.[36,37] Our failure also raises the question of to what extent the two acetic acid side chains flanking the meso carbon atom between the A and D rings of the uroporphyrinogen type III structure may be mechanistically required in order for a corrinization process in this series to be successful. This could ultimately relate to the reason why the ligand structure of vitamin B_{12} is contracted specifically between the rings A and D and not at one of the three other meso positions. But what our difficulties have led to foremost is the question of why vitamin B_{12} has the structure of a corrin, and not that of a corresponding corphin, since the latter, after all, lies chemically as well as biosynthetically along the way to the former. Do thermodynamics and a functional preference for a corrin structure as against a corphin point etiologically in the same direction?

3.2. Why a Corrin and Not a Corphin?

Raising this question shifts the focus of inquiry into the etiology of the B_{12} structure from *formation* to *function*: to probing chemical reasons for the superiority (taken as a given) of the corrin ligand over the corphin ligand system for the biological function of B_{12}-coenzymes. The matter comes to a head in the question of what chemical properties in fact distinguish a cobalt corrin complex from a cobalt corphin complex. From the perspective of what our model studies have shown regarding the structural differences between the two ligand systems, the question can be expressed more pointedly: how do corrin and corphin complexes differ in the axial coordination chemistry that develops due to differing dimensions of the coordination spaces these two ligands have to offer a cobalt(III) or nickel(II) ion?

The question alone points already to the hypothesis that imposes itself as part of an answer: Ni^{II} ions, square planar-coordinated in "oversized" corphin ligands, as compared with the "optimally sized" corrin ligand, are coordinatively unsaturated: they show a greater tendency toward additional axial coordination than in corrin ligand. Analogously, axial ligands in octahedral cobalt(III) corphin complexes are bound more tightly than in corresponding cobalt(III) corrin complexes (Figure 13).

This difference has been demonstrated, at least qualitatively, in a number of cases; it is most readily observable with Ni^{II} complexes. Thus, the model Ni^{II} complex of 1,2,2,7,7,12,12,17,17-nonamethylcorrinate in methanolic solution is diamagnetic, while the Ni^{II} complex of the corresponding 2,2,7,7,12,12,17,17-octamethylcorphinate is paramagnetic, and therefore—in contrast to a corrin complex—axially coordinated (Figure 13).[49,50] For the octahedrally coordinated cobalt(III) complexes of synthetic corrins and corphins the difference is less conspicuous, but nevertheless qualitatively unequivocally apparent, for example in the facility of axial ligand exchange. Thus, a dicyanocobalt(III) corphin complex loses its cyano ligands hydrolytically much more slowly than a corresponding corrin complex.[50] Unfortunately, we did not at the time pursue the chemistry of cobalt(III)–corrin and cobalt(III)–corphin complexes to an extent such that we now would know the difference in dissociation energies for alkyl cobalt(III) complexes of the two ligand systems (as opposed to needing to estimate them qualitatively). As Halpern[51] first pointed out, the general biochemical function to be carried out by coenzyme B_{12} is to serve as an enzymatically accessible source of 5′-deoxyadenosyl radicals. Indeed, relative to other bond dissociation energies, that for the cobalt–carbon bond in coenzyme B_{12} is conspicuously low (ca. 30 kcal mol^{-1})[51,52] and crucial to the

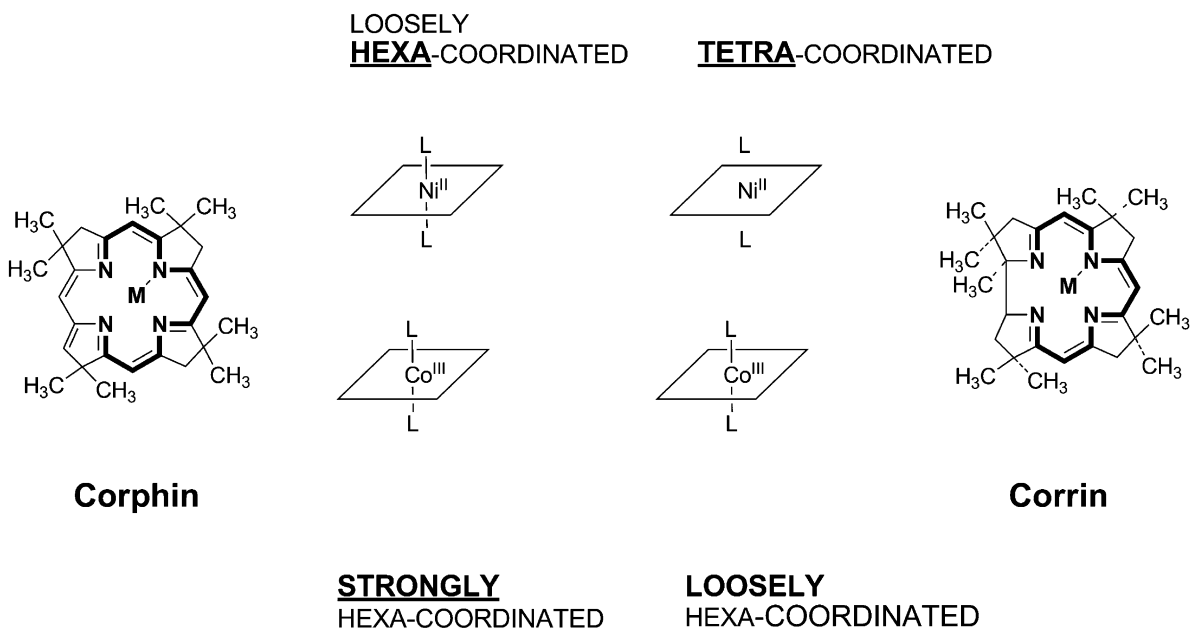


Figure 13. Ni^{II}- and Co^{II} and Co^{III}–corphin complexes bind axial ligands more tightly than corresponding corrin complexes.[31]

cofactor's enzymatic function. Therefore, a corphin/corrin comparison involving this specific bond-dissociation energy is missing in the pursuit of the question why nature "required" the biosynthetically more challenging corrin ligand instead of "making do" with a corphin.

3.3. Coenzyme F-430

In this context, special attention is warranted to a Ni^{II} complex, found about three decades after the discovery of vitamin B_{12} to be present in methanogenic bacteria. It derives once again from uroporphyrinogen, and was designated factor "F-430" by its discoverers.[53] Determination of its structure was accomplished in the author's research group at ETH by Bernhard Jaun and Andreas Pfaltz, in collaboration with the group of the microbiologist Rolf Thauer in Marburg, one of the cofactor's discoverers. The origin of this very fruitful collaboration was the microbiologist's conjecture that factor F-430 "might have something to do" with vitamin B_{12}, which led him to seek help from the ETH B_{12}-chemists. What we found was a confirmation as well as a surprise: the chemical structure of the coenzyme turned out to be that of the nickel(II) complex of a ligand system that is a derivative of a *corphin*, specifically a tetrahydrocorphin (Figure 14).[54,55] The magnetic and spectroscopic data, as well as the chemical coordination properties of this natural Ni^{II} complex, corresponded entirely with what had been observed in our previous studies on synthetic corphin and hydrocorphin complexes, and what had been concluded from comparing these with data from corresponding corrin complexes. Impetus derived from the novel structure of the new cofactor led us to develop a synthetic access to a nickel(II) model complex with the chromophore F-430 structure (Figure 15). This in turn allowed us to incorporate the model complex into our comparative study of physical and chemical properties of nickel(II) complexes of corrin, corphin, and tetrahydrocorphin ligands.[56]

In a way, we perceived the corphinoid structure of factor F-430, upon its discovery, to be something like a "commentary from nature" on our quest for an etiology of the vitamin B_{12} structure. It seemed as if the question "Why a corrin and not a corphin", posed with regard to the B_{12} structure, had been joined, so to speak, by its mirror image for the sake of a complementary inquiry into the structures and properties of the two coenzymes, B_{12} and F-430.

The corphinoid cofactor F-430 has a prominent place in contemporary microbiology. The role of the coenzymatic function of its nickel ion for methanogenesis from CO_2 and H_2 (and recently the reverse as well: *anaerobic* oxidation of methane) remains today a problem in mechanistic enzymology. Important contributions to its solution have come from the laboratory of Bernhard Jaun at ETH,[57] one of today's leading experts on the chemistry of coenzyme F-430.

4. Etiology of Nucleic Acid Structure

Just as is true of the molecular structure of coenzyme B_{12}, several simpler cofactor structures are conspicuous in containing elements that are components of the structure of the natural nucleic acids, or at least resemble such components. These include, to begin with, classical cofactors essential to the survival of even the most primitive organisms: thiamine pyrophosphate, pyridoxal phosphate, flavin adenine dinucleotide, nicotinamide adenine dinucleotide, folic acid, and coenzyme A. As already indicated, the notion has arisen that such fundamental biomolecules, or structural parts of them, could represent molecular fossils of biogenesis.[40,42] This idea implies a primordial scenario in which such cofactors, in association with nucleic acid strands capable of replication, might have functioned as "organocatalysts" that guided the

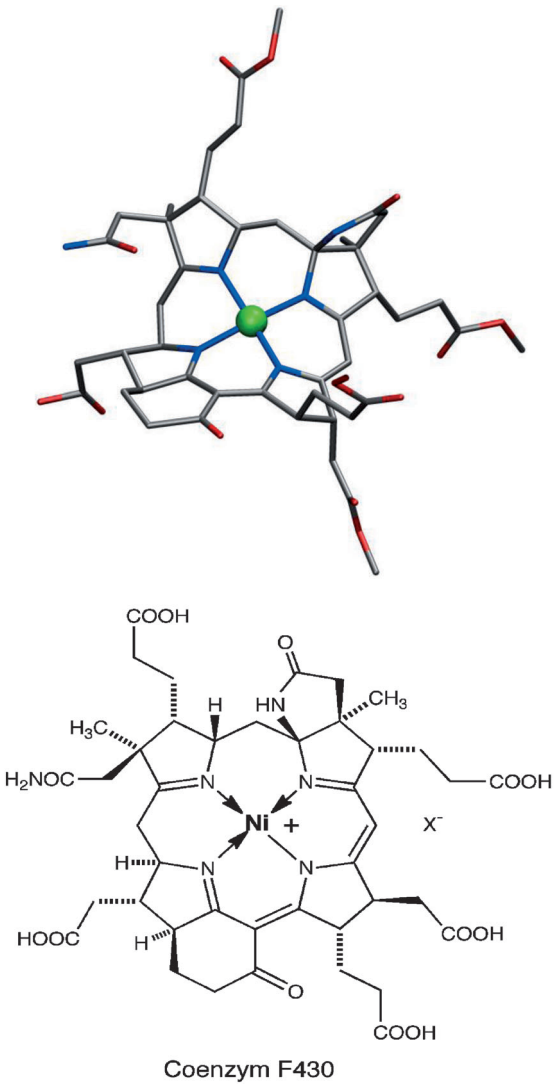


Figure 14. Structural formula for coenzyme F430[54] and X-ray structural image of 12,13-Di-epi-F430-pentamethylester[55] (X-ray structural image provided by Bernhard Jaun, ETH).

chemical processes of a primitive metabolism. From this standpoint, the nucleic acid structure is the most significant potential fossil of such a hypothetical primordial world, and is thus a correspondingly important object in a study on the etiology of potentially primordial biomolecular structures. Questions regarding the origin of the nucleic acid structure need to be directed primarily toward RNA, since all the relevant criteria—chemical, biochemical, and biological in nature—point in the same direction. RNA is the older type of structure, with DNA arising through evolution from RNA.[58,59]

In the early 1980s, in the wake of the work on the etiology of the B_{12} structure, we had taken our first steps into the realm of prebiotic chemistry.[29] These included a study of base-catalyzed aldolization of glycolaldehyde phosphate,[60] a project which, through its results, led us to ask: "Why pentose and not hexose nucleic acids?" This marked the beginning of all our further work. Actually, such a question shouldn't have needed any particular prodding for it to fall within the framework of a research project oriented toward an etiology of biomolecular structures, but we would probably never have posed it if we hadn't literally stumbled upon it through an astonishing observation, namely that base-catalyzed aldolization of glycolaldehyde phosphate can just as definitively lead to the principal product allopyranose-2,4,6-triphosphate as it can—in the presence of formaldehyde—to ribopyranose-2,4-diphosphate as the main product. In view of the impact of this finding on our work, there follows a presentation of its background.

4.1. Glycolaldehyde Phosphate

The "formose reaction" is a child of 19th century chemistry. Usually initiated with glycolaldehyde, it involves the base-catalyzed and autocatalytic formation of complex mixtures of monosacharides from formaldehyde.[61] Together with the creation of α-amino acids and the canonical nucleobases from the most elementary of starting materials,

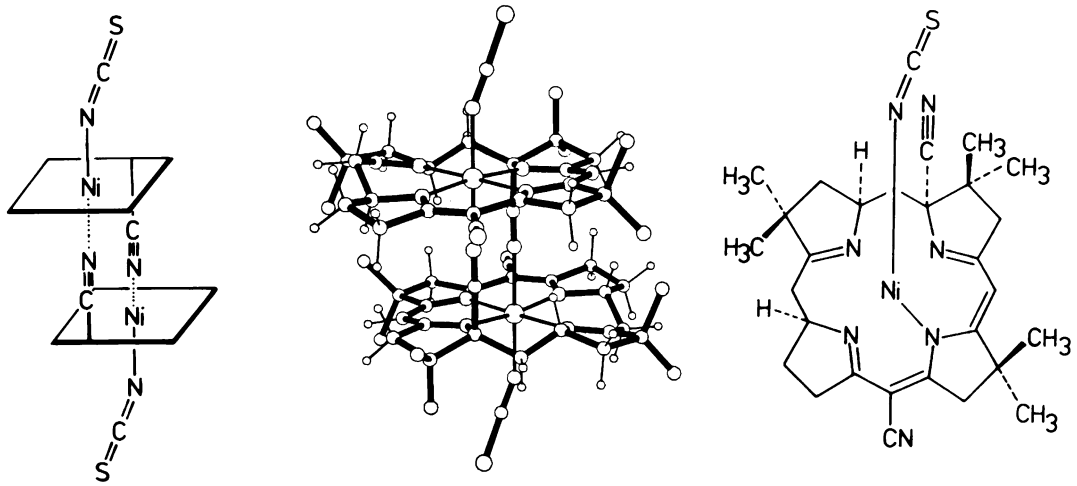


Figure 15. Structural formula and X-ray structural image for a synthetic Ni^{II}-model complex of coenzyme F430.[56]

the formose reaction belongs among the three pillars of classical prebiotic chemistry.[7] Common to all three, however, are acute inherent problems associated with a notoriously low efficiency of product formation and a persistent uncertainty regarding each of these processes' true role on the pathway toward biogenesis. In the case of the formose reaction this takes the form of a problematic requirement for high formaldehyde concentrations if the process is not initiated by glycolaldehyde or other components of its own product palette, together with a low constitutional selectivity in product formation. The classical way of conducting the formose reaction leads to complex reaction mixtures, in which ribose is present in such meager amounts that relevance of the formose reaction to the origin of RNA has frequently been questioned, and rightly so.[62] In one often-cited piece of work from the laboratory of Peter Decker,[63] analysis by gas chromatography of a "formose" product mixture obtained after initiation of the reaction with glycolaldehyde and subsequent derivatization of the resulting mixture revealed the presence of over 40 components, 33 of which were assigned constitutionally. The chief constituent within the associated sugar mixture (obtained overall in ca. 20% yield) was *rac*-xylulose, making up ca. 2.5%.

It was the notorious ribose problem that induced us at the time to investigate the aldolization of glycolaldehyde *phosphate*. On the one hand, it was to be anticipated that, with substitution of a phosphate in place of free glycolaldehyde, all the reaction steps involving tautomerization of aldoses to ketoses in the corresponding aldolization network would disappear, thereby eliminating one source of constitutional complexity in the reaction mixture. Moreover, there was the well-known—but still no less remarkable—fact that in cellular biochemical processes, monosaccharides apparently never operate in the free state, but always in *phosphorylated* form. It is a short step from such considerations to the notion of a primordial scenario in which, again, phosphorylated and not simply neutral forms of carbohydrates would have been operative. In a self-organization process in a primordial environment, it may have been of primary importance for carbohydrate molecules to escape chemical chaos, finding themselves instead at concentrations suitable for chemical reactions, and in reaction spaces that would facilitate efficient chemical transformation. With respect to both requirements, phosphorylated sugar molecules would, through their electrical charges, have offered advantages over neutral, water-soluble carbohydrates in environments containing mineral surfaces or minerals with expandable layer structures.[64]

A typical result of our experiments on the aldolization of glycolaldehyde phosphate in strongly basic, aqueous solution, and in both the absence and the presence of formaldehyde, is provided by Figures 16–18, documenting GC analytical data for derivatized product mixtures, which, in conjunction with NMR spectroscopy of the underivatized reaction products, permits reliable assessment of the constitutional and configurative course of the reactions.[65] As these GC data convincingly (albeit indirectly) show, in the absence of formaldehyde the reaction products are mainly (racemic) hexose-2,4,6-triphosphates, with *rac*-allose-2,4,6-triphosphate as clearly the principal reaction product (46% of the total amount of the eight diastereomeric hexoses). Under identical reaction conditions, *but in the presence of a half-molar equivalent of formaldehyde*, the lion's share of what results—together with small amounts of hexose triphosphates—is the family of the four diastereomeric pentose-2,4-diphosphates. Surprisingly but equally noteworthy, the chief component is *rac*-ribose-2,4-diphosphate (ca. 50% of the total amount of the four diastereomeric pentoses).[60] Under otherwise largely identical conditions, aldolization of an equimolar mixture of glycolaldehyde phosphate and *rac*-glyceraldehyde-2-phosphate (see below) leads also to the four pentose-2,4-diphosphates, among which, and to a similar extent, the ribose derivative again dominates. The two tetrose-2,4-diphosphates survive hexose triphosphate formation under the conditions of glycolaldehyde phosphate aldolization only to a small extent; by reducing the basicity of the medium they too can become major products (in roughly equal amounts). If under otherwise standard conditions for the formation of the family of pentose-2,4-diphosphates one allows the reaction to run not simply for one week but rather for 23 weeks, the total amount of the four diastereomers is largely unchanged, but the ratio of the diastereomers is displaced such that the content of ribose diphosphate is increasingly diminished in favor of arabinose diphosphate, with the latter ultimately gaining the upper hand (Figure 19). This observation accounts for the conclusion that *the preferential formation of ribose-2,4-diphosphate under the reaction conditions employed is a result of kinetic reaction control*.

These investigations have shown that, in a formation of pentoses through a process related to the formose reaction, ribose can be the preferred product among the four resulting pentoses. To our knowledge, this provided the first evidence for an intrinsic formation selectivity of ribose. In earlier discussions about the formose reaction as potential prebiotic source of the RNA constituent ribose, such a selectivity would in fact have been regarded as most unlikely.[66] In the meantime, the characteristic of ribose being the most rapidly formed pentose in aldolization reactions under kinetic control has been noted as well by Zubay[67] in aldolizations involving free glycolaldehyde. Moreover, recent work by Benner[68] on borate complexes of ribose has revealed further possibilities of not only a kinetic, but also a thermodynamic preference for this sugar over other pentoses.

Attempts to understand on the level of a qualitative conformational analysis the observed configurational selectivity in glycolaldehyde phosphate aldolizations led to a result that was rather surprising, given the complexity of the system. Minimization of "Newman strain",[69] to be attributed to 1,5-interactions between (hydrated!) phosphate groups and free hydroxy groups, together with minimization of steric hindrance in the transition states of product-defining reaction steps, leads to a picture of preferred reaction paths that largely correspond to what is observed experimentally (for details cf. ref. [60]). This applies above all to formation of the principal products *rac*-ribose-2,4-diphosphate and *rac*-allose-2,4,6-triphosphate; both are the result of aldolization steps that proceed consistently in an erythroid fashion ("allose is the ribose of the hexose series").[70]

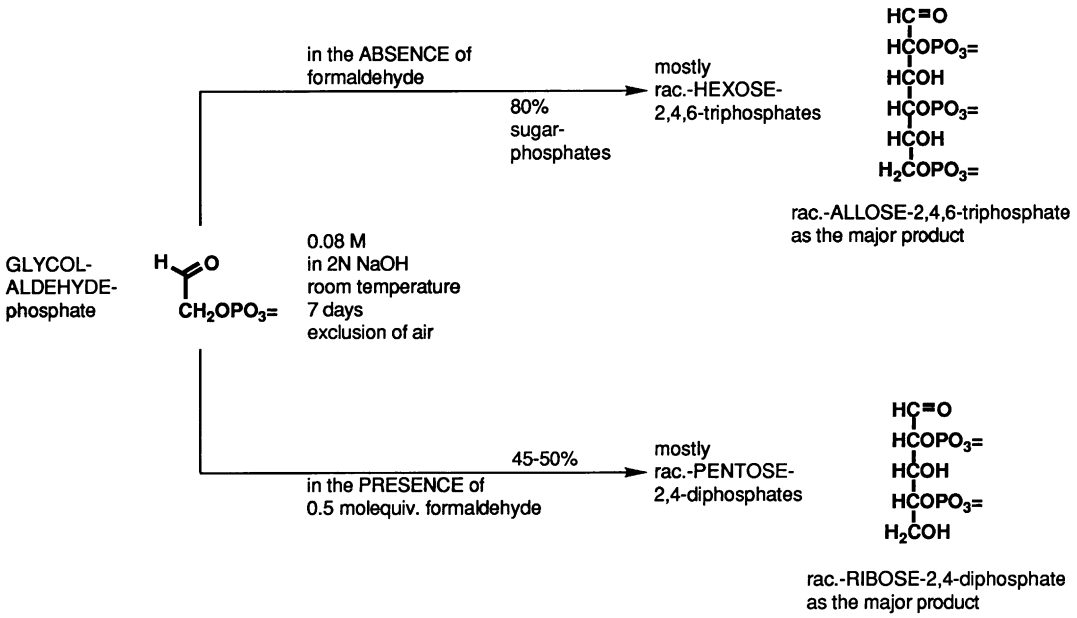


Figure 16. Aldolization of glycolaldehyde phosphate in aqueous alkaline solution in the absence and presence of formaldehyde.[60]

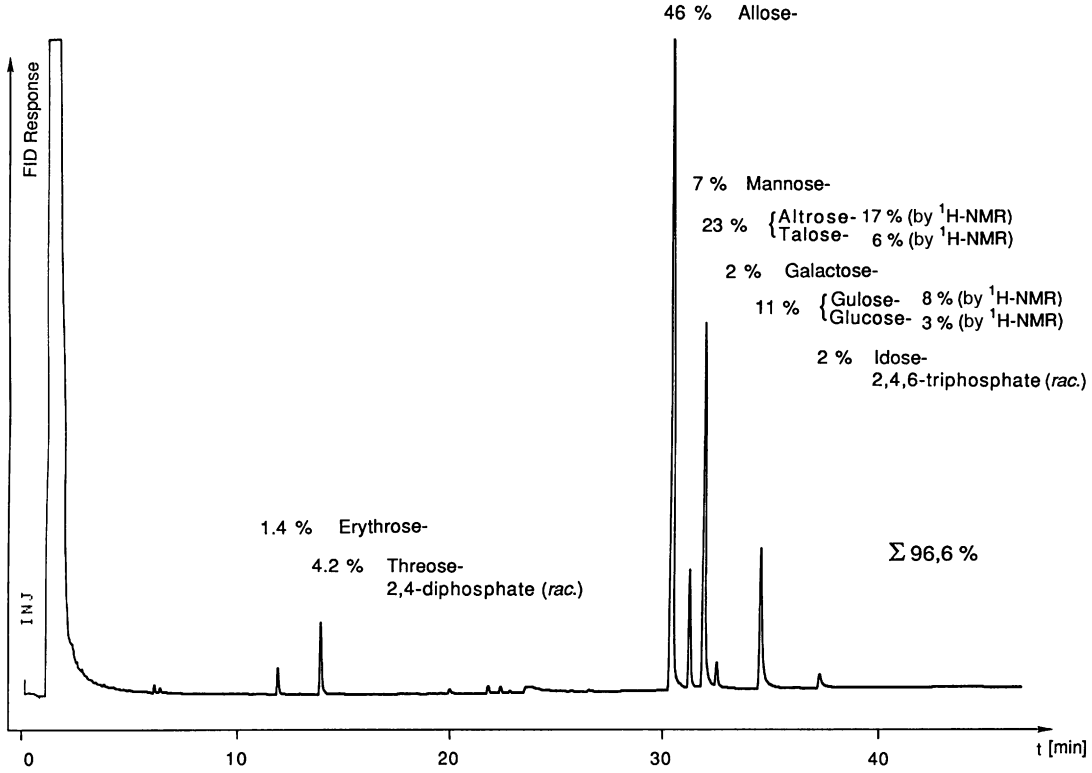


Figure 17. Gas chromatographic analysis of the product mixture from aldolization of glycolaldehyde phosphate in the absence of formaldehyde (Figure 16) after enzymatic dephosphorylation, reduction with $NaBH_4$, and acetylation: the derivative of allose as the principal component.[60]

Our work on the aldolization of glycolaldehyde phosphate was focused on optimization of product formation and analysis of product composition, with no attempt made to simulate "prebiotic reaction conditions." An extension of the project in the latter direction was undertaken, however, in collaboration with the research group of the geochemist Gustaf Arrhenius at the SIO.[71] Highly diluted in an aqueous 0.1N NaOH solution in the presence of expansive layer minerals of the hydrotalcite type (magnesium aluminum hydroxide carbonates),[72] glycolaldehyde phosphate aldolizes, presumably within mineral layers, to give a sugar phosphate mixture containing the two tetrose-2,4-diphos-

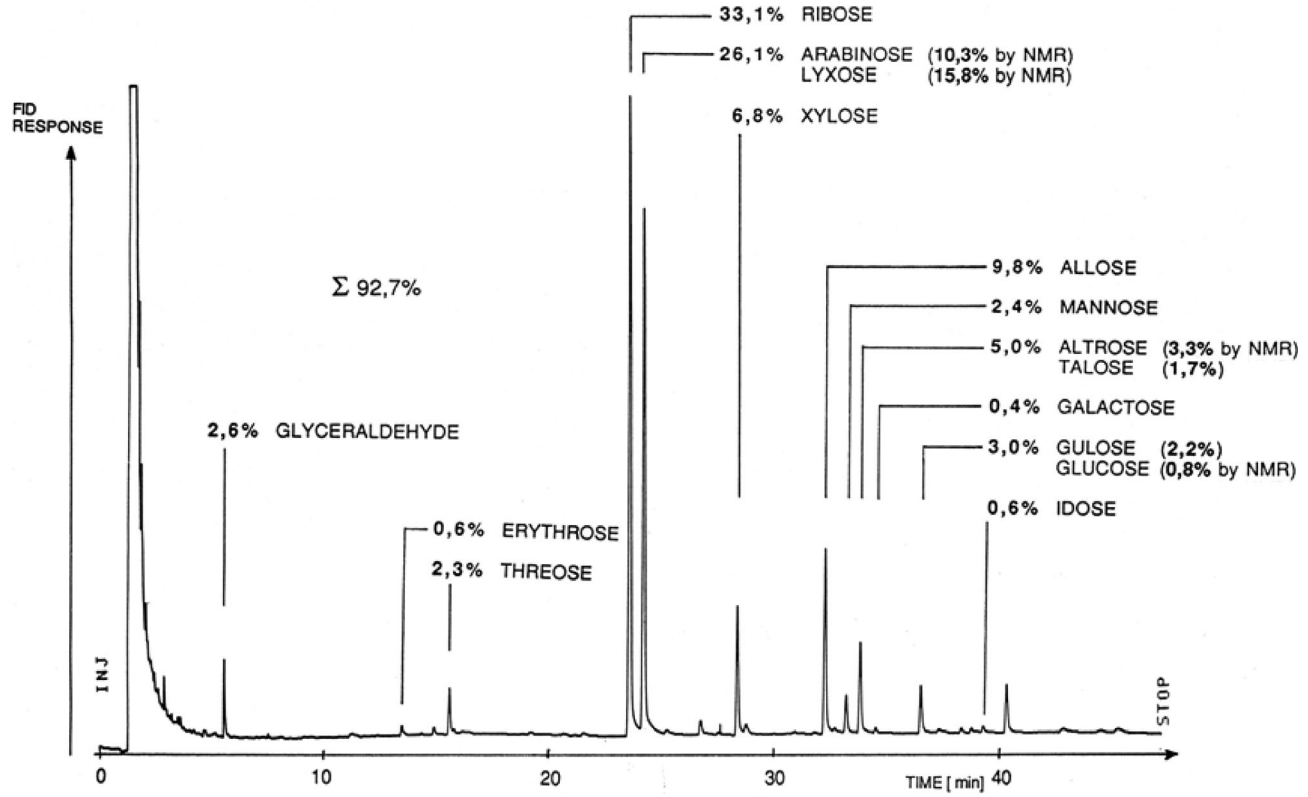


Figure 18. Gas chromatographic analysis of the product mixture from aldolization of glycolaldehyde phosphate in the presence of formaldehyde (cf. Figure 16) after enzymatic dephosphorylation, reduction with $NaBH_4$, and acetylation: the derivative of ribose as the principal component.[60]

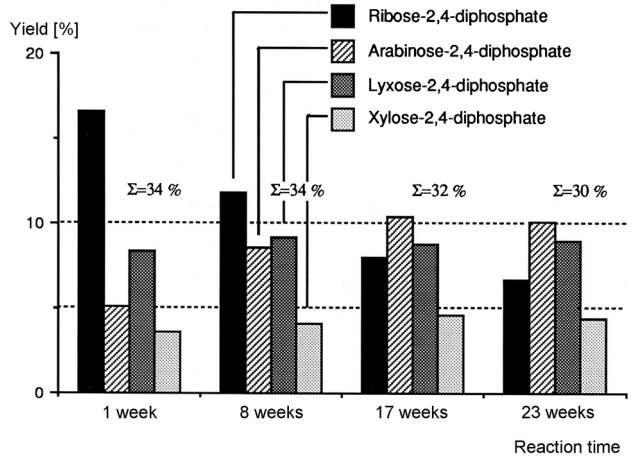


Figure 19. Product distribution in aldolization of glycolaldehyde phosphate in the presence of formaldehyde (cf. Figure 16) as a function of length of the reaction: the preferential formation of *rac*-ribose-2,4-diphosphate with a short reaction time (Figures 16 and 18) is the result of kinetic control of the course of the reaction.[60]

phates and once again the eight hexose-2,4,6-triphosphates, including prominent amounts of the allose isomer.[73] By contrast, in the absence of layer minerals, but under otherwise identical conditions, one observes absolutely no aldolization of the glycolaldehyde phosphate. Experiments in the presence of formaldehyde under these same conditions produced pentose-2,4-diphosphates in only very limited amounts. However, it was observed in a follow-up study at the SIO that, from a highly dilute, neutral solution, layer minerals caused *rac*-glyceraldehyde-2-phosphate and glycolaldehyde phosphate to co-aldolize to pentose-2,4-diphosphates in substantial yield (ca. 25%, with a ribose-2,4-diphosphate content of ca. 50%).[74]

Might glycolaldehyde phosphate be regarded as a potential prebiotic reaction component? Not at least on the basis of modes of formation for this compound as reported in the literature.[75] At ETH, one of our early studies on potentially prebiotic reactions (studies which today we regard as outdated) was focused on the chemistry of the oxide of acrylonitrile.[76] This epoxide was found to react rapidly and regioselectively in aqueous phosphate solution at room temperature to give glycolaldehyde phosphate cyanohydrin, which by cyanide transfer to for example, formaldehyde was readily transformed into glycolaldehyde phosphate.[77] The actual challenge, however, was ultimately the direct phosphorylation of glycolaldehyde in aqueous solution. Extensive efforts at the SIO[78] to achieve this goal with cyclotriphosphate[79] as a potentially prebiotic phosphorylation reagent were unsuccessful. Studies were subsequently resumed in the laboratory at TSRI that in the meantime had been established jointly with my former postdoctoral collaborator Ramanarayanan Krishnamurthy, this time using as reagent the ammonolysis product of cyclotriphosphate, amidotriphosphate. This substance suggested itself[80] as a reagent for regioselective phosphorylation of α-hydroxyaldehydes on the basis of its reaction potential as illustrated in Figure 20. It

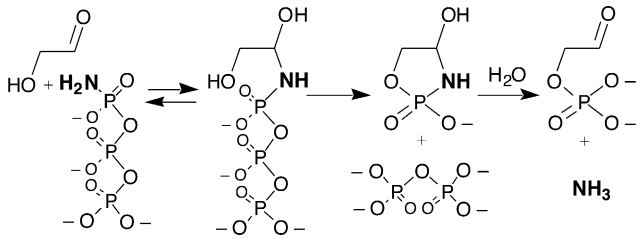


Figure 20. Principle of intramolecular phosphorylation of glycolaldehyde with amidotriphosphate.

fulfilled expectations to a remarkable degree (Figure 21), and brought closer to a possible answer the question of whether glycolaldehyde phosphate could have been a prebiotic molecule.[81] The success of this method of regiospecific phosphorylation is based on the principle of targeted conversion of an intermolecular into an intramolecular product-determining reaction step (cf. ref. [19a]).

Amidotriphosphate as an 0.1 molar, nearly neutral aqueous solution, in the presence of magnesium ions, transforms 0.025 molar solutions of glycolaldehyde within days at room temperature into glycolaldehyde phosphate in (analytically) nearly quantitative yield (Figure 21).[81] The reaction still runs with great efficiency (^{1}H- and ^{31}P NMR) even at substrate concentrations as low as ca. 30 μM and a reagent concentration of 60 μM, whereby this phosphorylation fulfills one of the principal demands one would set for it to be classed as potentially prebiotic. The phosphorylation of glyceraldehyde with amidotriphosphate leads, under the reaction conditions employed, both regioselectively and in high yield to glyceraldehyde-2-phosphate (Figure 21).[82]

From the very beginning of our work on etiology of the nucleic acid structure we attempted to bypass questions of *formation* and restrict ourselves to *function* in order to avoid the uncertainties of "historicizing chemical research". We tried thereby to conform to a byword we had adopted for ourselves (after learning the hard way) in the field of prebiotic chemistry:[29,76] Whenever possible, experiments on the etiology of potentially primordial biomolecular structures should be planned such that the results might not only have etiological relevance, but also class as valuable contributions to chemistry in general. Etiological relevance will ultimately remain uncertain; if sometime later it should succumb to new insights, the results still survive as a contribution to chemistry. The experiments on phosphorylation of glycolaldehyde with amidotriphosphate present an example: they arose from a question about the prebioticity of glycolaldehyde phosphate, and their outcome is of chemical interest, this irrespective of the survival of the conjecture that glycolaldehyde phosphate may have played a role in biogenesis.

4.2. "Why" Questions

"Why questions" of the type "Why a corrin and not a corphin?" or "Why pentose and not hexose nucleic acids?" are not posed in an absolute sense, but strictly in relation to an alternative.[17,83,84] This point seems critical to making it possible to address such questions in a sensible way. Frank Westheimer,[83] in his classic essay "Why Nature Chose Phosphates", drew attention to both metabolic processes and the structure of the nucleic acids, aiming directly at alternatives such as sulfates, silicates, or arsenates as formally possible, but—as he argued—functionally unsuitable phosphate alternatives. Experiments permitting a comparison between biomolecular structures and potential alternative structures may address either the *formation* or the *function* of the biomolecule, or both. If an answer is intended from the outset to relate to function, then it must be possible in in vitro experiments to correlate the latter with a chemically measurable characteristic of the biomolecule. The comparison of a biomolecular structure with an alternative structure with reference to a characteristic property of the biomolecule can make apparent key elements of a rationale for why this specific structure was selected in preference to the alternative,

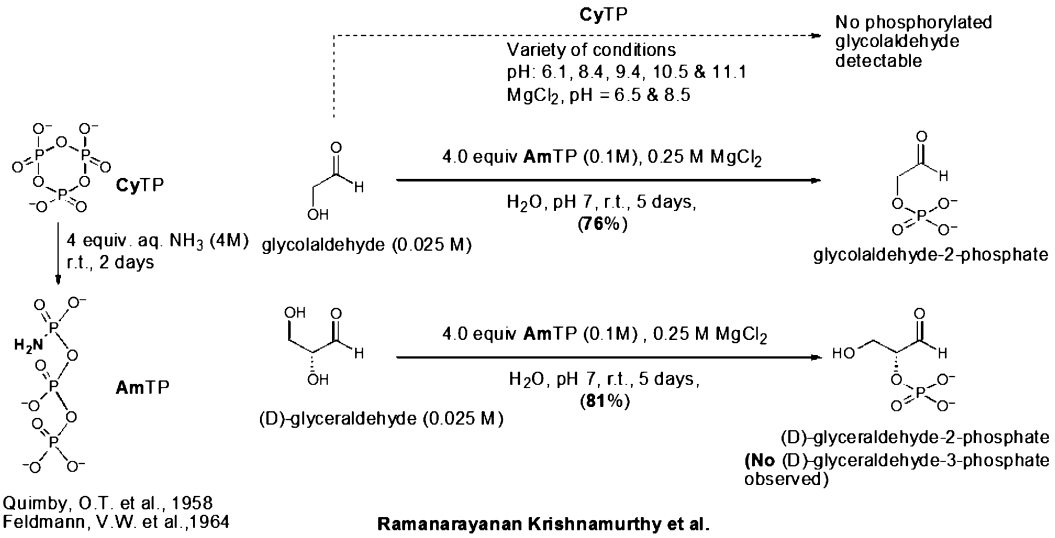


Figure 21. Regioselective phosphorylation of α-hydroxyaldehydes with amidotriphosphate.[81,83]

especially if the latter can be classed as a "potentially natural" structure. This is supposed to be the case if, from a chemical perspective, the types of processes presumed to underlie formation of the biomolecular structure would have had the potential for also producing the alternative structure.

The nucleic acid structure fulfills in nearly ideal fashion the requirements for addressing experimentally the "why this, and not that " question. Above all, there can be no doubt what kind of chemical property is to mediate between the structure of the biomolecule and that of the alternative: it is clearly the capacity for informational base pairing. Characteristics that are variable include the specific sugar module, the structure of the nucleobases, and the linkage between mononucleotide units. Most obvious from both chemical and etiological standpoints is a variation in the sugar module. The repeating pattern in the constitutional arrangement of functional groups in the $(CH_2O)_{n=4,5,6}$ aldose sugar family is such that ribose can be replaced by alternative aldoses without other RNA structural elements necessarily being changed at the same time. That permits fulfillment of the requirement that alternatives should have "potentially natural" structures. Although we of course know nothing about the circumstances that originally generated the RNA structure, it can be assumed that those circumstances had the potential to permit inclusion into an oligonucleotide structure not only of ribose but, by chemically analogous paths, other aldoses as well.

In principle, this expectation implies an evolutionary scenario of selection of the RNA structure from a natural offering of various oligomeric structures. Systematic variation of the sugar modules by synthetic chemical means, and comparison of the resulting variants with RNA from the standpoint of efficiency of informational Watson–Crick base pairing, has the character of an experimental simulation of such a potentially natural process. The comparisons involved refer to the function of RNA, not its formation; *thus, the etiological relevance of the project is independent of whether the natural processes were of a prebiotic or a biotic nature.*

4.3. "Homo-DNA"—A Model Study

The number of possibilities formally to be taken into account through variation of sugar modules within the nucleic acid type of structure (trioses, tetroses, pentoses, hexoses; furanose and pyranose forms) is so large that the first task was to establish selection criteria for setting priorities in the choice of variants to be addressed experimentally. The most important of these criteria was the likelihood to be assigned to a given variant that it might turn out to involve an informational[85] Watson–Crick pairing system. When the project was initiated,[86] the literature offered no guidance with respect to such an evaluation. Since we had no access to a theoretical treatment of the problem (quite apart from the questionable efficacy in those days in predicting an oligomer system's potential for Watson–Crick pairing by computer-assisted modeling) we tackled the matter of base-pairing prognoses for oligonucleotide single-strands by using qualitative conformational analysis. What under no circumstances would have seemed possible for nucleic acids based on furanose modules, promised to be attainable with an oligonucleotide containing pyranose modules. This is because it had, after all, been the lucid chair conformation of cyclohexane as opposed to the intricate conformation of cyclopentane that—since Derek Barton's classic paper—had been the very icon for success in qualitative conformational analysis in organic natural product chemistry.

The outset of our experimental involvement in the nucleic acid project was marked by the so-called "homo-DNA",[87,88] an oligonucleotide system composed of D-2,3-dideoxyallopyranose instead of DNA's D-2-deoxyribofuranose as its sugar building block. The only difference between homo-DNA and DNA consists in an extra endocyclic CH_2 group in the sugar module of the former; both the nucleotidic connections of the (canonical) nucleobases, and the phosphodiester bridges between sugar rings correspond constitutionally and configuratively (both relative and absolute) to the structure of natural DNA (Figure 22).

Dideoxyhexoses, in contrast to normal hexoses, are far from classifying as "potentially natural" sugars in the sense defined above.[89] Therefore, from an etiological perspective,

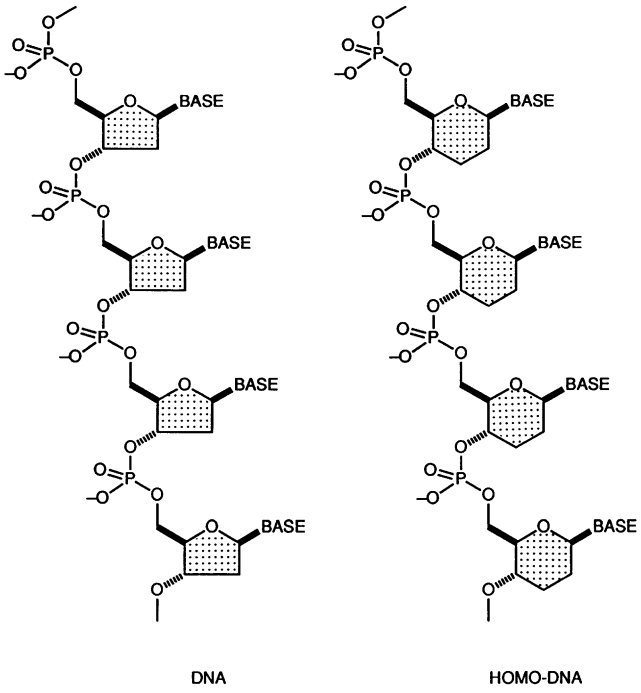


Figure 22. The only difference between the chemical structure of a "homo-DNA" single strand and that of a DNA single strand is a CH_2 group in each of the rings.

the work on homo-DNA was regarded as a model study only, albeit one that proved to greatly facilitate our entry into base-pairing prognoses, the development of methodology for oligonucleotide preparation, and the interpretation of base-pairing characteristics of alternative nucleic acids. Figure 23 recalls the principle and results of the first pairing prognosis undertaken in this context. The analysis was based strictly on considering only so-called "idealized" conformations, which we defined as all single bonds in a constitutionally repetitive

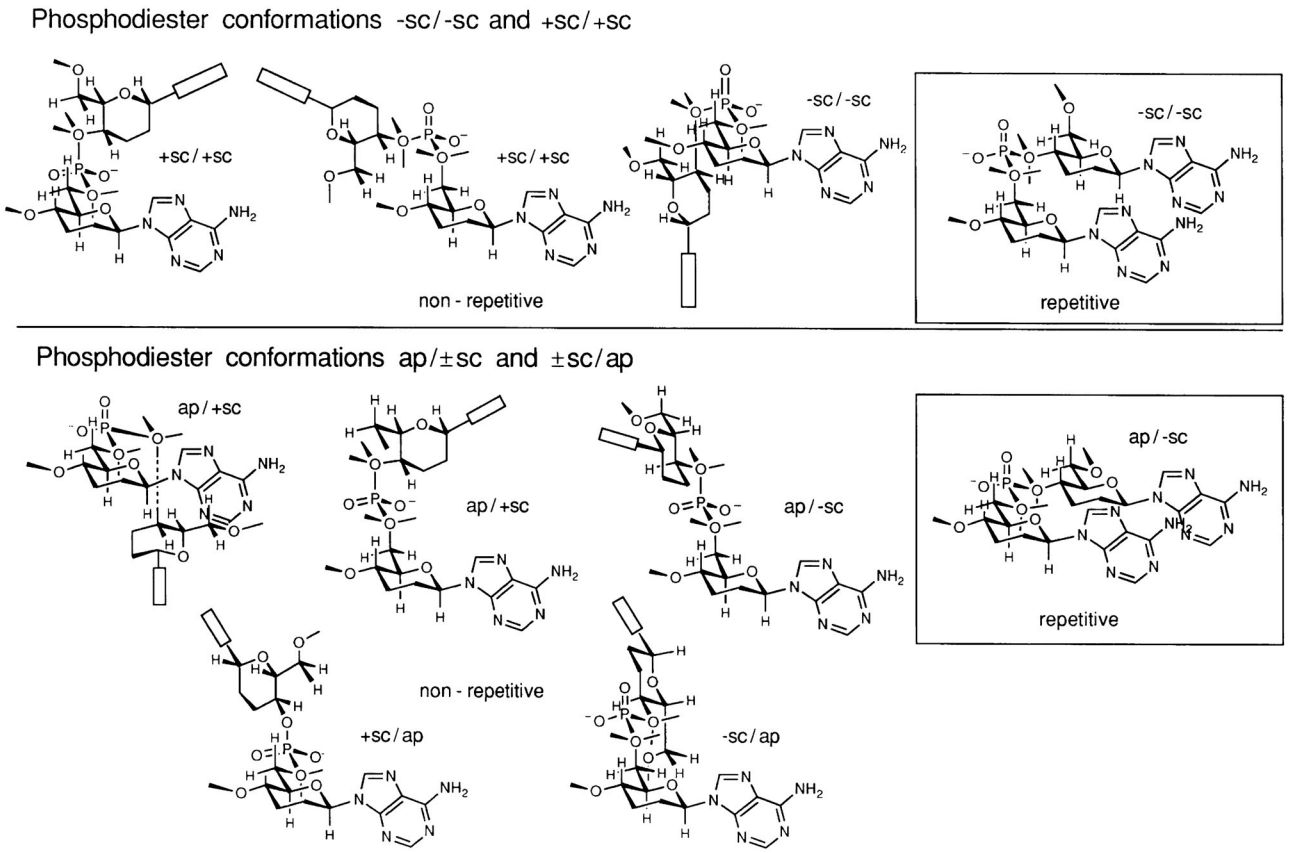


Figure 23. Pairing prognosis for homo-DNA by qualitative conformational analysis on the basis of idealized conformations: Of a total of 4 sterically least-hindered conformations of a backbone unit with two-fold synclinal phosphodiester groupings, only one conformation is "repetitive" (top of figure); this conformation is expected to be the paring conformation. If one also takes into account conformations with *trans*-synclinal phosphodiester groupings (bottom of figure), there will be two possible pairing conformations.[90]

pyranose-phosphodiester unit being ideally staggered. For such a repetitive unit, the set of enthalpically most stable conformations is then established through the criterion of minimizing the number of steric 1,5-repulsions ("Newman strain"),[69] as well as excluding doubly antiparallel phosphodiester conformations, using stereoelectronic reasoning.

A judgment is then made as to whether, within such a set of most stable conformations, at least one is "repetitive", i.e., would identically repeat itself when built into a single-strand oligomer. If this is in fact the case, then pairing is assumed to be possible, and the repetitive conformation of the monomer unit is expected to be the pairing conformation for the corresponding oligomer system.[90,91]

The set of most stable (idealized) conformations for the repeating homo-DNA unit consists of a total of ten conformations, two of which are repetitive (Figure 23). These two differ in spatial orientation of the substituents of the phosphodiester group (synclinal/synclinal vs. synclinal/antiparallel). If, based on qualitative stereoelectronic considerations (anomer effect), one follows the postulate that a synclinal (±sc) conformation of a phosphodiester arm should be favored over the antiparallel (ap) conformation, then the decision falls to the (−sc/−sc) conformation as the homo-DNA pairing conformation (Figure 23).[90]

At the time, it was with some satisfaction that we took note of the fact that this type of (idealized) conformation turns out to correspond precisely to the A-type strand conformation familiar from DNA duplex X-ray structure determinations, and thus also to the pairing conformation of RNA (Figure 24). Not only did this appear to be welcome evidence justifying the pairing-prognosis approach, but conversely it also represented a rationalization on a qualitative

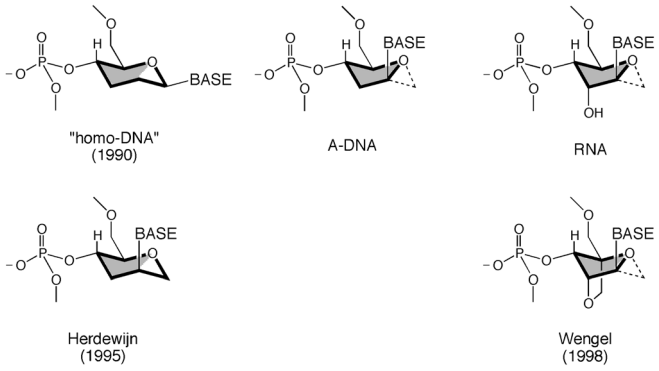


Figure 24. Conformational analogy between the structures of homo-DNA, Type-A DNA, and RNA, as well as the structures of Wengel's "locked-RNA"[94] and Herdewijn's "hexitol-NA".[95]

level of the pairing conformation of RNA. That is to say, since the decisive selection criterion in derivation of the pairing conformation for homo-DNA was minimization of the number of steric 1,5-repulsions, a correspondence between the pairing conformations of homo-DNA and RNA therefore suggests that the conformation of natural RNA likewise results from minimization of this type of steric repulsion. Thus, the characteristic A-type conformation of the RNA oligomer backbone in duplexes seems imposed not primarily through spatial demands of Watson–Crick base pairs, but may rather be an intrinsic property of the backbone. To the extent that this is true, the RNA single strand is pre-organized toward the conformation it assumes in hairpins and duplexes, and thus toward Watson–Crick pairing. Figure 25 illustrates why, despite differing constitutions of the sugar modules, such a conformational inference regarding RNA can be made from homo-DNA.

Describing and interpreting the pairing conformation of nucleotidic nucleic acid systems at the level of "idealized"

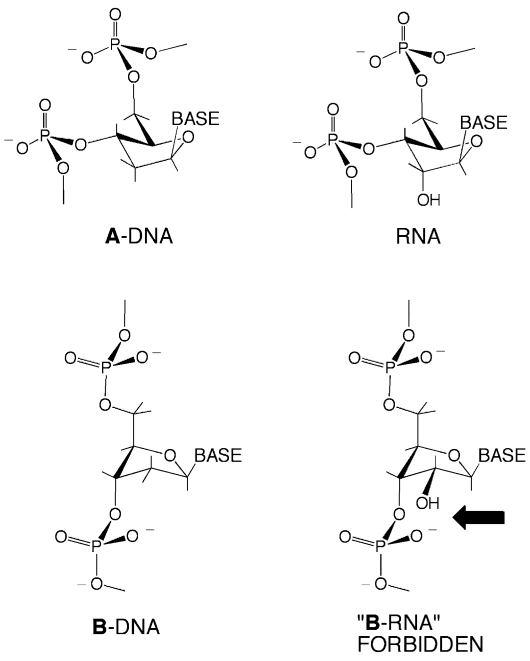


Figure 25. Conformational considerations on the basis of idealized conformations make it possible to see why DNA is accessible as Type A and Type B pairing conformations, but RNA only as Type A.

conformations can simplify recognition of relationships that in fact often appear blurred in a more realistic presentation of nucleic acid structural data. The loss in reality accompanying rendition of such an idealization is frequently compensated by greater visibility and clarity of essentials. For example, the conformational transition from type A to type B DNA becomes, for the organic chemist, readily recognizable as the flipping of one "half-chair" form of a five-membered ring into the other "half-chair" form (Figure 25). Similarly, the "Van der Waals clash", noted by the discoverers of the DNA-double helix[92,93] to be responsible for the fact that, in contrast to DNA, a type B double helix would not be feasible for RNA, becomes clearly identifiable as steric 1,5-repulsion of the Newman type (Figure 25). The increased base-pairing strength—relative to RNA—of the "locked RNA" described by Wengel[94] reveals itself just as conspicuously as the ability of Herdewijn's homo-DNA isomer (called "HNA")[95] to undergo—in contrast to homo-DNA itself—cross-pairing with natural nucleic acids (Figure 24) (see below).

As will be further elaborated below, homo-DNA strands with even relatively short complementary base sequences form Watson–Crick duplexes that are much more stable than the corresponding DNA-duplexes. On the other hand, homo-DNA does not undergo cross-pairing with DNA or RNA. This feature was at the time not necessarily met with surprise, since the conformational (idealized) model of a homo-DNA single strand (in contrast to the corresponding model of a DNA strand) had indicated from the outset that a homo-DNA double strand would not be helical, tending instead to be linear. An NMR spectroscopic analysis of the structure of the duplex of the self-complementary homo-DNA base sequence 6′-A_5T_5, carried out by Otting in the laboratory of Wüthrich,[96] revealed a structure that was indeed largely linear. While in fact showing the predicted sc/sc type phosphodiester conformation to be present, this structure proved more complex than anticipated, in that it was necessary to interpret the conformation of the two strands within the duplex as being in dynamic equilibrium between (−sc/−sc) and (ap/−sc) forms (Figure 26, left side).

When crystallization of the homo-DNA duplex containing the self-complementary base sequence 6′-CGAATTCG was achieved, it was anticipated that X-ray structural analysis would permit conclusive resolution of the problem. Instead, in the hands of more than one expert in X-ray analysis, the structure determination turned out to represent a puzzle so complex that nearly two decades passed before Martin Egli[97] was able to provide the solution. His achievement entailed systematic preparation and screening of a series of derivatives, culminating in a duplex of the selenophosphodiester sequence $CGA_{PSe}ATTCG$. The latter revealed itself to have a crystal structure of capricious complexity: namely, one of duplex-dimers held together by inter-duplex hydrogen bonds between two nucleobases that are "dislocated" from the normal base-pairing position within their duplex molecules (Figure 26, right side). Despite this irregularity and the shortness of the oligonucleotide strands, the originally postulated characteristics of the homo-DNA structure were seen to be present in the structure's regular parts; these included: antiparallel strand arrangement within the duplex, chair conformation for the hexopyranose ring bearing equatorial nucleobases, sc/sc type phosphodiester conformation, and prevailing interstrand stacking of nucleobases as a consequence of a distinct inclination between (local) strand and base-pair axes.[90]

Since we had recognized quite early on that in homo-DNA we had in hand for the first time an oligomer system capable of self-pairing but not of cross-pairing with DNA, i.e., a system capable of communicating with itself using a "base-pairing language" different from the "language" of the natural nucleic acids, it was not least for this reason that we undertook an extensive experimental study of its pairing characteristics.[98–101] The most important of these character-

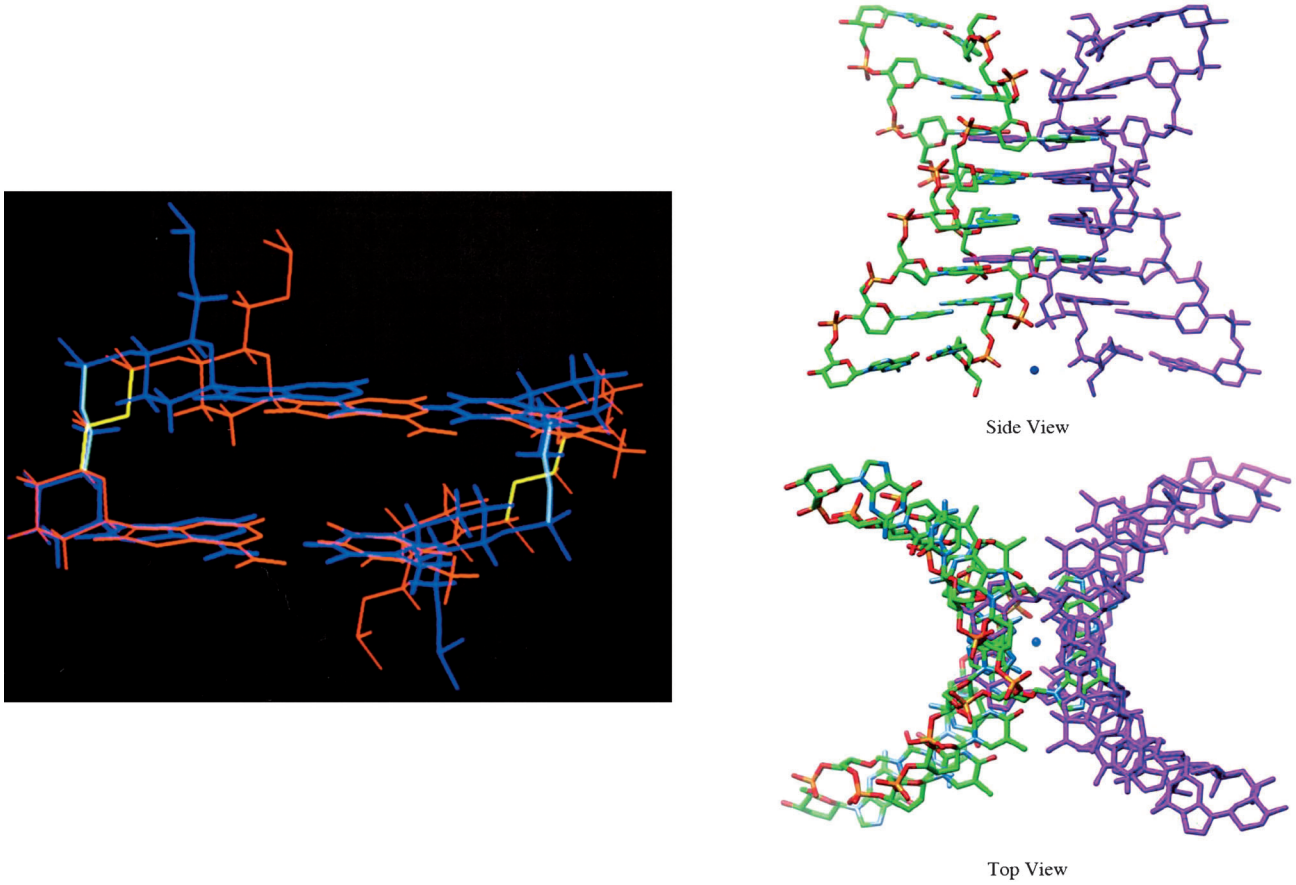


Figure 26. NMR-spectroscopic (left) and X-ray structure analytical (right) information on the structure of homo-DNA duplexes: The sc/sc- and the ap/sc pairing conformations (Figure 23) are co-present in the duplex of the antiparallel self-complementary base sequence 6′-A_5T_5;[96] the X-ray structure of the self-complementary sequence 6′-CGAATTCG is complicated by lateral dimerization of the duplex.[97]

istics are the following: As a consequence of strong backbone inclination, no base-pairing is observed in parallel strand orientation; all four canonical Watson–Crick pairs are stronger than in DNA; homobasic sequences of adenine show strong adenine–adenine pairing in the reverse-Hoogsteen mode, as do corresponding sequences of guanine

HOMO – DNA : PURINE – PURINE PAIRING

MELTING TEMPERATURES (°C) OF HEXAMER – DUPLEXES:

15-20μM Nucleotide
150mM NaCl
10mM Tris pH 7

		A	H	D	X	I	G
ADENINE	A	47					
HYPOXANTHINE	H	<7	—				
2,6–DIAMINOPURINE	D	41	<5	36			
XANTHINE	X	14	—	**63**	—		
ISOGUANINE	I	43	12	39	16	42	
GUANINE	G	<15	<15	<15	18	**61**	38

dd-BBBBBB BBBBBB-dd
dd-BBBBBB **BBBBBB**-dd
dd-BB**BBBB** **BBBB**BB-dd
dd-BBB**BBB** **BBB**BBB-dd

~40 HOOGSTEEN (BIDENTATE)
~60 WATSON – CRICK (TRIDENTATE)
— NO PAIRING OBSERVED

Figure 27. Purine–purine pairing in homo-DNA duplexes (T_m-values): Watson–Crick pairing of xanthine with 2,6-diaminopurine and of guanine with isoguanine, as well as self- and cross-pairings in the reverse-Hoogsteen mode of the purines adenine, 2,6-diaminopurine, isoguanine, and guanine (cf. Figure 28).[100]

(Figure 27); and both of these purine–purine pairs are stronger than the canonical adenine–thymine pair, but the guanine–guanine pair is weaker than the canonical guanine–cytosine pair. This means that—depending upon the base sequence—homo-DNA's pairing selectivity is different from that of DNA.

In homo-DNA, duplexes have for the first time been observed in which the strands are held together by purine–purine base pairing in the Watson–Crick mode: xanthine (X) and guanine (G) form strong (presumably) tridentate purine–purine base pairs (X/D and G/I) with their non-canonical partners 2,6-diaminopurine (D) and isoguanine (I), respectively (Figure 28). In addition, 2,6-diaminopurine and isoguanine can pair both with themselves and with each other, as well as with the two canonical purine bases, all in the reverse-Hoogsteen mode; such pairs are significantly weaker than the Watson–Crick purine–purine base pairs A/D and G/I[102] (Figure 27).

The thermodynamic data reproduced in Figure 29 suggest an entropic advantage for canonically paired homo-DNA duplexes relative to corresponding DNA duplexes. This points—in comparison to DNA—to a higher degree of pre-organization of the homo-DNA strand toward the type of conformation it will adopt in a duplex. Such a view is in line

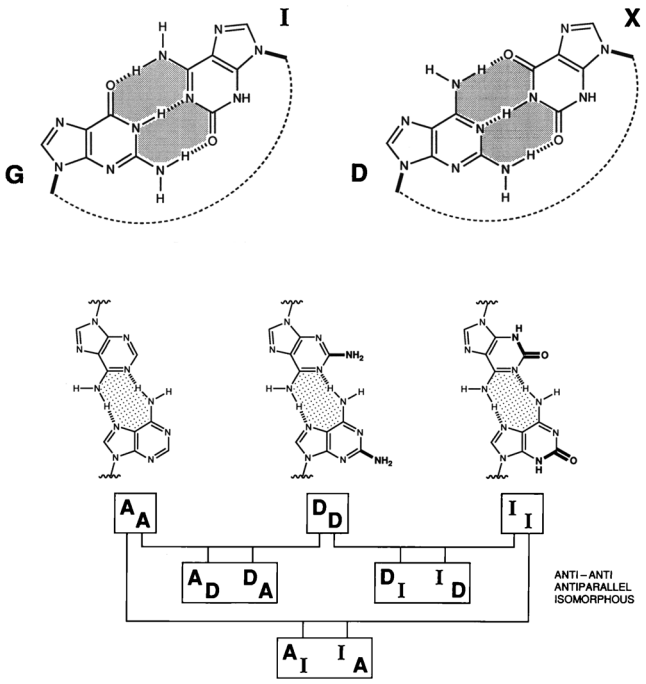


Figure 28. Constitutional types of the purine–purine base pairs present in the duplexes of Figure 27.

with the conformational rigidity of a pyranose chair relative to the high flexibility of a furanose ring.

The phenomenon of strand pre-organization, documented through thermodynamic data among base-pairing systems for the first time in the homo-DNA series, proved to bear fruit in a field of medicinal chemistry that was gaining momentum at about that time. In research on artificial "anti-sense" oligonucleotide systems, using the pre-organization criterion as a planning aid in the search for new oligonucleotide systems made it possible to foresee the structure of oligomers that would show stronger base-pairing than RNA and DNA. The concept was applied for the first time by Christian Leuman[103,104] in his "bicyclo-DNA" project.[105] Probably its most impressive success is the role it may have played in initiation of the synthesis of LNA ("locked nucleic acid"; Figure 24) by Wengel.[94] Herdewijn, who had—independently of the ETH group—also synthesized homo-DNA oligomers,[88,106] made the important discovery of the novel pairing system known as HNA ("hexitol-nucleic acid") (Figure 24).[95] This is a constitutional isomer of homo-DNA that undergoes efficient cross-pairing with RNA and DNA in the Watson–Crick mode in an antiparallel strand arrangement, demonstrating that a six-membered ring structure of a sugar unit in oligonucleotides does not as such rule out formation of a helical duplex structure of the natural type. Further information on this point comes from Herdewijn's comprehensive studies of the chemistry of pyranose oligonucleotides,[107,108] which included the synthesis of so-called α-homo-DNA.[109] This anomer of the original homo-DNA[110] undergoes, in contrast to the latter, pairing with RNA, yet

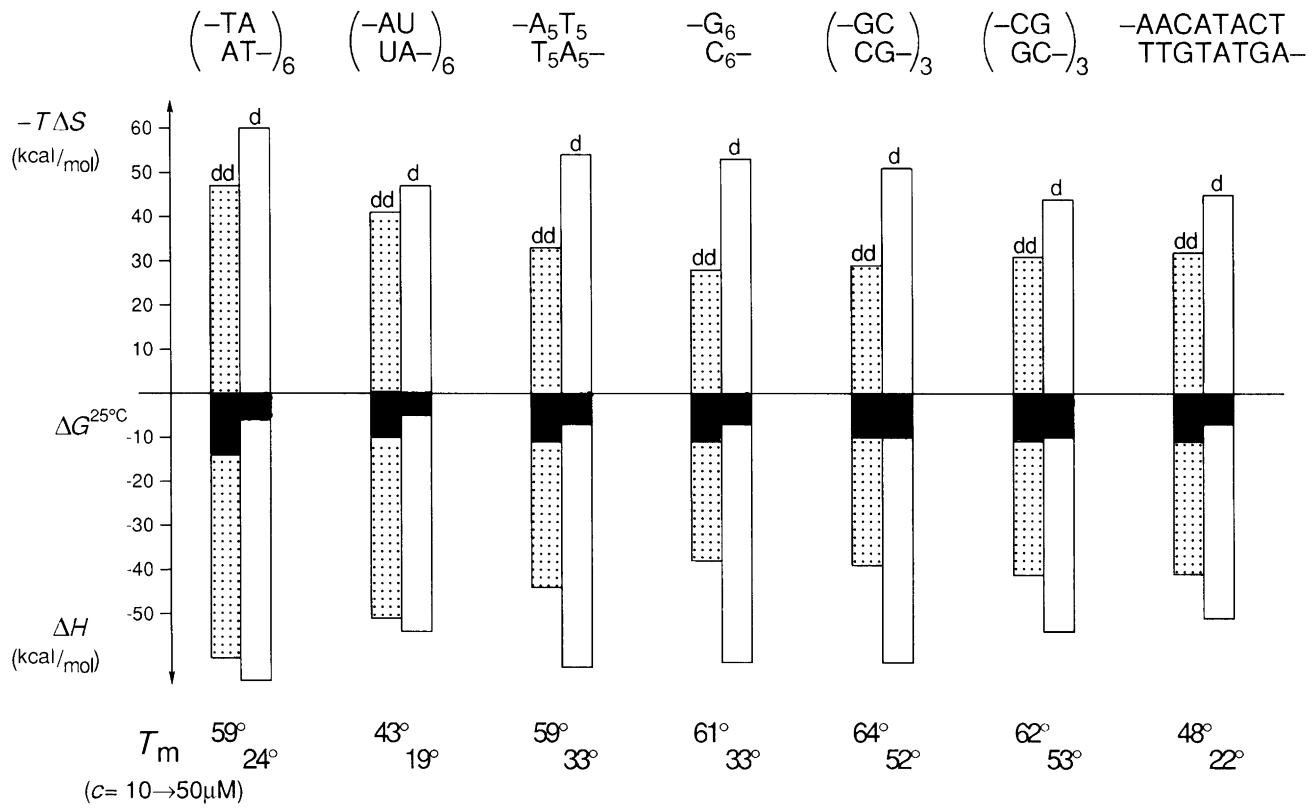


Figure 29. Comparison of thermodynamic data for duplexing of single strands with the same base sequences in the homo-DNA and DNA series. The almost uniformly more negative ΔG values in the homo-DNA series are due to corresponding differences in ΔS values.[99,100]

only weakly so, and in a parallel, not antiparallel, strand orientation.[111,112]

Oligonucleotide systems like bicyclo-DNA, HNA, and LNA, as well as numerous other pairing systems, the synthesis of which had been stimulated by the pre-organization criterion,[94,104,107,108,111,113–115] may be relevant in a chemical, biological or medical context, yet they have no direct connection with the etiological problem; they are all artificial systems with molecular structures that do not fulfill the demanding criteria (cf. above) for "potentially natural" nucleic acid alternatives.

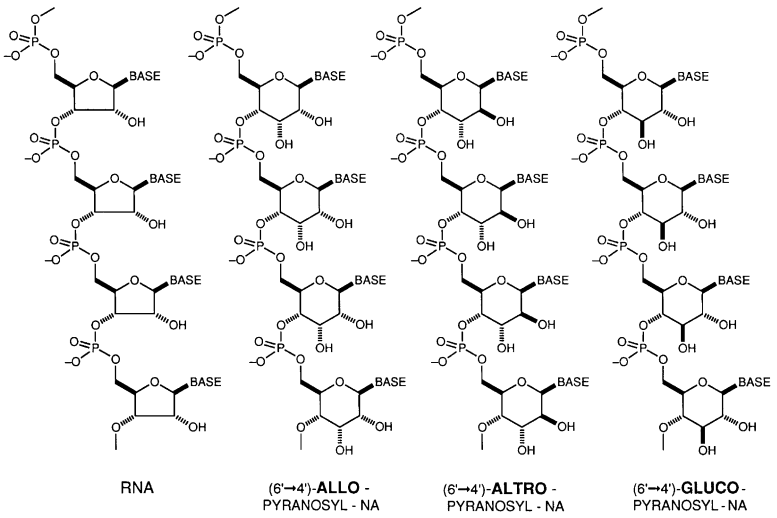


Figure 30. Constitution, configuration and chirality of the experimentally investigated potential RNA alternatives in the allose, altrose, and glucose series.

4.4. Hexopyranose Nucleic Acids

Parallel to the studies on pairing characteristics of the model system homo-DNA, we had begun work that should have led us to "proper" nucleic acid alternatives, that is, oligonucleotides with sugar modules recruited from the family of the natural aldohexoses. All other constitutional elements of the repeating units of oligomers (glucosidic attachment of the nucleobases in the equatorial position, (6′→4′) instead of (5′→3′)-phosphodiester bridges) should correspond as nearly as possible to those of RNA. Of the total of eight diastereomeric hexopyranoses, only two are to be considered if one wishes oligonucleotides derived from them to have backbone structures and pairing conformations corresponding to those of homo-DNA. These are D-allose and D-altrose. Only with these two isomers will the nucleobase and the two backbone substituents assume an equatorial, and the free hydroxy group at position C-3′ an axial conformation. The latter is required, because only so will this hydroxy not cause steric distortion of the pairing conformation, as it undoubtedly would were it in an equatorial position. Our first choice fell to D-allose, since among all the hexoses this is the one most closely related configurationally to ribose, with respect to both relative and absolute configuration. The *rac*-allose derivative was the principal product in the hexose product fraction of aldolization experiments with glycolaldehyde phosphate, just as was the case for the ribose derivative in corresponding aldolizations in the presence of formaldehyde (cf. above).[60,70,116] Oligonucleotide variants based on D-altrose and D-glucose were added to the project later (Figure 30), the former in response to observing that the configuration of a hydroxy group at position C-2′ is critical to the pairing capability of such oligonucleotides, and the latter because glucose is supposedly the thermodynamically most stable among the aldohexoses.

Since experimental observations regarding the chemistry of allo-, altro-, and glucopyranosyl-(4,6)-oligonucleotides are (regrettably) available in their entirety only in dissertations[102,117–121,125] and postdoctoral reports,[122–124] they will be treated here in somewhat greater depth; only excerpts have previously appeared in print.[4,15,84,126–129] Our studies of these three oligonucleotide systems were, incidentally, less comprehensive than those dealing with homo-DNA, because their properties had soon made it clear that the systems would have had no chance of being evolutionary competitors of RNA.[130]

Whereas the synthesis of homo-DNA oligonucleotides[98] could rely upon DNA synthesis as a model, this was not so in the synthesis of hexose analogues of RNA, in which two vicinal hydroxy groups, not involved in phosphodiester bridging, pose a protective group problem. Its solution for the two vicinal *cis*-hydroxy groups in positions C-2′ and C-3′ of the allose unit was the isopropylidene acetal protective group which, though acid sensitive, turned out capable of withstanding hydrolytic cleavage under the (milder) acidic conditions employed in removal of the 5′-trityloxy protective group in solid-phase oligomer synthesis.[118–120] In subsequent work in the altrose and glucose series it turned out, surprisingly, that the corresponding protective-group role could be filled in both series by the benzoate group, to be cleaved under basic conditions.[102,121] Apart from these measures necessary for taking care of system-specific protection problems, the methodology employed for oligomer synthesis was the same as that used in the homo-DNA series.[98]

The experimental data shown in Figure 31 attest to how fortunate, at the time, our decision had been to first study the model system homo-DNA before becoming involved in investigations of hexopyranose nucleic acid alternatives that we supposed to be "potentially natural". Comparison of the melting temperatures of duplexes of the self-complementary base sequence 6′- (or 5′)-CGCGAAUU(TT)CGCG (Figure 31) in the DNA-, homo-DNA and allose-NA series (Figure 31) reflects our general finding that purine–pyrimidine pairing in allose-NA, if it occurs at all, is very much weaker than in homo-DNA. Dependence of the melting temperature (T_m) on pH in allose-NA is the inverse of that in homo-DNA; increase in pairing strength with increasing acidity such as that observed for allose-NA points to a pairing mode that presupposes protonation of the cytosine partner, as is the case for its pairing in Hoogsteen mode. Linear dependence of the T_m-value from pH for the duplex of the alternating sequence 6′-allo-$(GC)_5$ supports this assumption

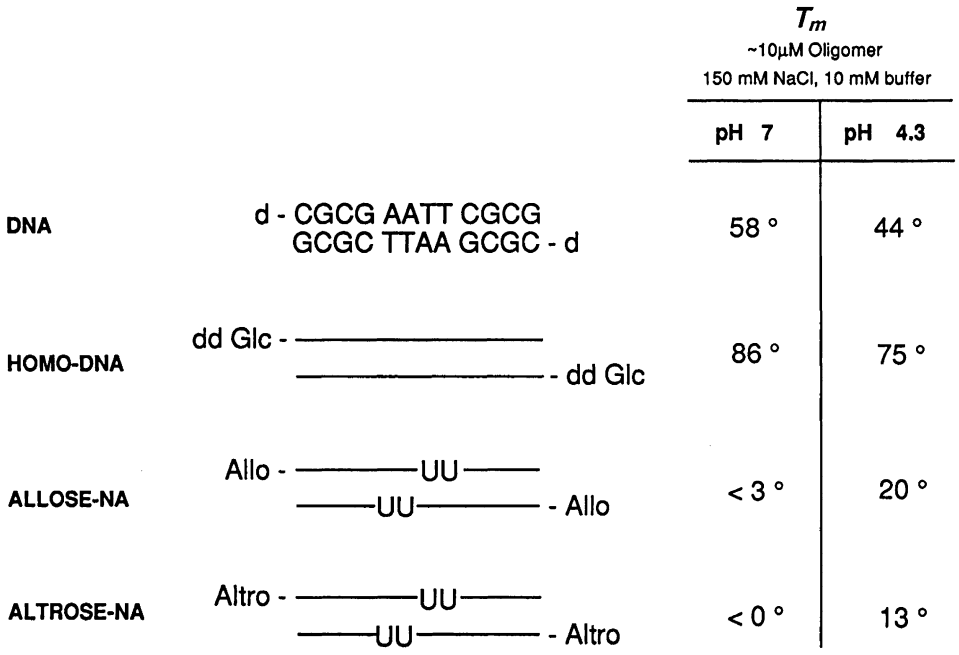


Figure 31. Comparison of the pairing strengths of duplexes of an antiparallel, self-complementary dodecamer sequence in the DNA, homo-DNA, allose-NA, and altrose-NA series.[118,119]

(Figure 32). On the other hand, the pH dependence of the T_m-value in the case of the block sequence 6′-allo-(G_5C_5) is more complicated: at a pH of 5–5.5 it clearly indicates an abrupt change from a pH-independent to a pH-dependent pairing mode. The (weak) pairing observed in the neutral region quite probably represents purine–purine- rather than purine–pyrimidine pairing.[119]

Extensive efforts were directed toward establishing the existence of an adenine–uracil pairing for the formally self-complementary sequences 6′-allo-$(AU)_6$, 6′-allo-A_6U_6, and 6′-allo-U_6A_6, as well as the sequence pairs allo-A_{12} and allo-U_{12} and the (antiparallel-complementary) sequence pair 6′-allo-$U_2A_4UA_2U_3$ and 6′-allo-$A_3U_2AU_4A_2$ through observation of melting temperatures and temperature-dependence of CD spectra. These studies led to the conclusion that in allose-NA such pairing must be exceedingly weak, if it exists at all, and—if it does—heavily sequence-dependent, since it

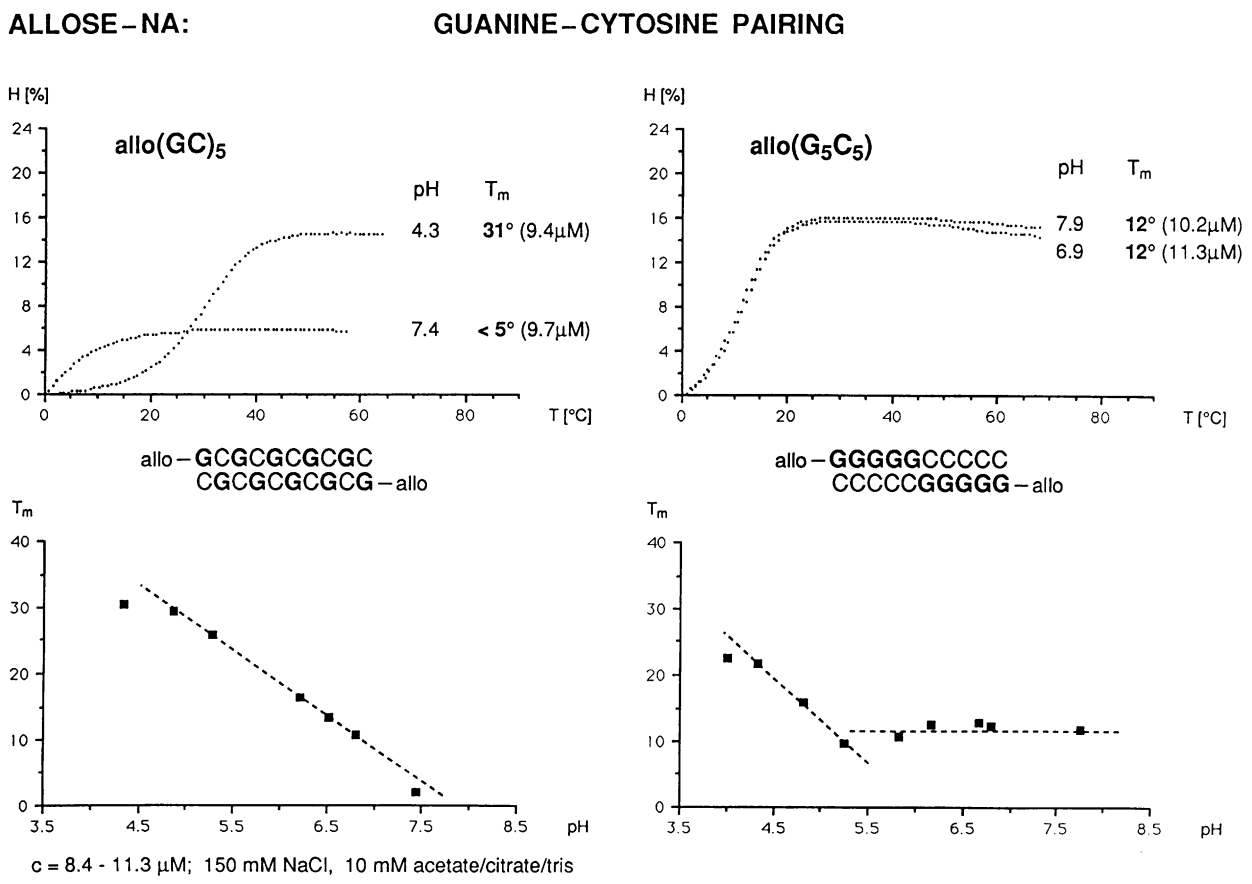


Figure 32. Sequence-dependence of the pH-dependence in pairing behavior of two isomeric guanine- and cytosine-containing allose-NA decamers: The "RNA of the hexose series" is not a Watson–Crick pairing system.[118,119]

would need to compete with adenine–adenine pairing (cf. below).[118,119]

Homo-basic purine–purine sequences in allose-NA behave rather like corresponding sequences in homo-DNA: they pair with themselves, but much more weakly. It had been established that in homo-DNA the guanine–guanine and adenine–adenine pairings in such sequences utilize reverse-Hoogsteen mode in antiparallel strand orientation.[99,100] But whereas, for example, the structure of temperature-dependent CD spectra and T_m-values for self-pairing homobasic A_{12} sequences are very similar in the homo-DNA ($T_m = 43\,°C$) and altrose-NA ($T_m = 39\,°C$) series (cf. below), the CD spectrum of the same sequence for allose-NA ($T_m = 29\,°C$) shows a drastically different structure. Which pairing structure this difference mandates (Hoogsteen-pairing in parallel-strand arrangement or *syn*-arrangement of the adenine cores in antiparallel reverse-Hoogsteen pairing) remains undetermined. The combination of guanine with isoguanine in sequences of allose-NA shows a much more complex behavior than with such sequences in homo-DNA.[121,119] Under standard conditions, mixtures of equal equivalents of the homobasic sequences allo-G_8 und allo-$IsoG_8$ lead to mixtures of triplex aggregates with, for some of them, unusually high T_m values. Self-complementary sequences such as 6′-allo-G_4IsoG_4 and 6′-allo-$IsoG_4G_4$ appear to form higher associates,[121] whereas complementary sequence pairs of G/IsoG-containing hexamers and octamers unite to give apparently "normal" duplexes, albeit low-melting ones. Experimentation with allose-NA sequences at concentrations higher than what is customary for UV and CD spectroscopic characterization proved generally difficult as a consequence of solubility problems; structure determinations by NMR spectroscopy seemed beyond reach, not least due to these problems.

In summary: substituting D-allopyranose for D-ribofuranose in RNA's oligonucleotide structure results in almost total loss of the capacity for informational Watson–Crick base pairing. This is in stark contrast to homo-DNA, in which this capacity—albeit in, so to speak, a different "language"—is retained, and indeed bolstered. The more pronounced tendency toward purine–purine pairing in homo-DNA relative to RNA is retained in allose-NA, even if to a significantly lesser extent. With such results it became important to pursue the question of why and how the two additional hydroxy groups of allose-NA compromise the pairing capability of homo-DNA to such a degree. An answer arose out of the pairing characteristics of 2′- and 3′-deoxy derivatives of allose-NA, which were synthesized for this specific purpose (Figure 33).[122]

Model considerations based on NMR structural data for homo-DNA[96] (Figure 26) had led us to suspect that an equatorial hydroxy group in position C-2′ of the homo-DNA duplex structure would encounter steric hindrance from the nucleobase of the adjacent downstream nucleotide unit. Experimental T_m data for self-paired duplexes of the homobasic adenine hexamer sequences of homo-DNA, allose-NA, and their 2′-deoxy as well as 3′-deoxy analogues are in agreement with this interpretation (Figure 33). These data convincingly show that the extensive loss in base-pairing

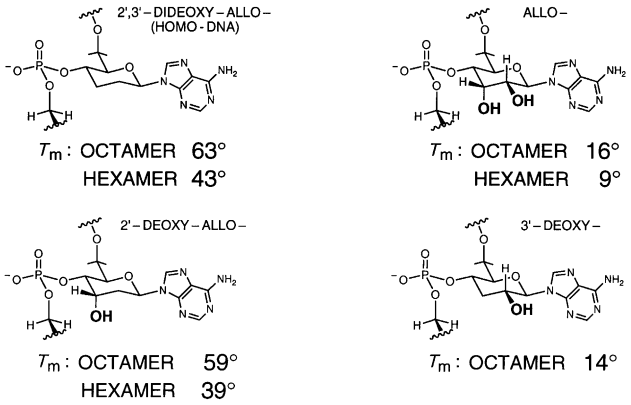
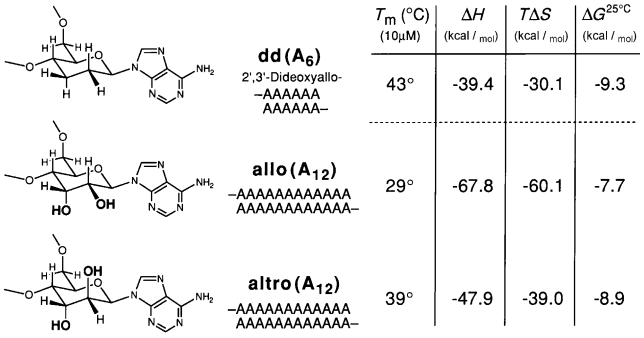


Figure 33. Self-pairing of hexopyranosyladenin-oligonucleotides ($c \approx 10$ μM, 0.15 M NaCl, pH 7, 260 nm). Comparison of the pairing strengths (in reverse-Hoogsteen mode) of (A_n)-oligomers ($n = 8$ and 6) in the allo-, 3′-deoxyallo-, 2′-deoxyallo-, and 2′,3′-dideoxyallo (= homo-DNA) series: The equatorial hydroxy group in the 2′ position is the source of steric hindrance in the pairing conformation of systems of the homo-DNA type.[122]

power of allose-NA relative to homo-DNA is due to the equatorial hydroxy group in the 2′ position, and not the axial one in the 3′ position. This interpretation eventually gained explicit support from the molecular dimensions in Egli's X-ray structure of a homo-DNA duplex.[97]

Our observation on allose-NA and interpretation of its behavior led us to altrose-NA, which differs from allose-NA solely in the configuration of the critical hydroxy group at position C-2′. The altrose variant promised to bring us closer to ultimately encountering a "potentially natural" hexapyranose alternative to RNA, one exhibiting a base-pairing efficiency similar to that of homo-DNA. But to the extent it was put to the test, this turned out not to be the case. In the altro-series, Watson–Crick pairing of the self-complementary dodecamer sequence of Figure 31 is as weak as in the allo-series—if, once again, it exists at all. On the other hand, homobasic adenine duplexes of altrose-NA are indeed somewhat more stable than is the case in allose-NA, but nonetheless still lower melting than in homo-DNA (Figure 34).[102] According to the temperature-dependent CD spectra, their pairing structure corresponds to that of homo-DNA, as the axial position of the 2′-hydroxy would lead one to expect, and



	T_m (°C) (10μM)	ΔH (kcal / mol)	$T\Delta S$ (kcal / mol)	$\Delta G^{25°C}$ (kcal / mol)
dd(A_6)	43°	-39.4	-30.1	-9.3
allo(A_{12})	29°	-67.8	-60.1	-7.7
altro(A_{12})	39°	-47.9	-39.0	-8.9

Figure 34. Self-pairing of hexopyranosyl-oligonucleotides: thermodynamic data. Comparison of pairing strengths in reverse-Hoogsteen mode of (A_n)-oligomers in the homo-DNA, allose-NA, and altrose-NA series.[117]

not to that of allose-NA. Guanine–cytosine pairing in the self-complementary sequences 6′-altro-$(GC)_5$ and 6′-altro-(G_5C_5) is, in contrast, just as weak as in allo-series, and the pH dependence of the T_m-values is largely analogous to that shown in Figure 32 for allose-NA.[123] In retrospect, we believe the reason for this unexpected pairing weakness to lie in steric clash between the axial 2′-hydroxy group and the nucleobase of the same nucleotide unit, thus forcing the nucleosidic torsion angle to deviate from the value of about −120° in homo-DNA. Indeed, from the X-ray structure of the adenosine of the altro-series, one observes a nucleosidic torsion angle of −150.3°[102] (Figure 35). The way such a rotation around the nucleosidic bond relates to the inclination between backbone- and base-pair axes is indicated in Figure 35; such rotations affect the distance between base pairs and, therefore, the efficiency of base stacking.[102]

What, after all these observations, the pairing prognosis based on conformational analysis led us to anticipate in advance with respect to the pairing ability of glucose-NA (with both free hydroxy groups equatorial) was unambigously confirmed experimentally: in oligonucleotide strands with D-glucose as sugar unit and adenine and uracil as nucleobases, UV and CD spectroscopic methods failed to reveal either adenine–adenine or adenine–uracil pairing (Figure 36).[120]

The answer to the originally posed question "Why pentose and not hexose nucleic acids?", based on these experimental observations, is straightforward: RNA variants featuring, rather than a D-ribofuranose, one of the three hexopyranoses D-allose, D-altrose, or D-glucose as sugar module could not have competed with RNA in an evolutionary selection process, because these variants lack the functional prerequisite for a genetic system (Figures 36 and 37).[130] Although our chemical experience with these systems is limited to relatively short base sequences, and one might anticipate more stable complexes with greater sequence length, the extreme sequence-dependence of the (weak) base-pairing observed with short sequences is not compatible with the kind of information transfer we know from natural nucleic acids. The latter demands the greatest possible invariance of pairing strength with respect to variance in base sequence within a given set of nucleobases.

	AT WATSON-CRICK	GC WATSON-CRICK	AA REV. HOOGSTEEN
homo-DNA	strong	strong	strong
	(stronger than DNA)		
ALLO-pyranosyl	very weak	weak	weak
	(much weaker than RNA)		
	very weak	—	weak
	(similar to ALLO-pyranosyl)		
	strong	—	strong
	(similar to homo-DNA)		
ALTRO-pyranosyl	very weak	weak	weak
GLUCO-pyranosyl	none	—	none

Figure 36. Qualitative survey of pairing properties of hexopyranosyl oligonucleotides ("–" means: not experimentally investigated).[117–123]

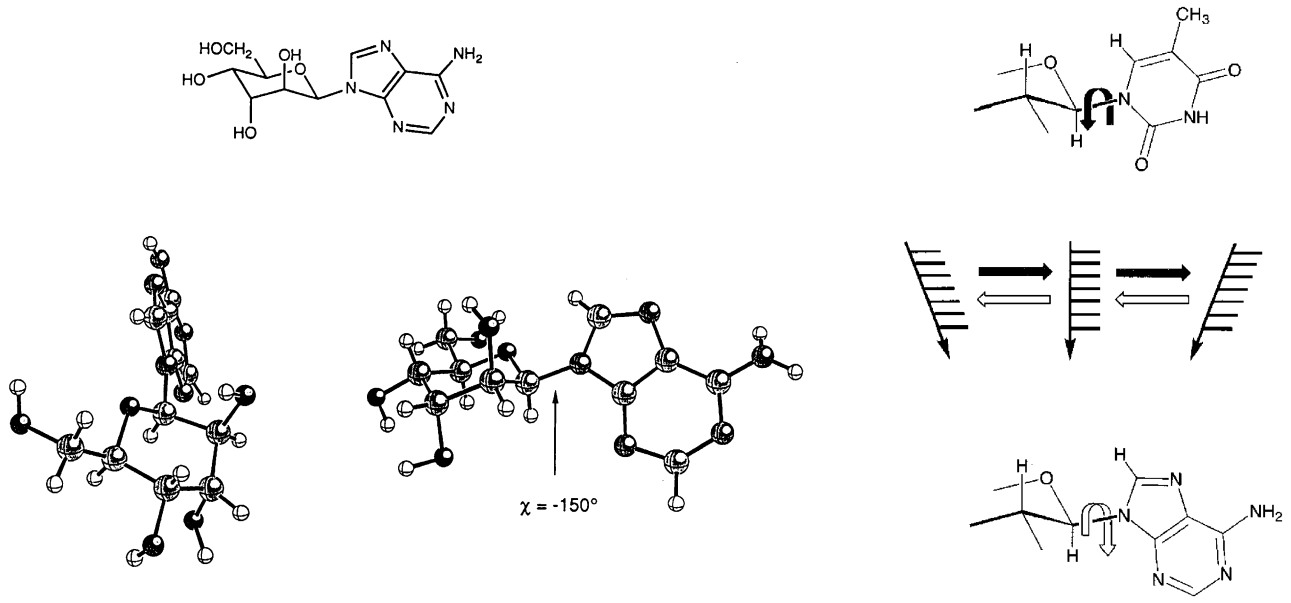


Figure 35. X-ray structure of the "adenosine of the D-altrose series",[102] and schematic illustration of dependency between nucleosidic torsional angle and inclination of the oligomer backbone relative to the nucleobase.

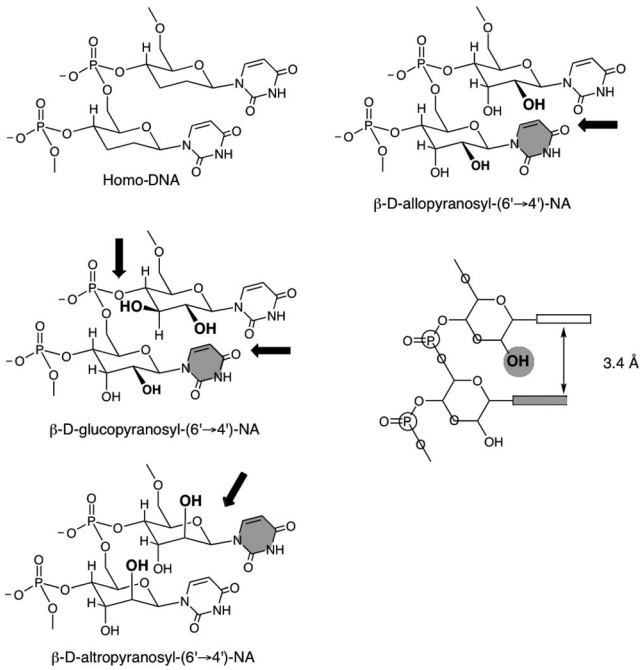


Figure 37. Qualitative overview of steric hindrances relevant to pairing in pairing conformation of the homo-DNA type in hexopyranose-NA oligomers.

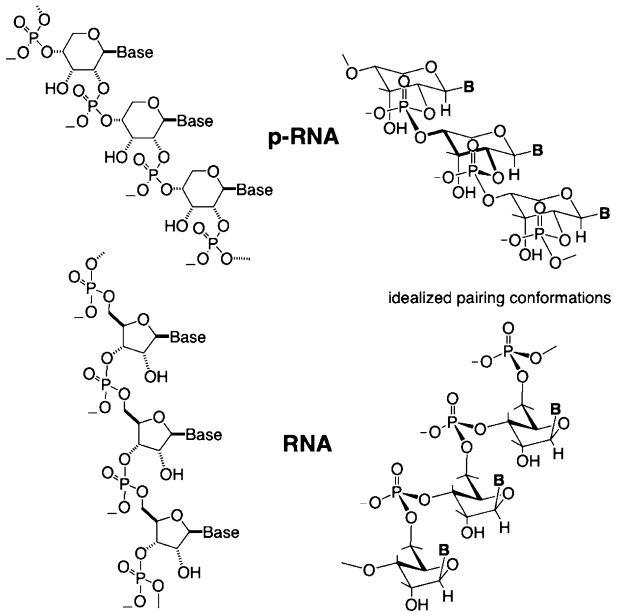


Figure 38. Chemical structure and predicted pairing conformation for the pyranose form of RNA ("p-RNA").[133]

4.5. Pyranose Form of RNA ("p-RNA")

Rationalization of the drastic differences in pairing characteristics between the model system homo-DNA and its "potentially natural" dihydroxy analogues may best be rendered as a paraphrase of a legendary exclamation:[131] "Too many atoms!" The two extra hydroxy groups in the three hexopyranose nucleic acid systems investigated interfere sterically with and distort what in homo-DNA is the pairing conformation; this insight led us to turn back in our work from hexoses to pentoses. The question was raised whether there existed an RNA alternative, one capable of pairing, which contained ribopyranose units in place of ribofuranose as a sugar building block. Whereas the structure type of a (5′→3′)-phosphodiester-linked ribofuranose module can be translated with ease into that of a (6′→4′)-linked hexopyranose module, such a translation into a ribopyranose module requires a more drastic constitutional change: the sole possibility for maintaining 6-center periodicity of the oligonucleotide backbone, at the time regarded to be essential,[132] was phosphodiester bridging via the hydroxy groups at positions C-4′ and C-2′ of the ribopyranose ring (Figure 38). A qualitative conformational analysis of this type of structure, adopting the procedure that had been successful for the pairing prognosis of homo-DNA, revealed a total of nine most stable conformations for the ribopyranose-(4′→2′)-phosphodiester unit, of which a single one is repetitive.[133] Therefore, this single repetitive conformation was predicted to represent the pairing conformation of a hypothetical Watson–Crick base-pairing system that would constitute a pyranose form of RNA ("p-RNA"). Models of duplexes of this novel oligonucleotide system clearly suggested a largely linear structure with a strong backbone inclination oriented in the sense opposite to that in homo-DNA (Figure 38). One element of uncertainty for such a structure was the necessity that the phosphodiester group assume a synclinal/antiparallel conformation; it could not assume a doubly synclinal conformation as in RNA, which had initially been judged (see above) to be the lower in energy between the two possibilities.[90]

An alternative nucleic acid structure, constructed from the same sugar module as RNA, but in an isomeric form, and one that, from a chemical perspective, could well have been assembled by reaction steps of the same kinds as those that must have led to RNA, appeared of particular interest with respect to an etiology of the RNA structure. Thus, our experimental involvement in the "p-RNA" project became correspondingly thorough (Figure 38).[133–141]

The pyranose form of RNA turned out to be a superb pairing system that displays highly instructive and very special pairing characteristics. The structure of the duplex of the self-complementary sequence 4′-pr-CGAATTCG was established through NMR spectroscopy by Bernhard Jaun and co-workers,[136,142,169] and modeled using molecular dynamics by Romain Wolf.[136] The data reveal a left-handed, weakly helical ladder structure with antiparallel strand orientation, and a strong inclination between the backbone- and base-pair axes in an orientation opposite to that of homo-DNA (Figures 39 and 40).[139] This type of structure corresponds to a remarkable degree to that predicted for a p-RNA duplex using conformational analysis based on an idealized conformation, as delineated above.[133]

The pairing characteristics of p-RNA differ sharply from those of RNA and DNA, above all as a consequence of the strong backbone inclination of p-RNA's ladder structure.[133,134,141] Duplexes here are paired in Watson–Crick mode, and consistently melt at higher temperatures than

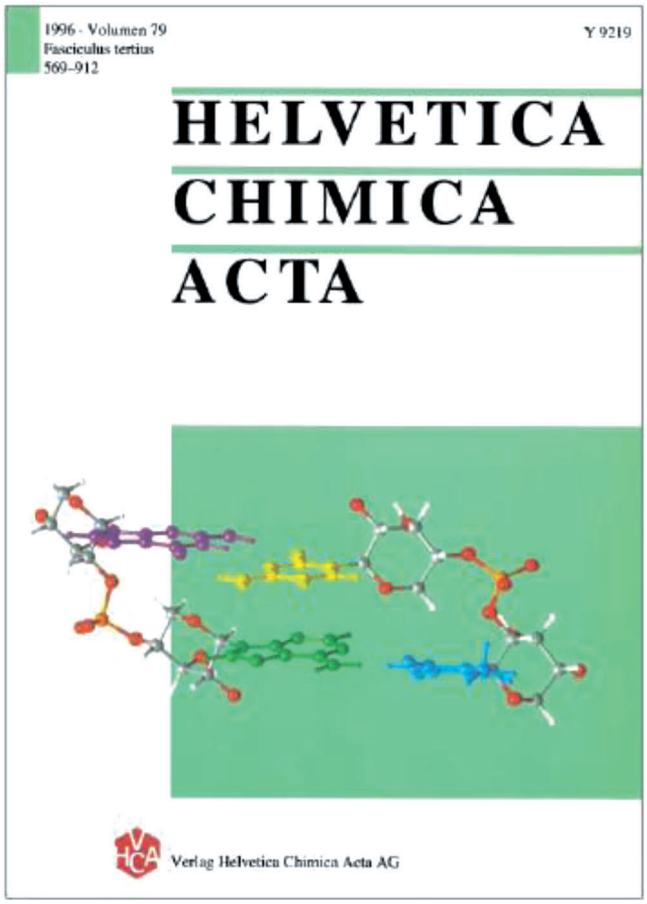


Figure 39. Pairing conformation of *p*-RNA as observed with NMR spectroscopy.[136] Base stacking is of the inter-strand type.

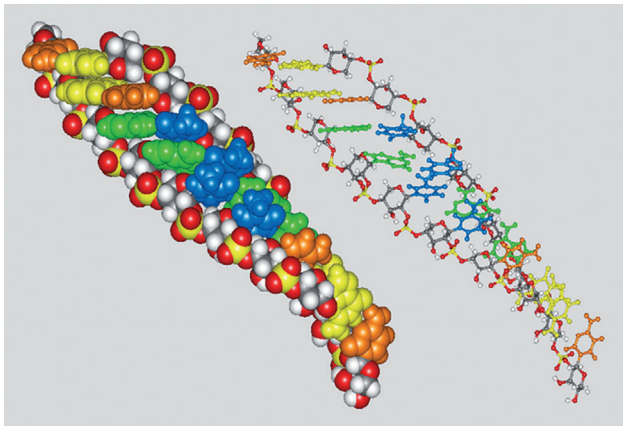


Figure 40. The structure (as established by NMR spectroscopy) of the p-RNA duplex of the antiparallel self-complementary base sequence 4′-CGAATTCG is a weakly left-helical ladder structure with a strong backbone inclination.[136, 169]

corresponding duplexes in the natural series. As with homo-DNA, base pairing in p-RNA is orthogonal to that in RNA and DNA, and orthogonal as well to base-pairing in homo-DNA. In an important difference relative to both homo-DNA and DNA, p-RNA base sequences—at least in the range of concentrations and sequence-lengths studied—show no pairing in any Hoogsteen mode, neither between purines and pyrimidines, nor between purines and purines.[133, 134, 141] As is the case in homo-DNA, the p-RNA preference for forming duplexes with antiparallel (as opposed to parallel) strand orientation is much more pronounced than with DNA: a hexadecamer sequence composed of adenines and thymines in the p-RNA series, one that undergoes pairing with the antiparallel-complementary partner strand to form a duplex melting as high as 60 °C (under standard conditions), did not display any pairing with the strand of a parallel-complementary base sequence.[141] This is in agreement with the expectation that Watson–Crick base-pairing systems with strongly inclined backbones should for geometric reasons be incapable of undergoing base-pairing in parallel strand orientation,[99] a prognosis that has experimental support in the homo-DNA series as well.[99]

Another consequence of the strong backbone inclination in the p-RNA ladder structure is interstrand base stacking. This is a structural aspect whose significance for the properties of pairing systems we fully appreciated only in the course of work on p-RNA, although the backbone structure in homo-DNA duplexes is also strongly inclined.[96, 97, 140] In interstrand stacking, neighboring nucleobases stack pairwise, whereby the two bases involved do not belong to the same strand; in contrast, intrastrand stacking operates between neighboring nucleobases on the same strand, and is not limited to interactions between pairs of bases. In the standard B-type structure of DNA, base stacking is strictly intrastrand, whereas a hybrid of the two stackings characterizes an A-type DNA (as well as RNA); in p-RNA duplexes the stacking is purely interstrand (Figure 39). In principle, one would expect interstrand base stacking to contribute more efficiently to the stabilization of a duplex than intrastrand stacking, since the former vanishes with "melting" of the duplex structure, and because this need not necessarily be the case with intrastrand stacking. The expectation that the contribution of interstrand stacking to the stability of a p-RNA duplex would be greatest for purine–purine stacking and least with pyrimidine–pyrimidine stacking, and that stability differences between isomeric base sequences could largely be rationalized in this way, has been confirmed to a remarkable degree (Figures 41–43).[137, 139, 140] The same applies to the expectation that so-called "dangling bases" in a p-RNA duplex would be stabilizing only if positioned at the 2′-end of a strand, and that this effect would be stronger when the dangling base is a purine rather than a pyrimidine (Figures 43 and 44). The reliability of these relationships is strongly supported by observation of their systematic reversal in corresponding complexes of homo-DNA, in which backbone inclination has the reverse orientation (Figures 41–44).[140]

Not a small part of our experiments on p-RNA had been dictated by reasoning about etiological aspects, such as what requirements a p-RNA would have had to fulfill as a primordial genetic system. A sideways glance at the concept of an "RNA world" (see below) reminds us that such a system would need to have had the potential for expressing itself in a wide diversity of structural forms linked to a corresponding diversity of base sequences. Not least for this reason we have tested whether p-RNA is capable of forming hairpin com-

		p-RNA		homo-DNA	
Sequence Motifs	Self-complementary Sequences	T_m °C	ΔΔH kcal/mol	T_m °C	ΔΔH kcal/mol
py_n-PU_n	-TTTTAAAA	40	+11.8	34	+3.8
PU_n-py_n	-AAAATTTT	27		38	
	-TTTTTAAAAA	54	-4.3	45	-3.2
	-AAAAATTTTT	43		50	
	-CCCGGG	68	+7.2	49	-1.8
	-GGGCCC	58		53	
$(py\text{-}PU)_n$	-TATATATA	40	+5.0		
$(PU\text{-}py)_n$	-ATATATAT	38			
	-TACGTA	36	+4.9	27	-0.6
	-ATGCAT	26		37	
	-CGCGCG	65	+7.4	55	-2.0
	-GCGCGC	62		53	

Figure 41. Comparison of pairing strengths in duplexes with antiparallel complementary base sequences from the p-RNA and homo-DNA series: Opposing orientations of backbone inclination lead to opposing influences of interstrand base stacking on duplex stability (cf. Figure 42).[141] Estimated error for $\Delta H = \pm 10\%$; 10 μM, 150 mM NaCl, 10 mM Tris·HCl, pH = 7.0.

plexes. In fact, despite the lower flexibility to be assumed for its pyranose sugar backbone, formation of hairpin complexes, even with short sequences, was observed to occur with an ease comparable to that of corresponding RNA sequences (Figure 44).[137]

Whereas in sharp contrast to homo-DNA no (homochiral) purine–purine self-pairing is observed in the reverse-Hoogsteen mode in p-RNA (see above), extensive studies about the chiroselectivity of p-RNA base-pairing revealed the existence of pairing combinations in other modes where purines are involved. First, what applies to homo-DNA (Figure 28) is also true for p-RNA: guanine and isoguanine form a Watson–Crick pair, albeit one that is not as strong as in homo-DNA. Observations made on interactions of homochiral D- and L-strands, both containing homobasic sequences G_n ($n = 6, 8$), with homochiral D-strands of the homobasic sequences C_n and I_n (I = isoG), led us to conclude that in p-RNA, homochiral base-pairing in Watson–Crick mode is strictly enantioselective.[135] Also observed was heterochiral purine–purine and purine–pyrimidine pairing between homochiral strands of opposite sense of chirality, presumably in reverse-Hoogsteen mode. Notable among these interactions is the remarkably high stability of the duplex resulting from heterochiral pairing between homochiral L-I_8 and homochiral D-C_8 strands. Finally, weakly paired duplexes are formed from homobasic L-I_n and homobasic D-I_n sequences, presumably involving two readily accessible tautomers of isoguanine as pairing partners. Such enantiomeric strands may pair with each other in the reverse Watson–Crick mode, thereby forming pseudo-racemates.[135,141]

We have extensively examined[134,138,139,141] the potential of p-RNA sequences to become replicated by way of template-controlled ligation.[143] Promising in this context was the important property of p-RNA not to engage in homochiral self-pairing of purines in reverse-Hoogsteen mode (cf. above), since, in his pioneering work on template-controlled oligomerization of activated mononucleotides in the RNA series, Leslie Orgel[144] had observed that self-pairing of guanine seriously interfered with non-enzymatic replication of RNA by way of withdrawing templates containing guanine sequences from interacting with cytosine-containing ligands. Such interference is not to be expected with p-RNA.

All our observations on replication of p-RNA sequences were made in experiments where activation of the 2′-phosphate group of a nucleotide or oligonucleotide unit consisted in its conversion into the mildly activated five-membered (2′,3′)-cyclophosphate group, using a water-soluble carbodiimide as the dehydrating agent. Template-controlled ligation of ligands activated in this way leads to the proper (4′→2′)-phosphodiester coupling with very high regioselectivity; no isomeric (4′→3′)-phosphodiester coupling is observed. This is because of what can be described as a "fortuitous" arrangement of reaction centers in the ligation complex, where the 2′-hydroxy group of the ligand acting as nucleophile is perfectly positioned in line with the P–O3′ bond of the other ligand's cyclophosphate group, such that nucleophilic substitution of the 3′-oxygen function with inversion at the phosphorous

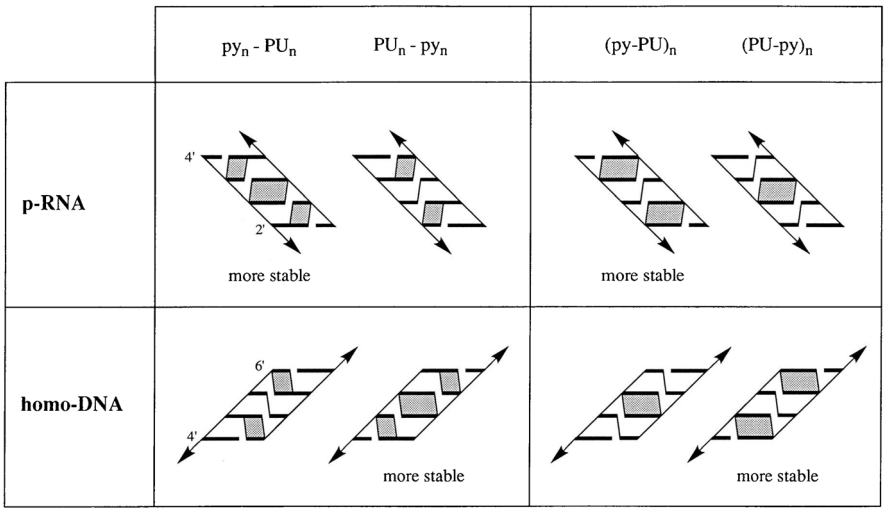


Figure 42. Schematic explanation of the opposing influence of the opposite orientations of backbone inclination for p-RNA and homo-DNA on interstrand stacking (py = pyrimidine, PU = purine), and thus on stabilities of duplexes with identical base sequences.[141]

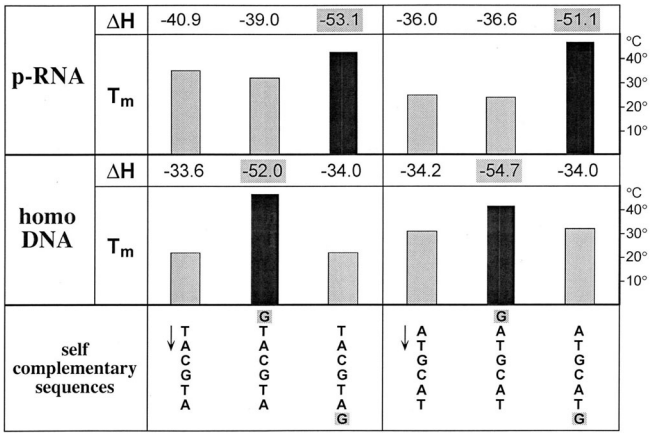


Figure 43. The opposing orientations of backbone inclination in p-RNA and homo-DNA duplexes leads to opposing influence of dangling bases on duplex stability.[140] ΔH in kcal mol^{-1}; solutions: 3.5 μM, 100 mM NaCl, 10 mM Tris·HCl, pH = 7.0.

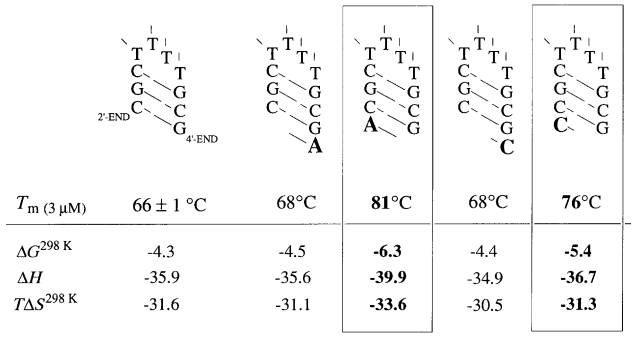


Figure 44. Influence of dangling bases on the stability of p-RNA hairpin complexes.[140,141]

center occurs selectively (Figure 45).[145] In the RNA series, by contrast, an analogous ligation process by way of activation of the (2′,3′)-cyclophosphate group accomplishes, again regioselectively, the non-natural (5′→2′)-coupling.[139,146] This means that non-enzymatic, template-controlled ligation via cyclophosphate activation leads from p-RNA again to p-RNA, but not from RNA to RNA (Figure 46). This very special property of p-RNA is once more a consequence of its characteristic backbone inclination.[147]

The experiment illustrated in Figure 47 is representative of a large number of experiments using octamers as templates and activated tetramers as ligands in order to establish, besides the regioselectivity of cyclophosphate ring opening, the sequence fidelity, sequence dependence, chiroselectivity, and dependence on medium of the replicative ligation reaction in p-RNA.[134,141,139] Under the concentration conditions employed (150 μmol template, 450 μmol tetramer-2′,3′-cyclophosphate, in the presence of LiCl), such ligations proceed slowly (within days) and are limited only (but markedly) by hydrolytic deactivation of the cyclophosphate groups. They are highly chiroselective, i.e., when combining the homochiral D-template with the homochiral enantiomeric ligand, no detectable ligation occurs under otherwise identical conditions. The same is in essence true for the group of diastereomeric ligands in which, in each case, only one of the four nucleotide units has the enantiomeric configuration. One exception is the ligand diastereomer that contains the enantiomeric unit at the 4′-end: this shows only an about 10-fold slower reaction than the homochiral ligand isomer. The coupling of ligands consisting of sequences containing as nucleobases only adenine and thymine proceeds correspondingly more slowly; however, such ligations proceed efficiently when corresponding octamer ligands are reacting on hexadecamer templates.

With hairpin-forming sequences as templates and hexamer ligands of the type illustrated in Figure 48, template-controlled ligations proceed with remarkable ease. It was in the series of such hairpin templates that we launched a search for what might amount to a "ligase" by studying various template and ligand sequences. Hairpin-forming templates might have a chance to reconstitute themselves from a product duplex in the course of the reaction process. However, our search remained unsuccessful; despite extensive efforts in appropriate directions we failed to achieve replication of any such hairpin sequence catalytically, that is, with a turnover number greater than one.[141] Attempts to overcome "product inhibition" by release of the ligation catalyst (= template hairpin) through periodic fluctuation in the temperature of the reaction medium, failed experimentally due to the complexity of the reaction mixture that built up as a consequence of parallel acceleration of hydrolytic deactivation of cyclophosphate groups. It remains an important open question whether ligase activity could have been observed at very much lower template and ligand

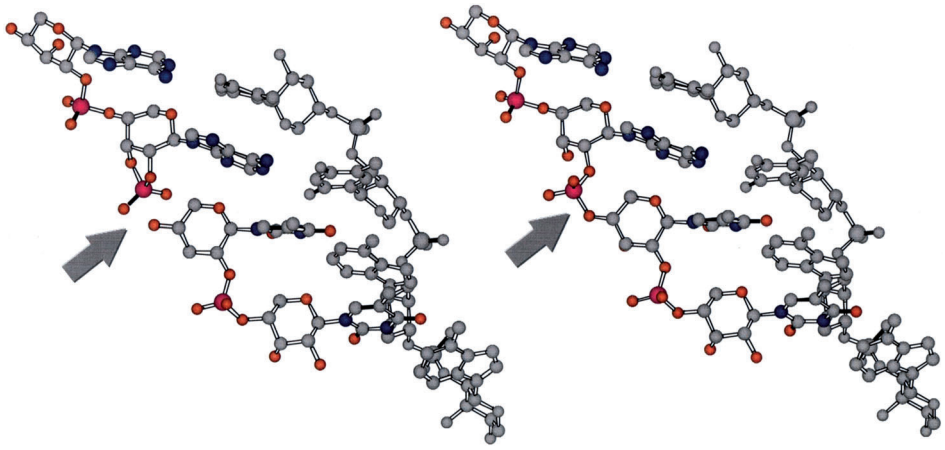


Figure 45. 2′,3′-Cyclophosphate-activated p-RNA ligands at antiparallel complementary RNA templates are structurally preorganized for regioselective ligation to 2′,4′-linked phosphodiester bridging.[145]

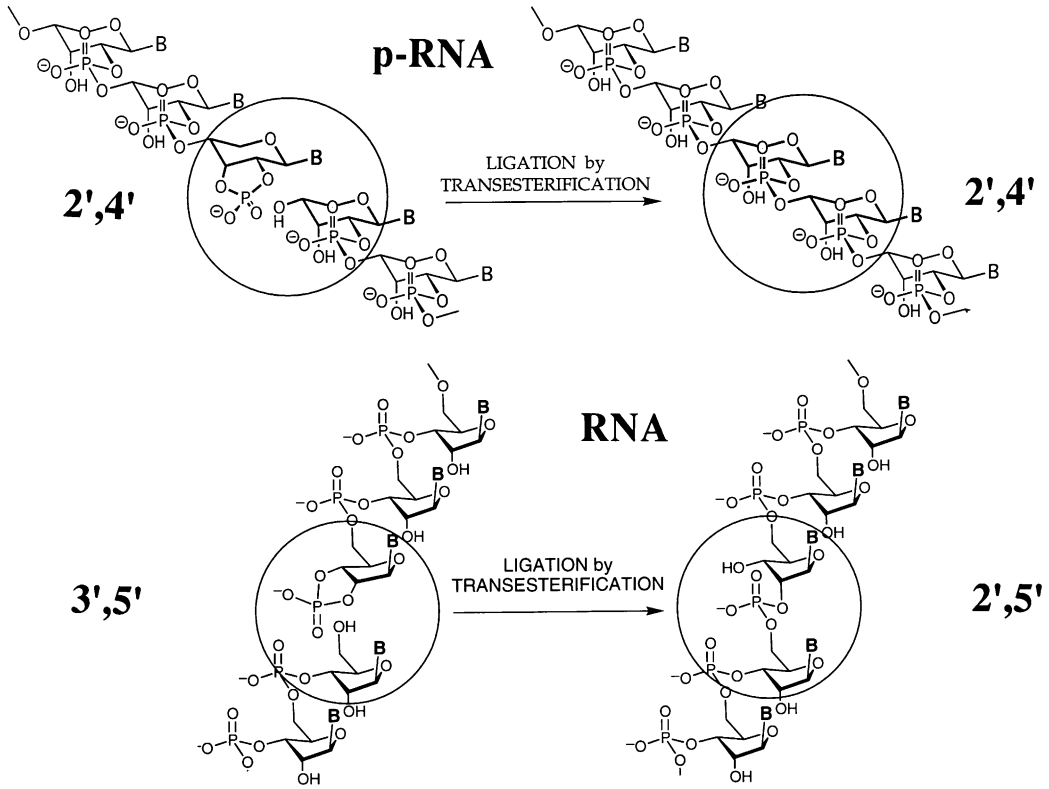


Figure 46. Template-controlled ligation of 2′,3′-cyclophosphate-activated ligands in the p-RNA series leads with high regioselectivity to 2′,4′-coupling of the ligands, in contrast to RNA, where the analogous ligation leads not to natural 3′,5′-coupling, but rather to the isomeric 2′,5′-coupling.[139]

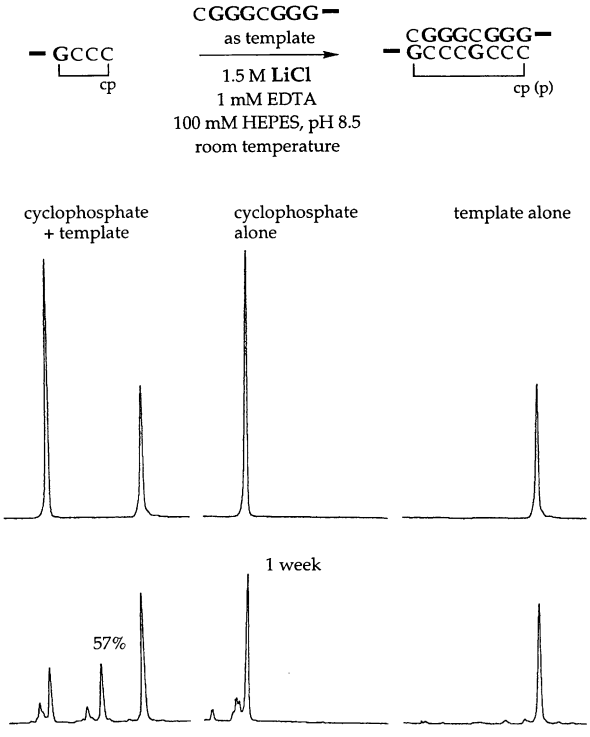


Figure 47. Following a ligation experiment in the p-RNA series by HPLC (together with test experiments). A side reaction is hydrolysis of the 2′,3′-cyclophosphate groups in both ligand and product to corresponding phosphate groups (signals accompanying the primary signals on their left).[139]

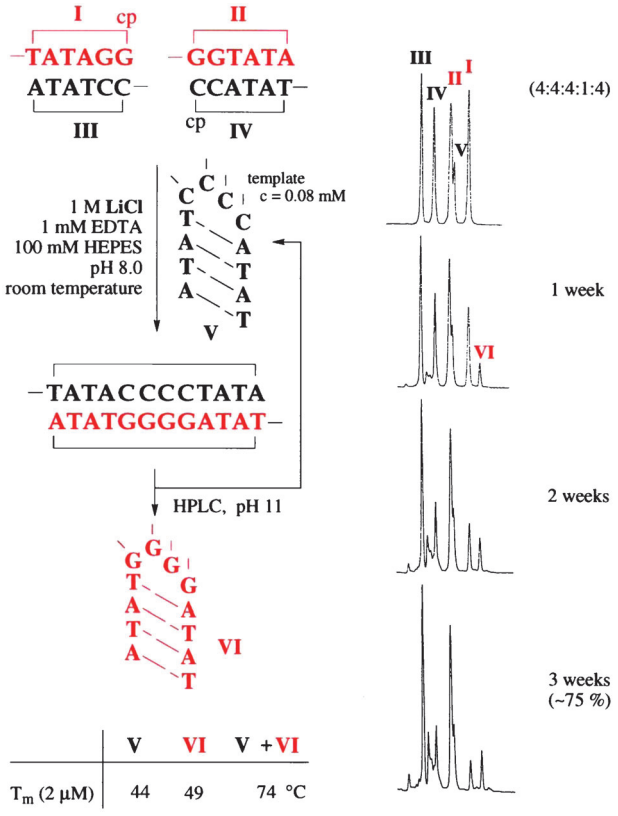


Figure 48. Potentially replicative ligation of p-RNA hexamers to p-RNA dodecamers that form hairpins (turnover number = 1).[141]

concentrations, far outside the analytical range we were able to control.

Perhaps the most noteworthy finding referring to ligation reactivity in the p-RNA series has been the self-assembly of higher oligonucleotide sequences within duplexes through self-templating oligomerization of half-complementary tetramer-2′,3-cyclophosphates.[138,141] In the very simplest case, such higher oligonucleotide sequences are self-complementary regular oligomers of a single half-self-complementary tetramer sequence (e.g., 4′-ATCG-cp; cp = cyclophosphate; Figure 49). They can also be two different regular oligomers

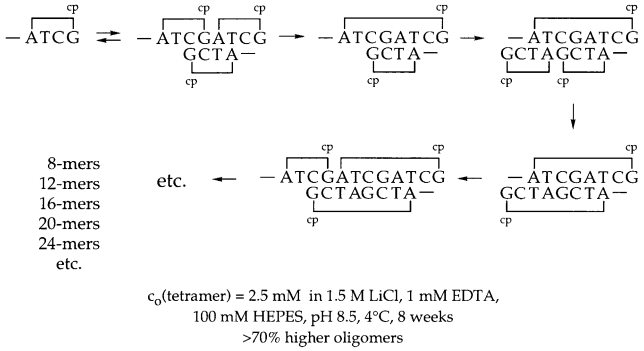


Figure 49. Ligative self-assembly of p-RNA duplexes starting from half-self-complementary tetrameric 2′,3′-cyclophosphates as exemplified by the tetramer pr-ATCG-2′,3′-cp.[138]

that are complementary to each other, made up of two different tetramers that are half-complementary to each other (e.g., 4′-ACGG-cp mixed with 4′-GTCC-cp). In the most general case (not dealt with experimentally) they could be a multitude of pairwise complementary oligomer sequences that (potentially) could arise from a quasi-stochastic co-oligomerization of a set of half-complementary and half-self-complementary tetramer sequences (e.g., 4′-ATCG-cp in a mixture with 4′-GCCG-cp).

Apart from several other tetramers, we studied the process most extensively for the single tetramer sequence 4′-ATCG-cp (Figure 49). Monitoring its oligomerization by HPL chromatography (by recording single strands) indicated a slow but steady formation of oligomers of tetramers of differing sequence lengths, thought of as resulting from the growth of corresponding duplexes, within which these oligomers in both strands assemble cooperatively. Worth noting is the pronounced chiroselectivity of this self-assembly process (Figure 50): of the total of 8 possible diastereomeric enantiomers of a racemic mixture of the activated tetramer,[148] the homo-chiral D-isomer reacts by far the most rapidly. All hetero-chiral diastereomers show barely detectable product formation, again with the exception of the tetramer that bears the enantio-ribose unit in the 4′-position, 4′-LDDD-cp;[149] the latter shows about one-tenth the reactivity of the homochiral tetramer.[150] If one permits oligomerization of the homochiral tetramer to occur in the presence of the heterochiral diastereomers, one finds the oligomerization efficiency to be essentially unaffected. Here again, what hampers progress of the ligation process is hydrolytic deactivation of cyclophosphate groups. Although an expectation like the following is remote from experimentally recorded reality, it is nevertheless worth recognizing that, from an equimolar mixture of altogether eight possible diastereomeric racemates of a half-self-complementary tetramer 2′,3′-cyclophosphate, there would arise a racemoid mixture of duplexes of primarily homochiral oligomer sequences.

The analytical capacity of the HPLC methodology we were employing was far too limited for us to be able to fathom the true potential of this type of reaction, namely the combinatorial self-assembly of higher p-RNA sequences starting from appropriate sets of half-complementary tetramer-2′,3′-cyclophosphates.[151] For the sake of illustrating the principle, Figure 51 formulates such an intrinsically consistent co-oligomerization set of 16 tetramer sequences, so defined that for every formally possible ligation step among these tetramers, the requisite template-tetramer is present within the set, but not any superfluous tetramer. Of the two highlighted subsets of 4 tetramers, each in itself also represents such a consistent co-oligomerization set. Through neglect of all the many factors that for partially obvious reasons would in reality make such reaction systems hope-

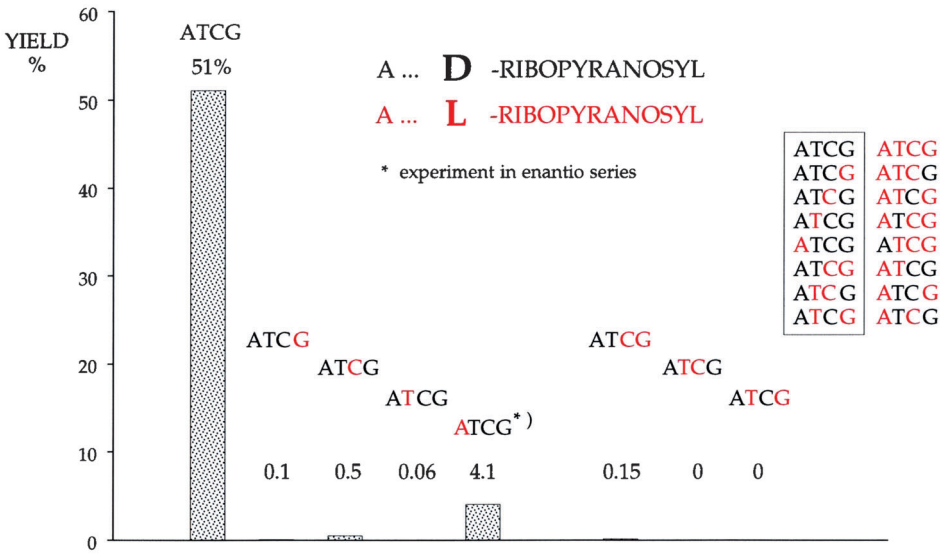
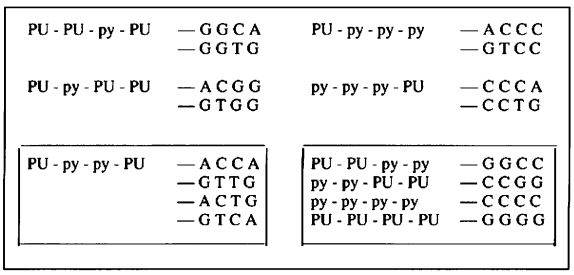


Figure 50. Chiroselectivity of ligative self-assembly (cf. Figure 49): substitution of D- for L-ribose in one each of the four nucleotide units of the all-D tetramer reduces the overall yield of oligomers by 100- to 500-fold, except with substitution at the 4′ end, where the yield is reduced only by a factor of ca. 10.[138] Given are the yields of higher oligomers; after 3 weeks, 0.0025 M, 1.5 M LiCl, 4 °C.



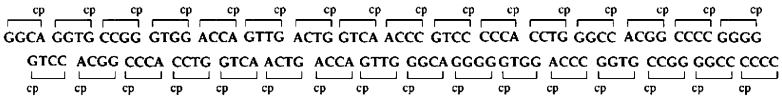


Figure 51. Ligative self-assembly of p-RNA duplexes starting with tetramers that would not be half-complementary to themselves, but rather half-complementary among each other.[151] Above: Co-oligomerization set of 16 tetramers that would contain the required template tetramer for every formally possible ligation step of every tetramer. Below: Example of a duplex from two of the 1.8×10^{19} sequences the self-assembly of which would be formally possible from these 16 tetramers (cp = 2′,3′-cyclophosphate end of tetramers).

lessly complex, one gets a glimpse of how and to what extent diversity in oligomer sequences in such a system might arise: From a co-oligomerization set of 4 (activated) tetramers, there could be generated 10^6 different sequences in duplexes of 40-mers (10-mers of 4-mers), and from a co-oligomerization set of 16 (activated) tetramers there would formally be 1.8×10^{19} different sequences in duplexes of 64-mers (16-mers of 4-mers). Two of these are shown (as duplex) in Figure 51 for illustrative purposes.

At this point, it seems worthwhile to reflect on a thought experiment of a stereochemical nature (Figure 52): In a radical simplification of reality, let us assume an ideally stochastic process of combinatorial co-oligomerization within a mixture of all 8 diastereomeric racemates of each tetramer out of a consistent co-oligomerization set of 16 tetramers, and assume further that not a library of various duplex lengths, but rather only duplexes of 100-mers were formed. Further, we will impose complete chiroselectivity upon the reaction (not simply "pronounced" chiroselectivity, as above). There would then be exclusive formation of two libraries of duplexed 100-mers, both containing the same amounts of matter, and the two would be homochiral, one D and the other L. *Although deriving from racemates, the two libraries together would not themselves constitute a racemate.* This is because the possible number of different sequences (ca. 10^{60}) would be incomparably greater than the maximum number of sequences actually generated (at most ca. 10^{24}, with molar amounts of starting materials). Statistically, each of the sequences generated would appear at most once, and in fact in only one of the libraries, not both. The constitutional makeups of the two libraries would differ, which rules out their being racemates. Such a scenario illustrates the notion that a chemical system cannot ultimately avoid breaking molecular mirror symmetry whenever a critical degree of constitutional diversity is exceeded as a consequence of increasing constitutional complexity (Figure 52).[138,152]

An experiment in the RNA series showed that template-controlled ligation of tetramer-2′,3′-cyclophosphates is nearly as efficient as the corresponding process in p-RNA (cf. above and Figure 46). A reaction type involving self-assembly of oligomers of tetramers within growing duplexes would also be conceivable here; because only template-controlled growth would be involved, and not replication, the repetitive formation of non-natural 2′,5′-phosphodiester bridges between tetramer units would presumably not deter the growth process. 2′,3′-Cyclophosphates would appear of interest in the context of any oligomerization process leading toward the generation of higher RNA oligomers, not only because they represent the simplest and mildest form of phosphate activation,[153] but also

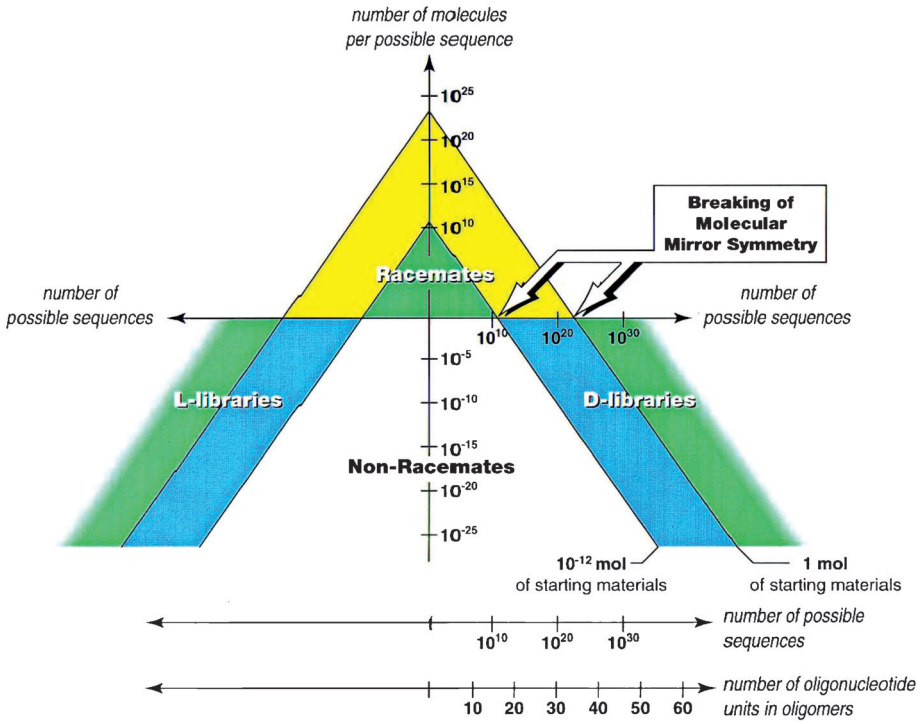


Figure 52. If, as a consequence of an increase in constitutional complexity (e.g. sequence length), the constitutional diversity of an inherently racemic mixture of components (e.g., oligonucleotide sequences) exceeds a critical limit, a breach in molecular mirror symmetry through deracemization is unavoidable.[138]

because the principle of their self-assisted oligomerization within duplexes appears modifiable in the most versatile ways with respect to sequence length of the activated ligation units.[154] These are reasons why one would be hesitant to dismiss this general type of self-assembly of constitutional diversity in oligonucleotide systems as etiologically irrelevant.

In retrospect, among all our studies on nucleic acid alternatives, the one on the pyranose form of RNA occupies a very special position. The sheer existence of an alternative informational oligomer system that could have been generated from the very same building blocks and along analogous chemical pathways as RNA and furthermore—as judged from its chemical properties—could well have been chosen as Nature's genetic system, represents in itself a provocative fact from an etiological point of view. Knowledge of the chemical properties of this RNA alternative can allow us to better recognize those properties of RNA that were significant at the chemical level for the system's biological selection (see below). As far as the chemistry of informational oligomer systems is concerned, p-RNA has stimulated a number of further research projects; one is that regarding constitutional δ-peptide analogues of p-RNA in the laboratory of Gerhard Quinkert in Frankfurt,[155] whereas others include studies pursued in our own laboratory in the wake of the p-RNA project. These are summarized below.

4.6. The Family of the Four Diastereomeric Pentopyranosyl-(4′→2′)-Oligonucleotides

Within the etiological perspective, the remarkable characteristics of the pyranose form of RNA almost inevitably led to the question "Why ribose, and not one of the other pentoses?" Any answer can scarcely avoid reference to the matter of the formation of ribose; this question was dealt with above (Section 4.1, as well as below); here we are concerned with the functional aspect. It would actually be just as important to consider the matter with respect to pentofuranoses along with pentopyranoses;[156] the choice fell to the pyranoses, because conformational reasoning had led us to suspect that, in the series of four pentopyranose systems with equatorial nucleobases (Figure 53), none of the three diastereomeric members would be able to compete with p-RNA with respect to pairing ability. If this were true, it could become part of a rationalization of ribose's unique position. In the β-xylopyranose isomer, the pairing conformation corresponding to that of p-RNA must be severely sterically hindered (arrow symbols in Figure 53), and in both the α-lyxo and α-arabino oligomers the pairing conformation must necessarily deviate significantly from that of p-RNA as a result of axial attachment of the phosphodiester bridge at C-4′.[133] Reality turned out to be far more complex. Eventually, we had to accept that experimental observations in fact presented essentially the antithesis of this prediction (Figure 54).[157] This was a harsh lesson, albeit an instructive and important one. Conformational reasoning clearly predicted strong steric hindrance in the pairing conformation of the α-arabinopyranose system, but this in fact turned out to be

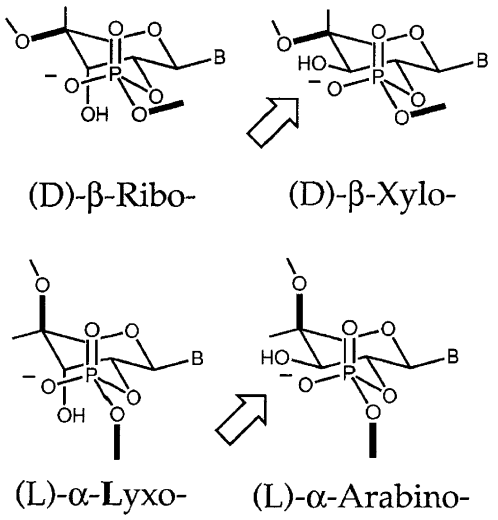


Figure 53. The family of the four diastereomeric pentopyranosyl-4′,2′-oligonucleotides. Arrows indicate sites of steric hindrance in the (idealized) conformations that would correspond to the pairing conformation of p-RNA.

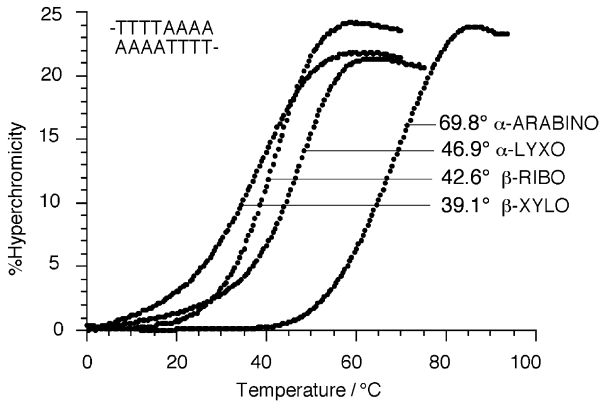


Figure 54. Melting points, determined by UV spectroscopy, for the four isomeric pentopyranosyl oligonucleotide duplexes of the self-complementary base sequence 4′-AAAATTTT.[84] $c = 10$ μM; 0.15 M NaCl, 0.01 M NaH_2PO_4, pH = 7.

the one showing by far the greatest pairing strength among all the four isomers.[157–159]

The duplex structure of the self-complementary sequence 4′-ara-p-CGAATTCG, again established NMR spectroscopically in the laboratory of Bernhard Jaun,[158,169] displays a ladder structure resembling that of the corresponding p-RNA duplex,[136] though one that is more clearly linear and somewhat more strongly inclined (Figure 55). This reminded us of something that also has to be part of a conformational evaluation of the prerequisites for Watson–Crick base pairing, namely that pairing is also possible with severely sterically hindered pairing conformations provided that all alternative conformations are also hindered. NMR spectroscopic data provide in addition an important indication as to why the arabino duplex shows a higher melting point than the corresponding p-RNA duplex: In duplexes of p-RNA and xylopyranose-NA, some degree of intra-strand hindrance between neighboring nucleotide units must be postulated as a

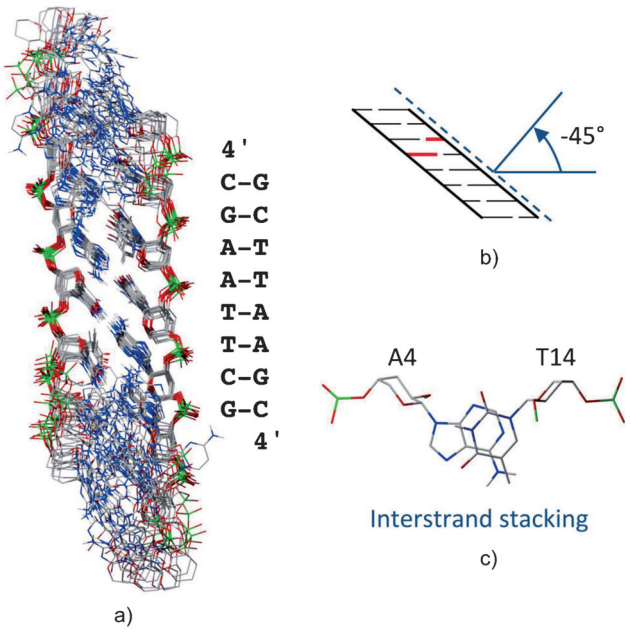


Figure 55. a) Structure established by NMR spectroscopy for the duplex of the self-complementary base sequence 4′-CGAATTCG in the L-α-arabinopyranose series;[158,169] b) Backbone inclination; c) Interstrand base stacking (structural images provided by Bernhard Jaun, ETH).

consequence of the diequatorial linkage of the phosphodiester bridge. This kind of steric constraint is absent in arabinopyranose- and lyxopyranose-NA due to axial–equatorial attachment of that bridge (Figure 56). Apart from evaluation of models, this interpretation rests on the differing NMR shieldings of those protons indicated in Figure 56, due to neighboring nucleobases.[158,169]

All four diastereomers of the pentopyranose nucleic acid family share the conspicuous feature that they engage in informational intra-family cross-pairing in Watson–Crick mode, apparently with no limitations of any sort (Figure 57).[160] None of them pairs with the natural nucleic acids, however, nor with homo-DNA. All four members appear to coincide with regard to the structural parameters that determine the "pairing language"; the most important of these are, besides a general backbone flexibility, the degree and orientation of the backbone/basepair-axes inclination. With regard to the melting temperatures of duplexes, all four members respond in the same way to the sequence-dependent occurrence of interstrand base stacking (Figure 58). Along with consistently high melting points of self- and cross-pairing complexes for complementary sequences (Figure 57) comes the phenomenon that, in all four systems, even short non-self-complementary sequences, such as 4′-TTAAAATA and 4′-TATTTTAA, engage in relatively strong self-pairing, presumably via a reading frame shift.[160]

For a few selected examples it was shown that self-complementary base-sequences that contain backbones consisting of diastereomeric pentopyranose units within the same strand [such as D-ribo-L-lyxo-(T_4A_4)] are capable of pairing with essentially the same efficiency as strands with configurationally uniform backbones. Strands need not have the same

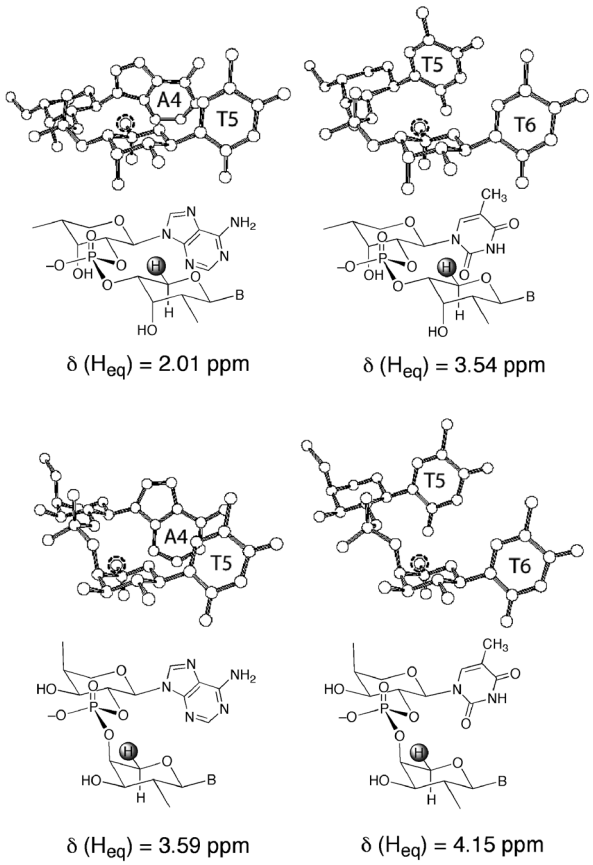
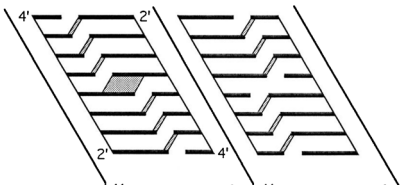


Figure 56. NMR spectroscopic differences in shielding of the equatorial 5′-protons by neighboring nucleobases in the ribo- vs. arabino-series correlate with relative pairing strengths in the two series.[159,169]

T_m values of **TTAAAATA** / **TATTTTAA** duplexes

SELF-PAIRING		-	15	10	23	48	
		RNA **r**	p**r**	p**x**	p**l**	p**a**	4' 2' **TTAAAATA**
-	RNA **r**	17	(-)	(-)	(-)	(-)	
22	p**r**	(-)	46	46	30	45	
23	p**x**	(-)	46	44	26	46	
20	p**l**	(-)	40	37	46	61	
49	p**a**	(-)	54	53	62	75	
4' 2' **TATTTTAA**							

Figure 57. Inter- and intrasystem cross- und self-base pairing in the family of the pentopyranose nucleic acids (T_m values; $c = 10$ mM in 1.0 M NaCl, 0.01 M NaH_2PO_4, 0.1 mM Na_2EDTA, pH = 7). r = D-β-ribo-furano, pr = D-β-ribo-pyrano, px = D-β-xylo-pyrano, pl = L-α-lyxo-pyrano, pa = L-α-ara-pyrano.[160]



	4′-TTTTAAAA-2′ AAAATTTT- T_m (10 μM)	4′-AAAATTTT-2′ TTTTAAAA- T_m (10 μM)	ΔΔG	ΔΔH
			kcal/mol	
RIBO	40 °C	27 °C	+2.5	+11.8
LYXO	47 °C	38 °C	+2.0	+ 6.3
ARABINO	69 °C	61 °C	+1.0	- 2.3
XYLO	39 °C	16 °C	+1.6	+20.7

Figure 58. Consistently differing stabilities within the pentopyranose series of duplexes of the self-complementary base sequences 4′-TTTTAAAA and 4′-AAAATTTT can be traced back to correspondingly differing stabilization through interstrand base stacking. According to relative stabilities of duplexes containing the sequence motifs (py_n-PU_n) versus (PU_n-py_n), all four pentopyranosyl-(2′→4′)-oligonucleotide systems have similar backbone inclinations. Conditions: 0.15 M NaCl, 0.01 M Tris·HCl, pH 7.0.

backbones, nor do the backbones need to be configurationally uniform, in order for two strands of the family of pentopyranose nucleic acid alternatives to be capable of informational *W*atson–Crick pairing; however, what is necessary is that the two strands be homochiral.

One of the decisive criteria for constitutional design of nucleic acid alternatives from the pentopyranose family had been the 6-center periodicity characteristic of the backbone structure of natural systems. The question arose in retrospect as to what extent this specific periodicity is in fact really essential to the pairing ability of an oligonucleotide system. For this reason, we synthesized and studied the base-pairing capability of the oligomer system isomeric to p-RNA in which (4′→2′)-phosphodiester bridging is replaced by a (4′→3′)-bridge, whereby the constitutional periodicity is reduced from 6 centers to 5 (Figure 59). To no great surprise, we found experimental confirmation for the prediction that such a system would show no pairing—even if only in consequence

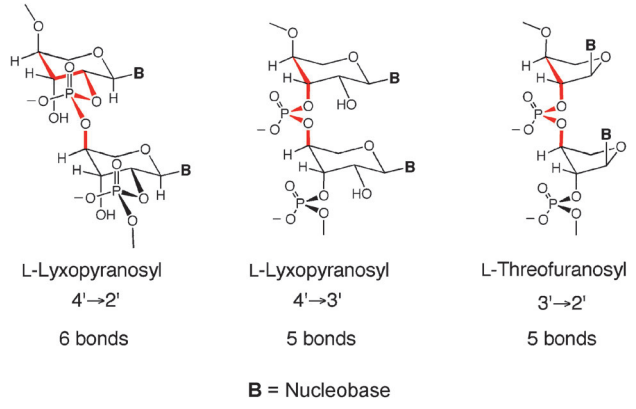


Figure 59. Pairing characteristics of L-α-lyxopyranosyl-4′,3′-oligonucleotides[161] provided the incentive for our work on L-α-*threo*-furanosyl oligonucleotides ("TNA").

of mutual repulsion of vicinal phosphodiester groups.[161] Since such repulsion would not be expected for a corresponding isomer with diaxial phosphodiester bridging, we extended our investigation to the lyxopyranose series (Figure 59). The result took us by surprise: L(4′→3′)-lyxopyranose-NA proved capable of (weak) base-pairing, and even able to engage in cross-pairing with DNA, albeit only weakly so, and with variable sequence selectivity.[161] This was the signal to abandon a hypothesis cherished from the outset of our work with nucleic acid alternatives, namely that it would be pointless trying to construct from a tetrose sugar a nucleic acid alternative capable of pairing, since there could be no 6-center periodicity in the backbone. This change of mind led to our project on α-L-threofuranose nucleic acid ("TNA") (Figure 59).

4.7. L-α-Threofuranose Nucleic Acid ("TNA")

TNA turned out to represent an outstanding Watson–Crick pairing system that pairs informationally not only with itself, but also with the natural nucleic acids (Figure 60).[165–167] The duplex structure of the self-complementary standard

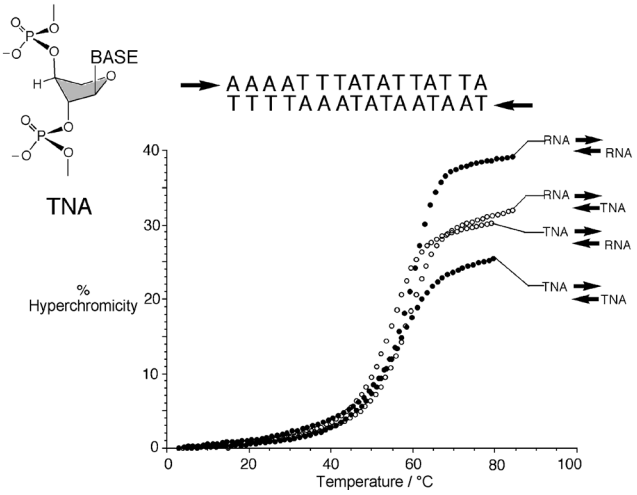


Figure 60. T_m curves for intra- and inter-system base pairing of two antiparallel-complementary hexadecamer sequences in the TNA and RNA series.[165,166] UV/T_m curves, 260 nm, ca. 10 μM in 1 M NaCl, 10 mM NaH_2PO_4, 0.1 mM Na_2EDTA, pH 7.

sequence 3′-CGAATTCG was established once again by NMR spectroscopy in the Jaun laboratory. It is a clockwise double helix, resembling in detail more nearly an A-type than a B-type DNA (Figure 61).[168,169] Consistent with such structural similarities, TNA engages in significantly stronger cross-pairing with RNA than with DNA (Figure 62).[170] The strand orientation in a TNA double helix is, at least for base sequences of the type tested in the experiment referred to in Figure 63, necessarily antiparallel;[165] in this regard, the pairing selectivity of TNA surpasses that of DNA. TNA, like DNA, shows relatively high resistance to hydrolytic strand decomposition (Figure 64).[166] In a comparative test, RNA is by far the most sensitive of all the systems

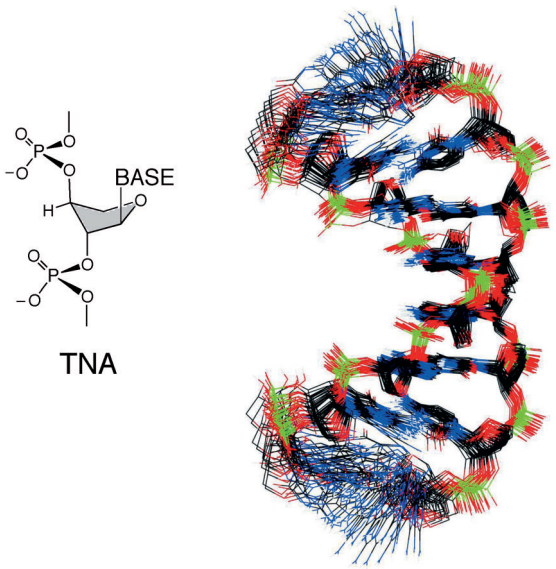


Figure 61. NMR-spectroscopically determined structure for the TNA duplex of the antiparallel self-complementary sequence 3′-CGAATTCG (structural image provided by Bernhard Jaun, ETH).[168,169]

	TNA	RNA	DNA	
TNA	42 **56** 53 31	28 **57** 57 39	32 **47** 43 25	A B C D
RNA	76 **58** 50 41	62 **59** 57 52	59 **44** 40 36	A B C D
DNA	68 **41** 36 26	47 **43** 41 35	55 **48** 43 36	A B C D

DUPLEX
A 3' - A_{16} - 2'
2' - T_{16} - 3'
B 3' - $A_4T_3ATAT_2AT_2A$ - 2'
2' - $T_4A_3TATA_2TA_2T$ - 3'
C 3' - $AT_2AT_2ATAT_3A_4$ - 2'
2' - $TA_2TA_2TATA_3T_4$ - 3'
D 3' - ATTCAGCG - 2'
2' - TAAGTCGC - 3'

T_m (10 μM)
1.0M NaCl, pH 7.0

	TNA	RNA	DNA
TNA	**-14.5** -73.7 -59.2	**-17.3** -93.9 -76.6	**-13.3** -77.3 -64.0
RNA	**-15.8** -95.2 -79.4	**-20.0** -131.9 -111.9	**-12.3** -73.9 -61.6
DNA	**-11.0** -57.0 -46.0	**-13.7** -98.0 -84.3	**-16.8** -129.2 -112.4

DUPLEX
B 3' - $A_4T_3ATAT_2AT_2A$ - 2'
2' - $T_4A_3TATA_2TA_2T$ - 3'

$\Delta G^{25°C}$
ΔH
$T\Delta S$

Figure 62. Overview of intra- and inter-system pairing of selected base sequences in the TNA, DNA, and RNA series; T_m values (left) and thermodynamic data (right).[165,166]

investigated, and in particular also more sensitive than p-RNA. The data from the test series shown in Figure 64 illustrate the destabilization of a phosphodiester group by a vicinal free hydroxy group, already familiar from RNA.

In our search for potentially natural nucleic acid alternatives, in TNA we for the first time came across a pairing system capable of communicating reliably[171] with RNA. From the etiological standpoint, this aroused interest not least

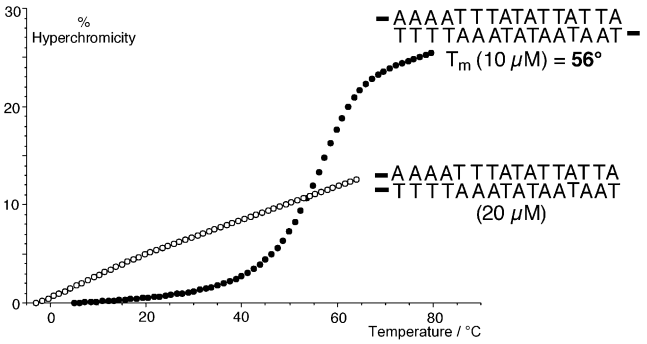


Figure 63. Comparison of the pairing behavior of a TNA base sequence (L-α-threo-furanosyl-(3′→2′)-oligonucleotide) for corresponding antiparallel-complementary and parallel-complementary base sequences.[165]

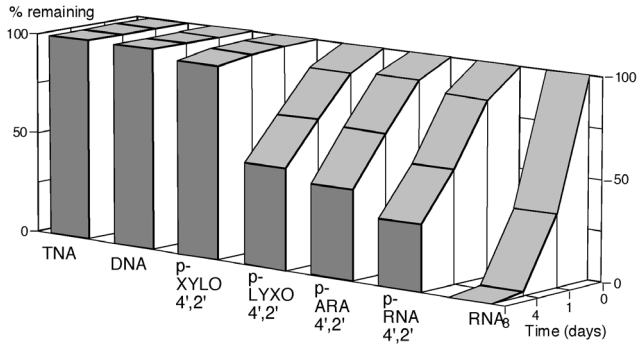
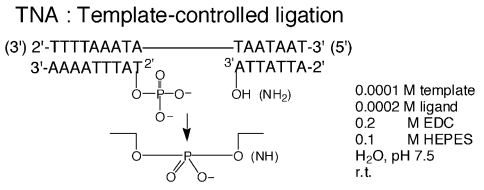


Figure 64. Hydrolytic stability of oligonucleotide single strands (T_8). With respect to hydrolytic decomposition, the stability of TNA is comparable to that of DNA.[166] Conditions: 250 mM $MgCl_2$, 1.0 M NaCl, 100 mM HEPES, pH 8, 35 °C (RNA: U in place of T).

because the sugar module in TNA with its "only four carbon atoms" is generatively simpler than the corresponding module in RNA.[172] Since intersystem communication between pairing systems is of prime etiological interest,[173,174] we undertook a rather extensive comparative study in template-controlled ligation of TNA ligands at TNA, RNA, and DNA templates (Figure 65).[175,176] In these experiments we took advantage of the greater pairing efficiency of 2,6-diaminopurine relative to adenine,[135,177] as well as the greater nucleophilicity in such ligation reactions of the amino group relative to the hydroxy group.[178,179]

The ligation experiments of Figure 65 (EDC = water-soluble carbodiimide) reveal a drastic acceleration of ligation by replacement of adenine with 2,6-diaminopurine in the two TNA ligands, using both the TNA and the RNA templates; the substitution remained ineffective with the DNA template that was "dead" under the reaction conditions selected. The efficiency of ligation reached high levels when the hydroxy group at the 3′ end of the TNA ligand (functioning as a nucleophile) was replaced with an amino group. This is true to an even greater extent if, at the same time, adenine in both TNA ligands is replaced by 2,6-diaminopurine.[175] In the absence of these accelerating measures, ligation of TNA ligands is a slow reaction with both TNA and RNA templates, at least when both the ligand and template sequences contain



LIGANDS		TEMPLATES			
TNA–2'-OPO_3^{2-}	HO NH_2–3'-TNA	TNA (A,T)	TNA (**D**,T)	RNA (A,U)	DNA (A,T)
(A,T)	HO–	4h 19%	14%	30%	4h —
(**D**,T)	HO–	1h 57% 4h >95%	51% 5h >95%	53% 4h >95%	4h —
(A,T)	H_2N–	10 min. 59% 30 min. 97%	52% 33 min. 95%	43% 50 min. 95%	10min. 2% 4h 37%
(**D**,T)	H_2N–	10 min. 94% 13 min. >95%	80% 20 min. >95%	60% 45 min. >95%	10min. 11% 5 h >95%

Figure 65. Overview of results for ligation experiments in the TNA series (D = 2,6-diaminopurine).[175]

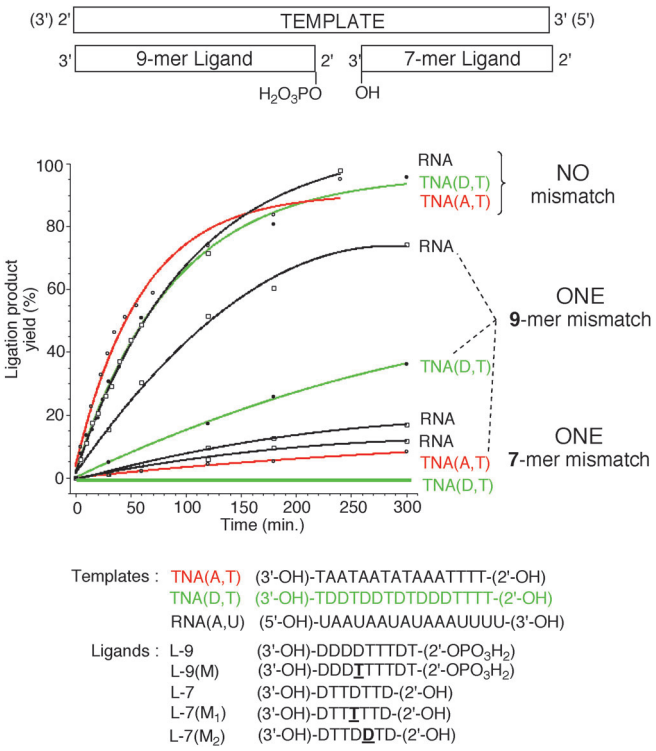


Figure 66. Overview of the influence of "mismatch" bases on template-controlled ligation in the TNA series.[175]

only adenine and thymine (or uracil) bases. An attempt to proceed along the path of information transfer in the opposite direction proved to be a failure: no ligation of the two corresponding RNA ligands could be observed on a corresponding TNA template under otherwise identical conditions. This should be seen in the perspective of the finding that ligation of two corresponding RNA sequences, again under the same conditions but with the RNA template, failed as well. Also known to be inefficient is chemical ligation of RNA sequences at DNA templates.[180] For the ligation of TNA sequences, RNA tends to be a more efficient template than TNA itself. This is seen as well from the results of experiments that demonstrate the influence of a mismatch of bases in the two ligands on the sequence-specificity of their ligation (Figure 66).[175]

Studies carried out in the research group of Szostak,[181–183] as well as in that of Herdewijn,[184] have shown that TNA sequences can function as substrates for enzymes. Selected polymerases were demonstrated to be capable of transcribing DNA sequences into complementary TNA sequences enzymatically,[181] as well as the reverse: TNA sequences into complementary DNA sequences.[182] Thus, there exists at least in principle the basic potential for an experimental in vitro evolution to catalytic TNA sequences.[183] Switzer[185] has observed that guanine- and cytosine-containing TNA sequences, apparently in contrast to ones containing only adenine and uracil (cf. above), can serve as templates for non-enzymatic transcription to RNA sequences.

Through its properties, TNA has shown that the search for alternative Watson–Crick pairing systems of the oligonucleotide type has available to it a structural latitude much greater than originally assumed. The next step along a path of simplifying the structure of nucleic acid alternatives would have been one to a system with backbone modules consisting of only *three* carbon atoms (Figure 67). While we took this step conceptually,[161] we did not do so experimentally. The reason for this was our self-imposed demand that a candidate structure is to be included in our study only if we can regard it as a "potentially natural" structure, that is, only if it could have arisen through the same type of chemistry as that

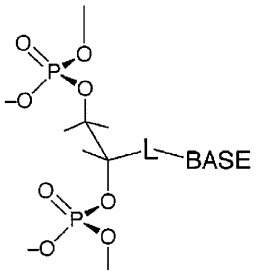


Figure 67. Extrapolation of the observed pairing behavior of TNA on the hypothetical pairing behavior of TNA-analogous aliphatic oligonucleotide systems.[161] L = ligand, e.g. CH_2.

through which RNA is supposed to have arisen. Such is not possible for a structure to be derived, for example, from glycerol. Fortunately, Eric Meggers,[186,187] with a different objective, did carry out this step, and was able to show that the oligomer system he calls "glycol-NA" ("GNA") (Figure 68) represents an outstanding pairing system, one that communicates efficiently with natural DNA. After Nielsen's "PNA",[188–191] Megger's "GNA" became a further important representative of Watson–Crick pairing systems that, despite the absence of cyclic structural elements in their backbones, are nevertheless capable of efficient pairing with both themselves and with natural nucleic acids. The two systems significantly enrich our understanding of the structural prerequisites for Watson–Crick base pairing.[192]

PNA and GNA are of interest as molecular instruments of informational chemistry and biology irrespective of whether they are or are not relevant to the etiology of RNA. For both

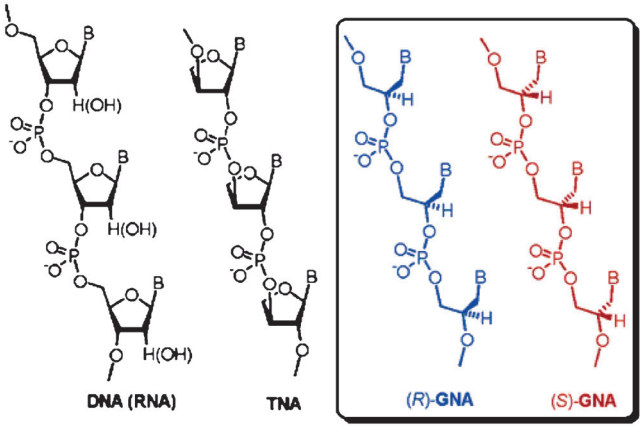


Figure 68. Erich Megger's first description of a "glycol-NA" (GNA), a non-cyclic "C_3-analog" of TNA, capable of intra-system pairing as well as cross-pairing with DNA and RNA.[186]

systems, their discoverers expressed the opinion[187–190] that oligomers could have been formed under prebiotic conditions, and that PNA and GNA might have been capable of functioning as genetic systems of a primordial life. Suggestions of this sort go to the heart of the quest for an etiology of nucleic acid structure, and thus—within the discussion presented here—are to be judged by the same criteria we have thought necessary to apply in the assessment of, for example, the etiological relevance of TNA (see below).

The section that follows begins by summarizing conclusions from the work described above on nucleic acid alternatives, and then deals more comprehensively with the question of the origin and role of RNA in biogenesis. In this context, and also with an eye toward the acyclic pairing system just alluded to, we then touch upon the question of RNA precursors. This was the question that led us toward a project directed far beyond the previous study. Presentation of that work then concludes this chapter on nucleic acids.

4.8. Why RNA?

Drawing lessons from a comparison of the pairing characteristics of potential nucleic acid alternatives with those of RNA seems simple, but at the same time also difficult. Difficult above all because we do not know whether the RNA type of structure is of biotic or abiotic origin. If biotic, then questions arise in what evolutionary phase and on the basis of what criteria a selection of this structure took place. The choice might have been based on the emergence and efficiency of the phenotypical functions of RNA as a genotype, a possibility on which commenting at the chemical level is hardly feasible with the current state of knowledge. Simple, by contrast, seems drawing lessons from what has been made apparent experimentally. The following can be concluded from these experiments:

1. The capacity for informational Watson–Crick base pairing is not a characteristic exclusive to the natural double-helix structure; on the contrary, this capacity is widespread among oligomer structures derived from modules in the immediate structural vicinity of ribose (Figures 69, 70).
2. The Watson–Crick mode of pairing can mediate informational communication between complementary base sequences via mutually orthogonal "languages". Among "language-determining" structural factors, apart from intrinsic backbone flexibility of the oligomer system, the most important are extent and orientation of the inclination between backbone and base-pair "axes" (Figures 42, 46, 55).
3. Oligomers from the hexose series that are constitutional RNA analogs—at least those derived from allo-, altro-, and glucopyranose—could not have been competitors of RNA in natural selection of a genetic system, because these oligomer systems lack the decisive prerequisite of a capacity for informational base pairing (Figures 30, 31, 36).
4. Nature did not select its genetic system using the criterion of maximization of pairing strength; otherwise it would have chosen a representative, for example, of the family of pentopyranose nucleic acids (Figures 54 and 57).
5. The pairing conformation assumed by RNA in duplexes is not imposed upon its backbone simply by the structural demands of Watson–Crick pairing. This special type of conformation belongs to the ensemble of sterically least strained conformations of an RNA single strand. To this extent, the RNA single strand is pre-organized for the structure it assumes in duplexes (Figures 23–25).
6. Among the four diastereomeric sugars in the aldopentose family, ribose occupies a singular position on a chemical level in the sense that this sugar represents the major product in processes related to the "formose" reaction, provided the formation proceeds under kinetic reaction control (Figures 16–19).

At the time when we began thinking about alternative nucleic acids, the notion that the key process in genetics, Watson–Crick base pairing, might be an exclusive property of the natural type of double helix, was not held uniformly, but was nevertheless widespread.[193] Our work, in conjunction with research carried out in medicinal chemistry laboratories on the "anti-sense project", has transformed the notion into its converse. In contrast to studies of the latter kind, with work aimed at nucleic acid etiology an observation such as "hexopyranose analogues of RNA do not constitute pairing systems" is in no way a disappointing finding; quite the contrary: because here the principle should apply even more sharply than usual that *disproof* of a notion may be possible, but not proof of its validity. Discovering that a particular alternative structure could not have fulfilled RNA's function is an etiologically more significant finding than a result to the contrary. Thus, for example, the notion that nature selected its genetic system using the criterion of maximization of the pairing strength of an oligomeric system would class as falsified, at least to the extent that such a thing is possible at all in an etiological context.

The argument that follows is intended to emphasize and illustrate the point that any attempt to rationalize the RNA structure at a chemical level has as a prerequisite a perspective for comparison, a perspective to be acquired

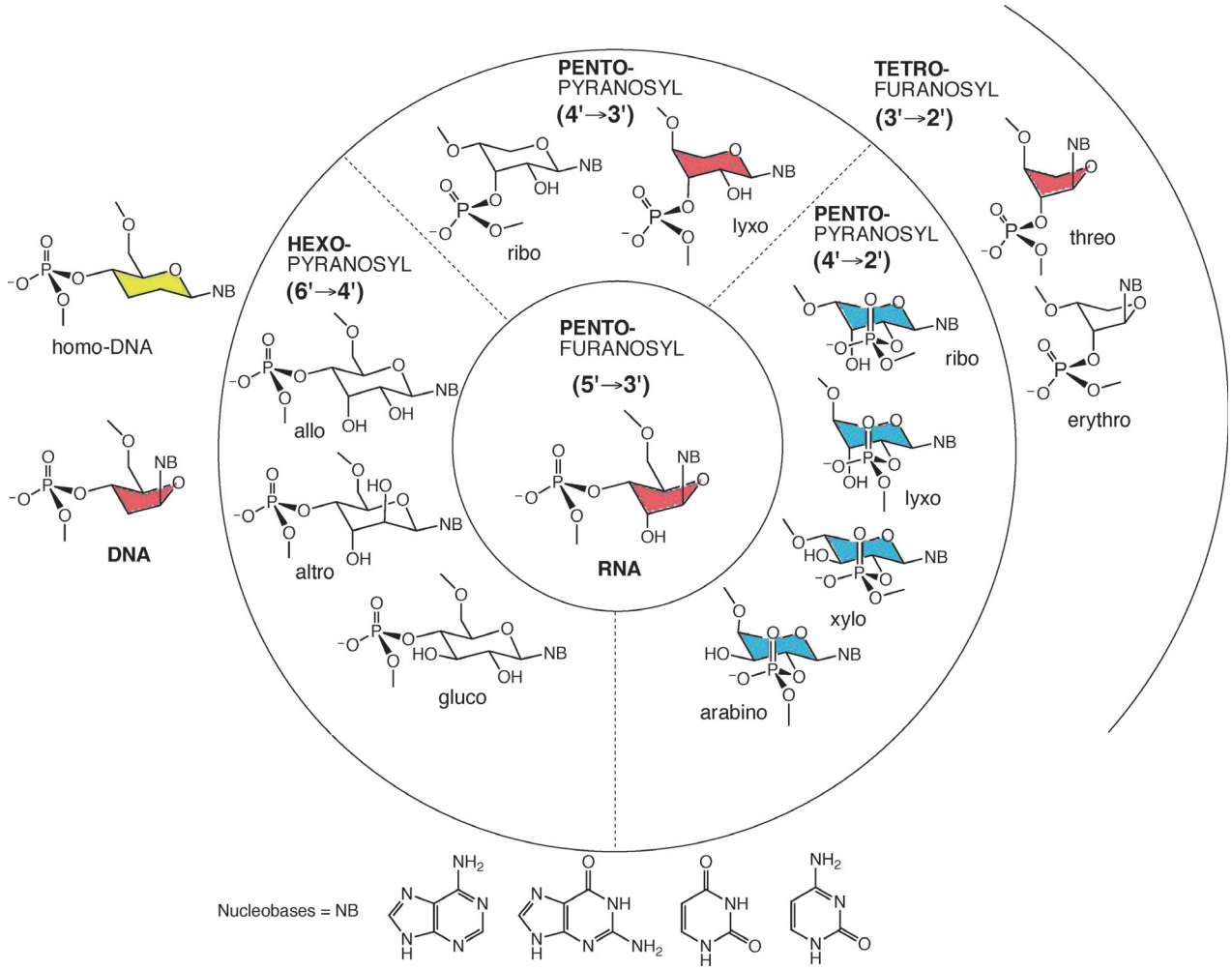


Figure 69. Overview of potentially natural nucleic acid alternatives in the structural vicinity of RNA that have been subjected to experimental study (with homo-DNA as a model system). Formulas of the repetitive units in idealized conformations. In color: Watson–Crick pairing systems; black-and-white: non-pairing systems; red, blue, and yellow: mutually orthogonal "pairing languages".

through knowledge of the properties of nucleic acid alternatives. The approach permits recognition of individual pieces of the jigsaw puzzle of RNA properties relevant to RNA's biological function, pieces that in their totality would be expected eventually to disclose those chemical characteristics that produce a functional optimum to coincide with a structure of the RNA type. In pursuit of such a goal, and in view of the properties of the pyranose form of RNA, it would appear sensible to reduce the question "Why RNA" to its narrowest form possible:

"Why furanose RNA and not pyranose RNA?"

Since these two systems consist of the same building blocks, posing the question in such a way eliminates from the outset any uncertainty regarding the primordial availability of essential ingredients. With the same building blocks, and deploying chemical processes of the same type, nature would have had a choice between two oligomeric systems, both of which could in principle, through Watson–Crick base pairing, have fulfilled the structural prerequisites of information storage and information transfer.

In aqueous solution, the chief fraction of each of the four diastereomeric aldopentoses exists in the pyranose form; in ribose, the pyranose fraction is smallest relative to the other pentoses, but still the major fraction. On a purely chemical level, fixation of the furanose form of ribose through monofunctionalization of the least sterically hindered C(5) hydroxy group, or cyclo-bifunctionalization of the vicinal C(2)/C(3) hydroxy groups, would be easy to achieve; in principle, however, etiological setting of a course in favor of either form would be conceivable.[194] While this question of whether the furanose form of the RNA structure has been selected in conjunction with RNA's first formation must remain open (cf. below), the issue to be raised here relates to the alternative possibility: selection on the basis of functional as against generational criteria.[195]

Among the characteristics of pyranose-RNA that might have played a decisive role in its non-selection, perhaps the most important is the consistently greater stability of its duplexes as compared to RNA. Excess duplex stability relative to RNA means less disposition on the part of p-

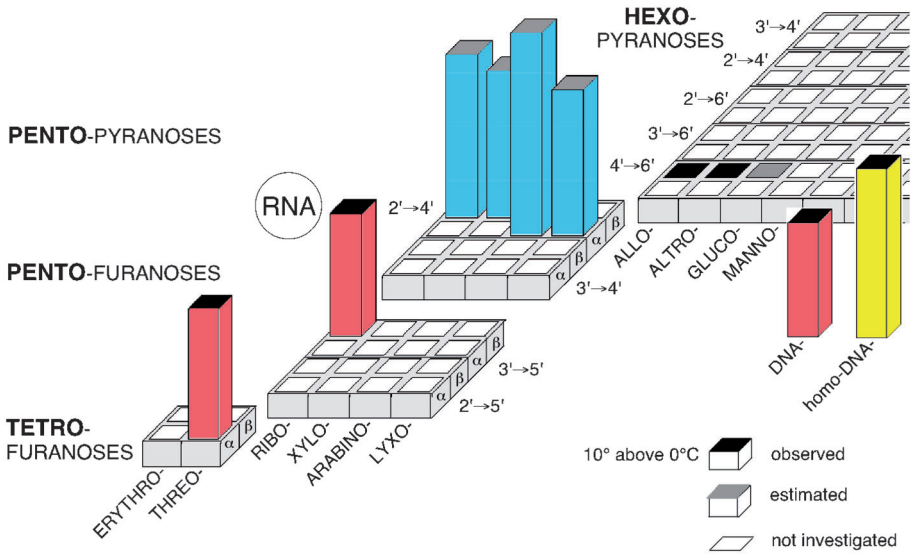


Figure 70. T_m values for duplexes of an antiparallel self-complementary dodecamer sequence CGCGAAUU(TT)CGCG, determined experimentally (or estimated from other pairing data), as a measure of the relative pairing strengths of the oligonucleodtide systems shown in Figure 69 (T_m values represented by column heights; $c \approx 10$ μM, 1 M NaCl, 0.01 M NaH_2PO_4, 0.1 mM Na_2EDTA, pH 7.0).

RNA to satisfy the demands of informational pairing; the data for self-pairing of non-self-complementary sequences in Figure 57 illustrate this point. Such an excess would be in diametric opposition to the greater pairing selectivity one would anticipate on the basis of other behavioral patterns of p-RNA. Among these are the exclusion of purine–purine pairing in reverse-Hoogsteen mode, or the exclusion of parallel-strand orientation in duplexes. But the most important constraint on a pairing system inherent in excess duplex stability has to do with the passing on of genetic information through replication of the information carrier. For this process, an optimum (and not a maximum) in pairing strength is critical. Any degree of excess pairing strength in an intermediary product-duplex could hamper turnover in the replication process, and in the extreme case make it impossible. Whether biological evolution on a molecular level of a "p-RNA replicase" might have been able to deal with a pairing strength much greater than that in RNA is an interesting question, but not one easily answered, because there is hardly a way to evaluate on a chemical level the chances for success of a hypothetical evolutionary process.

Even though p-RNA displays, as already indicated, some special properties that might be said to underscore suitability of p-RNA as a genetic system—a further example of such a property might be proper reproduction of the p-RNA structure through non-enzymatic replication via 2′,3′-cyclophosphates (cf. Figure 46)—evaluation of the characteristics of RNA that are absent in p-RNA seems more important. Among these there is above all the greater conformational flexibility of RNA (even though not documented experimentally; cf. the discussion related to Figure 44). Of the total of six single bonds in the repetitive backbone unit of RNA, only one is part of a ring, and thus restricted in its rotation, as opposed to two of, again, altogether six bonds in p-RNA (cf. Figure 38). This difference in backbone flexibility may be partially responsible for the difference in pairing strengths of the two systems (cf. above), but above all it should mandate a greater conformational diversity for RNA relative to p-RNA. Such a difference can be of decisive importance for chemical diversity in the catalytic potential of an informational oligomer system (see below), decisive above all in a genesis scenario in which the commitment to RNA was first consummated on a biotic level. There, greater efficiency relative to competitors in the production of catalytic functions providing positive feedback toward synthesis of the system could have been the dominant factor in selection. Such a scenario appears today not at all made up out of thin air, but—quite independent of whether RNA originated through prebiotic or biotic processes—in fact corresponds to the notion of an "RNA world" that is supposed to have preceded our present-day "RNA-DNA-protein world".[196]

4.9. An "RNA World"

The ideas and discoveries that in the last decades have led to scientific entrenchment of the concept of an "RNA world" represent a significant part of our present views on the early period of life. Discussion concerning this "world" was triggered by the discovery of ribozymes:[197] catalytic functions in present day organisms are accomplished not only by proteins (enzymes), but also by certain naturally occurring RNA oligomers (ribozymes). In principle, the "dogma" of biology, which in the eyes of biologists was "broken" by this discovery, need not have been a "dogma" in the first place, neither from a chemical nor from a molecular-biological perspective. As early as the 1960s, pioneers in molecular biology had expressed the notion that, in the critical initial phase of life, RNA molecules might have fulfilled not only genotypic but also phenotypic functions.[198] This concept was considered a possible solution to a dilemma: what around that time had been recognized to be the central process in biology—coded translation of the constitutional diversity of one type of biomolecule (RNA as genotype) into the corresponding diversity of a constitutionally completely different type of biomolecule (proteins as phenotype)—is so complex a process both chemically as well as biologically that it could not possibly have been a part of the origin of life. From a chemical perspective, the core of the idea of an RNA functioning both genetically and phenetically amounts to the simplest possible conceptual solution to the coding problem: *the constitution of an RNA sequence "codes" for its own*

conformation. Constitutional diversity of RNA creates conformational diversity of RNA, and therewith chemical diversity of its catalytic potential. One might designate this type of coding as "physical", in contrast to "chemical" as we find it proceeding in the ribosome. The bold ideas first proposed by Carl Woese, Francis Crick, and Leslie Orgel in the years 1967/68 originally arose out of etiological reasoning;[198] what they required be leveraged up to the rank of breakthrough was discovery of the actual existence of ribozymes.[199]

In view of the status the idea of an "RNA world" enjoys in our current beliefs regarding the etiology of life, the question of its experimental anchoring is important. Probably the strongest buttress is the actual revelation of X-ray structural analysis: "*the ribosome is a ribozyme*".[200] Another buttress is of a functional nature, and is the result of a systematic study of the in vitro evolution of RNA: the discovery of sustained pairwise self-replication of ligase ribozymes.[201] These two milestones are accompanied by a substantial and steadily growing number of experimental results acquired along the path of the in vitro evolution of RNA, which testify to the ability of RNA in vitro to catalyze organic chemical reactions of the most diverse sorts (aldolizations, amide formations, etc.).[201–205] That a primordial RNA could at least in principle have exercised alongside genetic functions also phenetic functions is in view of such facts no longer in question from a chemical standpoint. However, what we really are in need of is an experimentally-based scenario for the genesis and operation of such functional RNA and, above all, a convincing concept for its structural and functional incorporation into a chemical system that would be capable of evolving under RNA control.

4.10. Pre-RNA World(s)?

One cannot simply reject out of hand the possibility that "life" did not arise de novo in a chemical state like that with which we are familiar, but rather evolved from a "life" that from a chemical standpoint was constituted quite differently. Probably the most extreme notion of this sort—one not intended by its author to be science fiction, and persisting tenaciously in the scientific literature—is Cairns-Smith's scenario of a strictly inorganic, "mineral life".[206] This is supposed ultimately to have been transformed into our "organic" type of life through a stepwise process the author calls "genetic takeover". Underlying this important term is a concept that, quite independent of the original context, has significance also for a discussion of less ingenious proposals. This applies above all to the notion[207,208] that RNA may have had one or more antecedents which, though structurally and generatively simpler, nevertheless were capable of functioning as genetic systems; in other words, that the RNA world was preceded by "pre-RNA worlds", interconnected by "genetic takeovers". Such a scenario is not necessarily inconceivable: In a chemical sphere under the control of an evolving genetic system where constituents and catalysts would become successively more and more diverse, oligomer variants might emerge that are also capable of informational replication. Such a variant, while at first only supplementing the catalytic arsenal of the overall system,[209] might broaden its sphere of influence, and finally—on the strength of greater efficiency—prevail as the dominant genetic system. Conceptually, such a picture would be—at least in principle—more or less transparent if it were to imply a constitutional identity of genotype and phenotype, as envisioned in the notion of the RNA world. Either such overlapping genetic systems would communicate with each other by cross-pairing, the predecessor system thereby catalyzing the formation of the successor system in a direct way, or this would not be the case, and catalysis promoting formation of the successor system would have to occur via the detour of phenotypic catalysis. While direct transfer of sequence information would simplify a "genetic takeover", *"genetic continuity" would only seem to apply*. In principle, structural differences between two genetic oligomers require that direct transfer of constitutional information from the genotype would not simultaneously convey the "code" for the phenotype as well, because that code is specific to the particular molecular backbone structure of the genotype. Irrespective of its direct communication with the predecessor system, the successor system would have to evolve its phenotype de novo.[174,210]

The first explicit suggestion of a "pre-RNA world" came from pioneers of classical prebiotic chemistry.[207] The proposal may have represented a conclusion drawn from experience gained in the comprehensive experimental work in the Orgel school on the formation of ribonucleotides from their building blocks,[211,212] as well as the oligomerization of monomers to oligonucleotides.[213] Later efforts in the replication of oligoribonucleotides,[214,215] together with studies on the stability of components and intermediates,[62,216,217] conformed to this suggestion. That contrasting conclusions might also be drawn from such experiences is illustrated, among other places, in publications by Jim Ferris[218,219] on the formation of oligoribonucleotides at montmorillonite solid surfaces, starting from activated mononucleotides. Results of this work were judged unwaveringly by their author to be experimental support for the notion of a prebiotic origin of the RNA structure.[219] More recently, John Sutherland has taken up the question of a prebiotic origin for RNA in the sense of an imperative for synthetic organic chemistry: he challenged the latter to demonstrate experimentally that—and how—the RNA-type structure could have been formed under prebiotic conditions.[220] Quite recently, a publication from Sutherland's laboratory[221] caused a furor not only among "RNA-first" proponents, but also in the secondary-to-popular literature, because it was shown experimentally how a seemingly simple, off-the-beaten-path and chemically interesting synthetic route can lead to the ribonucleotide-2,3-cyclophosphate of cytosine, starting entirely with materials (cyanamide, cyanoacetylene, glycolaldehyde, glyceraldehyde, and phosphate) that, from the perspective of prebiotic chemistry in the Miller sense, would all class as "prebiotic" molecules.

As appealing as experimental results in prebiotic chemistry sometimes may appear, when it comes to the question of their etiological relevance they must be subjected to scrutiny using the criterion "the simpler is the enemy of the simple". This applies to structures of primordial molecules of potential

biotic function, as well as to paths that are supposed to have led to them. It stipulates as exhaustive a search as possible for alternatives as candidates for comparison. Thus, for adherents to the "RNA first" postulate, consideration of the broadest possible set of potentially prebiotic RNA syntheses is just as important as is, for advocates of a pre-RNA world, exposure and functional review of "all" conceivable RNA precursors that are prebiotically plausible. Potential diversity in structures and synthetic pathways can hardly be foreseen on the first go; this is a lesson taught by the history of organic chemistry. In the field of prebiotic chemistry, taking voices of confidence[219,221] seriously can be important, but equally so is to judge them with a measure of sustained skepticism. The same holds true with regard to fundamental criticism of prebiotic chemistry by prophets of gloom.[62,222]

Our work on nucleic acid alternatives was directed toward recognition of relationships between structure and function, and toward discovering criteria for an understanding of why nature may have chosen the RNA type of structure. We did not intend to delve into the possibility of finding in the structural neighborhood of RNA alternative systems that might have had the potential to arise more easily—irrespective of setting—than RNA itself. Structures were chosen for study because they were deemed to have been potential alternative products of the very same (hypothetical) chemical processes that led to RNA. Of all the alternative systems investigated, TNA attracted the most attention among biologists: far more, for example, than p-RNA, which may well be the more informative system from the chemical as well as the etiological point of view. In contrast to p-RNA, TNA is capable of interacting informationally with the natural nucleic acids, and its phosphodiester backbone is built with a "simpler" sugar module than is the backbone of RNA. This brought up the question as to whether the "simpler nucleic acid TNA" could have been a precursor of RNA.[172] Our project thus ended up precisely where we at first specifically intended *not* to go: namely, dealing with the issue of *formation* of alternative systems under primordial conditions.

The greater apparent simplicity of the TNA structure as compared to that of RNA, as well as the prospects for TNA as a potential precursor of RNA, may prove deceptive in the end. The sugar module is simpler in the sense that its formation would require only a single two-carbon precursor (glycolaldehyde), whereas the makeup of ribose requires in addition a C_1-component. But recent studies on the formation of ribose under (modified) conditions of the formose reaction (cf. above) indicate that its accessibility in no way lags behind that of threose.[60,67,68] Moreover, there is no basis for the assumption that along the pathway from threose to TNA some one of the reaction steps—including the notorious step of nucleosidation—would be simpler than the corresponding step in the process from ribose to RNA; indeed, the opposite seems more likely the case.

Pairing systems derived from a C_3-sugar, and thus formally even simpler than TNA, have recently come to the fore again.[187,223,224] Might such systems be candidates for a pre-RNA world? The notion according to which their formal simplicity represents, relative to RNA, a possibility for simpler formation under primordial conditions could prove to be still more deceptive than the analogous assumption for TNA. Not only does formation of the C_3-sugar module of, for example, GNA require a reduction step, the linkage between the nucleobase and the C_3-unit must involve a reaction step of the S_N2-type, a process simple enough from a purely chemical point of view, but from the standpoint of prebiotic chemistry as so far understood even more problematic than the notorious nucleosidation step for RNA or TNA.[225–227] Moreover, the structure of RNA with its *trans*-arrangement of the two phosphodiester bridge abutments at the furanose rings teaches us an important lesson: this specific arrangement is expected to inhibit spontaneous intramolecular deactivation, through cyclization, in non-enzymatic oligocondensations of activated monoribonucleotides without protective groups; such a reaction would deprive a template-directed oligomerization process of its activated building blocks. With a corresponding *cis*-arrangement of the two abutments, intramolecular deactivation is undoubtedly what would occur[227]—cf. xylofuranose NA.[156] Activated *monomers* of acyclic oligomeric systems lack the corresponding defense against self-deactivation through cyclization.

Finally, reasoning about RNA's origin suffers from the fact that the type of origin-scenario according to which relevant chemical processes are to be judged remains open: should our model of biogenesis be one of a heterotrophic or an autotrophic origin, and should that of the origin of RNA be one of abiotic or biotic emergence (cf. below)? Even the starting materials to be regarded as available would be very different. Thus, with initial formation of RNA in a heterotrophic scenario, a highly reactive molecule like cyanoacetylene[221] might seem permissible, although hardly for a scenario of the autotrophic type, and most certainly not for biotic first-formation of RNA.

All this notwithstanding, every experimental research project with the goal of demonstrating a "prebiotic" way of forming RNA or a potential precursor is a contribution to the development of a file of experimental facts, familiarity with which will be prerequisite toward any decisive step in the search for the chemistry of the origin of life. This is one of the reasons why Sutherland's "prebiotic" synthesis of activated ribonucleotides[221] is important, ultimately quite independent of the "half-life" of its etiological relevance. Pursuing questions of this sort in a critical way is essential not least because overly optimistic answers allow illusions to develop regarding our state of knowledge about life's origin, not only in the scientific community, but also within the non-scientific public.

4.11. Mapping the Landscape of Potentially Primordial, Informational Oligomer Systems

As indicated earlier, the structural palette of Watson–Crick systems extends far beyond oligomers of the phosphodiester type; how far, in fact, has been demonstrated by studies on chemical aspects of the antisense concept,[228] carried out over the course of the last two decades in medicinal chemistry laboratories.[104,107,113,189] Crucial was the discovery, also alluded to previously, of the properties of

"PNA" by Nielsen.[188] For reasons that range in nature from chemical to biomedical to etiological, it would seem desirable to be in command of as comprehensive an overview as possible of the structural landscape of informational oligomer systems. For the etiology of the RNA structure, only a subset of systems in such a landscape would necessarily be of interest, namely systems which, based on pertinent criteria, would classify as potentially primordial. In a search for such candidates—in contrast to our previous work on the etiology of RNA—*every structure should be admitted, both backbone as well as recognition elements*, so long as one could assume a structure *might* have been accessible under prebiotic conditions. Irrespective of the notoriously problematic nature of questions regarding the prebioticity of structures, reactions, and conditions, if one really wishes to pursue experimentally in a comprehensive way the notion of a pre-RNA world for the sake of ultimately arriving at a substantiated picture of RNA's origin, then trying to map that landscape would seem inescapable.

Whereas in our work carried out thus far the perspective of comparing was directed toward *function*, in the project of charting the landscape of potentially primordial information systems the main issue is *formation*—this with all the additional uncertainties that accompany distancing oneself from questions related to function. Here, to an even greater extent than before, it is necessary that there be selection criteria to assist in recognizing from among a multitude of possibilities the ones most promising to pursue experimentally. Basically, the task is a broadly conceived search for a chemistry that in the simplest possible ways could, under prebiotic conditions, have produced structurally simplest informational systems. In the best possible case, such systems could correspond in their properties to what "geneticists" and "metabolists" might accept as a "least common denominator" in their views on biogenesis (see below). There is nothing to be held against the important objection that we do not actually know what ultimately may count as "prebiotic", except our hope that eventually there will be feedback from increasing experimental experience on the extent of this very ignorance, experience that can only come from the aforementioned objection being at first ignored.

The legendary chemist Siegfried Hoffmann, to whom tribute was recently paid posthumously in *Nachrichten aus der Chemie,*[229] provided a comprehensive review of earlier work in a notable essay[230] published in 1979 entitled "Nukleinsäuremodelle". From today's perspective, if only a purely chemical one, these projects may be seen as precursors to the ones discussed today. Those experiments were carried out primarily in polymer chemistry laboratories and attempted, using the polymer chemistry of the time, to simulate, through coupling of nucleobases with backbones of the most disparate types, the intermolecular communication of DNA.[231,232] In his essay, Hoffman related this work metaphorically to ongoing "evolution", a thought worth remembering in the present context.[233]

Some of those earlier studies attempted a simulation of DNA pairing by substituting for the phosphodiester module an amidic unit.[234,232] Perhaps the first success in this direction was achieved by work done about two decades ago in medicinal chemistry by transforming the ribofuranose ring structure of RNA into a morpholine ring.[235] The decisive breakthrough for amidic backbones occurred with Nielsen's "peptide" nucleic acid, PNA[188] (Figure 71), in the wake of

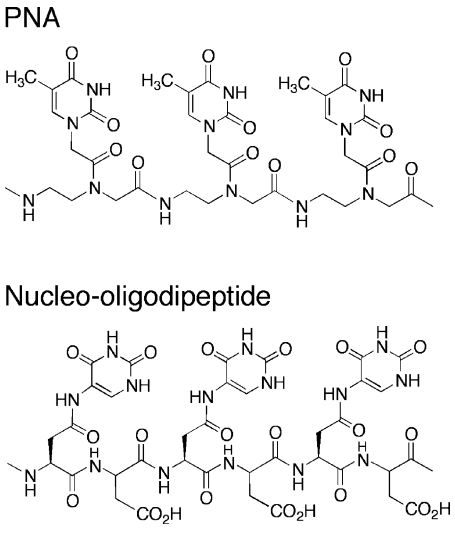


Figure 71. Structural nature of Nielsen's "PNA", and an example of a truly peptidic (actually dipeptidic) nucleo-oligomer with 5-aminouracil as alternative nucleobase.

which numerous pairing systems with PNA-like backbones were described in the antisense literature.[189,236,237] Nucleo-oligomers with true oligopeptide backbone structures have also been investigated,[238,239] among which of particular interest from an etiological standpoint are the heterochiral nucleo-oligopeptides consisting of alternating L- and D-nucleoamino acids described by Diederichsen.[240] As a result of too little constitutional distance between the recognition elements, these show no cross-pairing with natural nucleic acids, but strong pairing among themselves. A similar situation applies with respect to nucleo-β-oligopeptides, described by the same author.[241]

Stanley Miller entitled one of his publications "*Peptide Nucleic Acids Rather Than RNA May Have Been the First Genetic Molecules*", and described experimental results that were supposed to make plausible, from his perspective, the prebioticity of Nielsen's PNA.[242] Here again, a comparison with alternatives is called for. Thrusting themselves forward as alternatives are systems with backbones not just related to peptides, but systems that are "true" peptides. Compared to the modules and reaction types that a PNA oligomer would require for assembling itself, nucleo-oligodipeptides of the type formulated in Figure 71 (bottom)[243] would appear to be "simpler" candidates; this irrespective of the fact that, in view of the multitude of possibilities, one would neither want, nor be able, to commit oneself to a specific pathway by which oligopeptides should have arisen under prebiotic conditions.[244] The formation of peptides from α-amino acids is one of the few chemical transformations whose availability in the palette of prebiotic processes is not doubted under any of the current biogenetic working hypotheses.

In support of a thorough investigation of the nucleo-oligopeptide type of structure is its disguised conformational relationship to the RNA structure, as evidenced by considerations illustrated in Figure 72.[245,246] This relationship, based on criteria of qualitative conformational analysis, leads to the expectation that nucleo-oligo*di*peptides should be capable of base-pairing and possibly of informational cross-pairing with the natural systems. Not only should oligodipeptides bearing recognition elements at conformationally appropriate positions exhibit this property, but so also should, for example, oligomers with partially or completely saturated backbones that repetitively exhibit electronegative centers at the marked locations.[247] The constitution of PNA lies "between" the two variants.

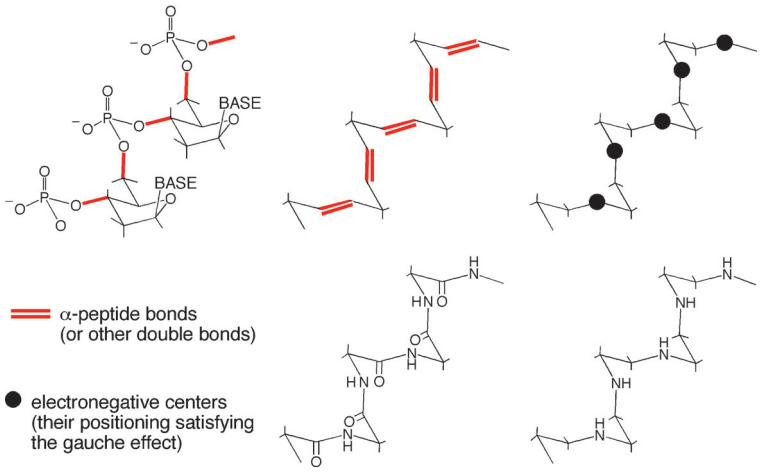


Figure 72. Extrapolation of the (+g/−g)-phosphodiester backbone (A-DNA) to α-oligopeptide- or (e.g.) ethylenediamine backbones. Conformational correspondence between the (idealized) pairing conformation of RNA and that of nucleo-oligodipeptides: bonding with a 180° torsion angle in the former (red) corresponds to peptide bonding in the latter.[246,247]

The work on nucleo-oligodipeptides summarized below represents one of the last projects I directed together with my younger colleague Ramanarayanan Krishnamurthy in our research group at the Scripps Institute in La Jolla. That these experimental studies led in fact to the results I am describing below is due in large measure not least to him.

The type of nucleo-oligodipeptides that we primarily investigated in these studies is shown in Figure 73, and presented there in the hypothetical (idealized) pairing conformation of the dipeptidic monomer unit. This consists of L-glutamic acid provided with a recognition element in its side chain, connected in peptide fashion to L-aspartic acid. The carboxyl function of the aspartic unit remains free in oligomers, ensuring their solubility in aqueous media. We combined the work on nucleo-oligodipeptides with a study of the properties of heterocycles that would class as potentially prebiotic alternative nucleobases, and limited our choice of recognition elements to such bases. Since both alternative and canonical recognition elements can easily be incorporated in oligodipeptide units, the notorious nucleosidation problem of prebiotic chemistry referring to the natural nucleic acids no longer applies.[67,199,211,248] Diversity and simplicity in joining backbone and recognition elements is one of the guidelines along which mapping of the landscape of potentially primordial alternative genetic systems must be pursued. This includes searching for both alternative backbones and alternative recognition elements. Thus, in our work on nucleo-oligodipeptides, we initially utilized 2,6-diamino and 2,6-dioxo derivatives of the *sym*-triazine nucleus as easily accessible derivatives of carboxyl functions, but later concentrated on the derivatives of 5-aminopyrimidine formulated in Figure 74.[249] This pyrimidine family distinguishes itself by the fact that in all four variants the amino group in the 5-position can be selectively acylated, providing a simple and highly variable means of joining the recognition element and a carboxyl function of an oligomer backbone. To such 5-aminopyrimidines can be attributed prebiotic availability at least as justifiably—if not more so—as to the canonical pyrimidines. A recently published long-term experiment carried out in Stanley Miller's laboratory on transformation of ammonium cyanide into nucleobases[250] produced (after hydrolysis of the reaction product) in addition to the canonical purines, more 5-aminouracil than uracil; this corresponds to expectation based on mechanistic analysis of possible pathways starting from the HCN tetramer.[251,252] From the standpoint of the chemistry of HCN, 5-aminouracil is easier to generate than the (constitutionally simpler) uracil.

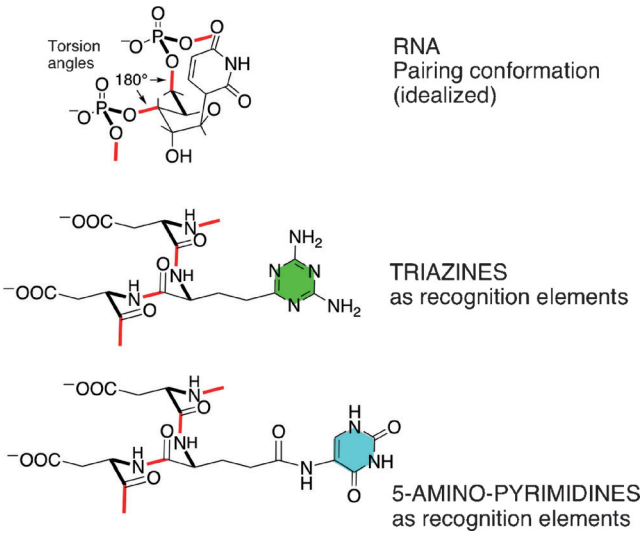


Figure 73. Nucleo-oligodipeptides with L-glutamic acid units as bearers of alternative nucleobases and L-aspartic acid units as spacers.

Among various nucleo-oligodipeptidic systems,[253] the one formulated in Figure 73, derived from oligomers of the dipeptide NH_2-L-glu-L-asp-COOH, showed the most coherent pairing behavior. The melting point curves reproduced in Figure 75 convey an impression of the strong cross-pairing between oligo-(T_n)s of the RNA, DNA and TNA series with

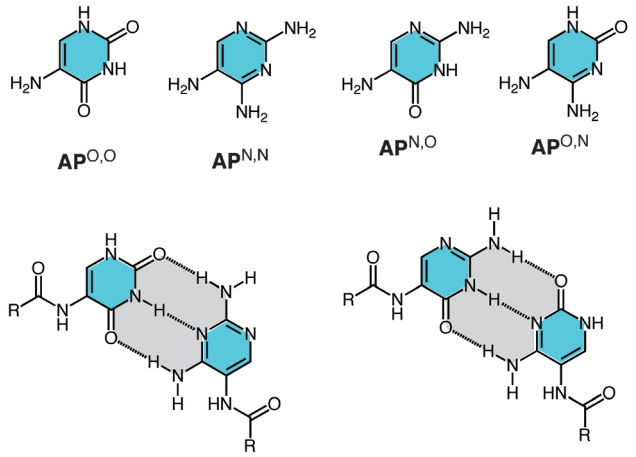


Figure 74. Family of the 5-aminopyrimidines as alternative nucleobases/recognition elementes in nucleo-oligodipeptides.

corresponding homobasic sequences oligo-($T^{N,N}$)s from the (Glu–Asp) series bearing the 2,4-diaminotriazine nucleus as alternative nucleobase. Melting temperatures ascertained by UV spectroscopy correspond to those from CD spectroscopy. It has also been demonstrated for the 16-mers that the stoichiometry of their complexation is consistent with duplex formation. The cross-pairing behavior of corresponding

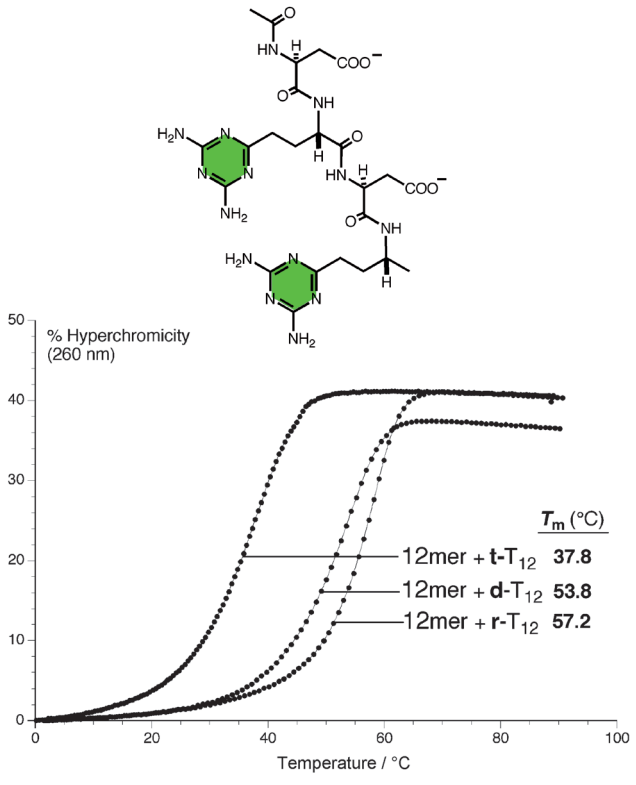


Figure 75. Strong cross-pairings of nucleo-oligodipeptides with 2,4-diaminotriazine as recognition element (with DNA, RNA, and TNA in aqueous medium).[253] Cross-pairing of nucleopeptide-dodecamers [glu-($T^{N,N}$)-asp]$_{12}$ with TNA (t-T_{12}), DNA (d-T_{12}), and RNA (r-T_{12}) (UV/T_m curves, ca. 5+5 μM in 1.0 M NaCl, 10 mM NaH_2PO_4, 0.1 mM Na_2EDTA, pH 7).

nucleodipeptide oligomers containing the 2,4-dioxo form of the *sym*-triazine nucleus as recognition element proved to be surprising, and (at first) disappointing: comparison of Figure 75 with Figure 76 illustrates how strikingly much

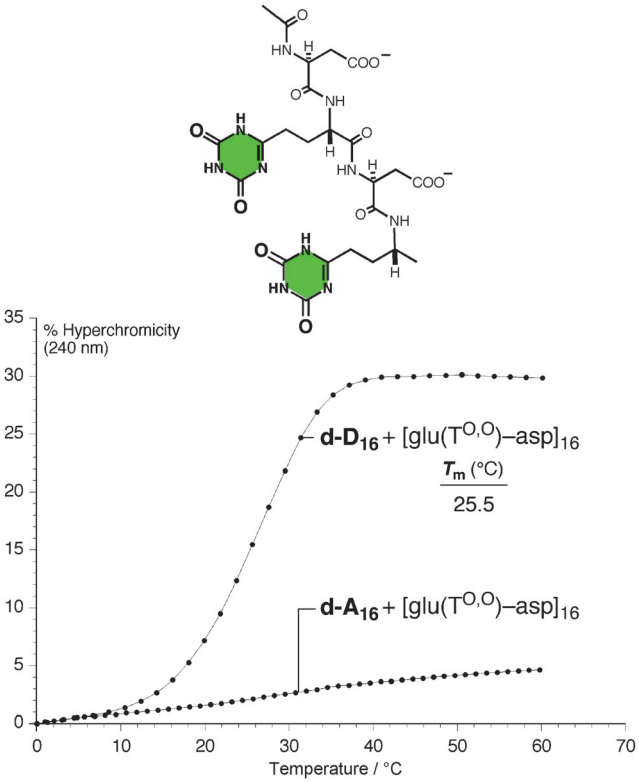
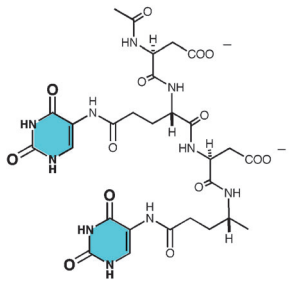


Figure 76. Weak cross-pairings of nucleo-oligodipeptides with 2,4-dioxotriazine as recognition element (with DNA in aqueous medium).[253] Nucleodipeptide-hexadecamer [glu($T^{O,O}$)-asp]$_{16}$ pairs with DNA-D_{16} but not with DNA-A_{16} (UV/T_m curves, ca. 5+5 μM in 1.0 M NaCl, 10 mM NaH_2PO_4, 0.1 mM Na_2EDTA, pH 7).

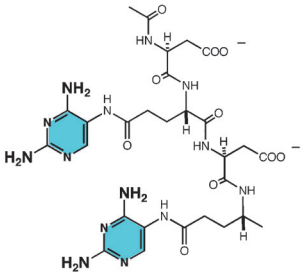
weaker the corresponding (isosteric) pairing of the 2,4-dioxo form turned out to be with DNA and RNA.[253]

The remarkable phenomenon of significantly reduced pairing strength for the 2,5-dioxotriazine/2,6-diaminopurine base pair compared with the 2,5-diaminotriazine/thymine pair has its "antisymmetrical" counterpart in a similarly noteworthy finding: with the corresponding bases from the 5-aminopyrimidine (in place of *sym*-triazine) series in analogous backbone combination, the base-pairing strengths are just as significantly different, *but this time in the reverse direction*. Here it is the dioxo form of the non-canonical nucleobase that pairs more strongly than the diamino form with correspondingly complementary DNA and RNA strands (Figures 77 and 78).[254] When homobasic oligodipeptide sequences, both with non-canonical but complementary bases from the *sym*-triazine and 5-aminopyrimidine series, were brought together to give doubly oligodipeptidic complexes, duplex formation at room temperature was found only between the strands carrying the bases which (with DNA) pair more strongly, namely, 2, 4-diamino-*sym*-triazine and 2,4-dioxo-5-(acyl)amino-pyrimidine. With the corresponding oli-



(°C)	12-mer	16-mer
	UV-T_m	UV-T_m
poly-**d**(A)	**61.1**	**64.7**
poly-**r**(A)	40.8	44.0
d(A_{12})	**51.4**	**54.7 d**(A_{16})
r(A_{12})	35.5	41.0 **r**(A_{16})
t(D_{12})	54.3	—

Figure 77. Strong cross-pairings of nucleo-oligodipeptides [glu($AP^{O,O}$)-asp]$_n$ with 2,4-dioxo-5-aminopyrimidine as recognition element (with DNA and RNA in aqueous medium; ca. 5+5 μM in 1.0 M NaCl, 10 mM NaH_2PO_4, 0.1 mM Na_2EDTA, pH 7, 255 nm).[254]



(°C)	12-mer
	UV-T_m
d(T_{12})	<10
poly-**r**(U)	<10
poly-**d**(T)	23.3

Figure 78. Weak cross-pairings of nucleo-oligodipeptides [glu($AP^{N,N}$)-asp]$_n$ with 2,4-diamino-5-aminopyrimidine as recognition element (with DNA and RNA in aqueous medium; ca. 5+5 μM in 1.0 M NaCl, 10 mM NaH_2PO_4, 0.1 mM Na_2EDTA, pH 7, 250 nm).[254]

gomer strands involving 2,4-dioxo-*sym*-triazine and 2,4-diamino-5-(acyl)aminopyrimidine, although formally equally complementary to each other, no pairing whatsoever is observed above 0 °C (Figure 79).[255]

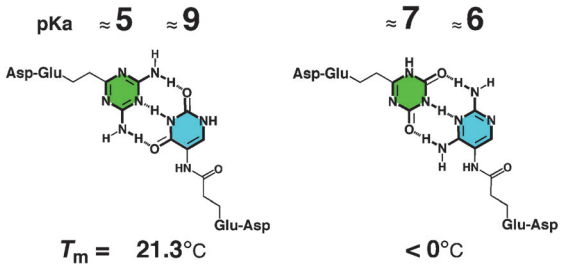


Figure 79. Relative pairing strengths; correlation with pK_a values (12-mers, 5+5 μM, 1.0 M NaCl, phosphate buffer, pH 7). With homobasic nucleo-oligodipeptide dodecamer sequences, which feature recognition elements of the triazine and 5-aminopyrimidine series, pairing was observed between 2,4-diaminotriazine and 2,4-dioxo-5-aminopyrimidine, but not between 2,4-dioxotriazine and 2,4-diamino-5-aminopyrimidine.[254]

Key to qualitative understanding of these results is comparing pK_a values of the four non-canonic and the four canonic bases. The strengths of corresponding pairs in aqueous medium correlate with pK_a differences between pairing partners, in the sense that the smaller this difference, the weaker—under otherwise identical conditions—the pairing (Figure 80 and 81). Ram Krishnamurthy, who first noticed this relationship in the course of the work at La Jolla, in a subsequent independent project extended the study to derivatives of orotic acid (6-carboxyuracil) as non-canonic nucleobases and a correspondingly modified oligodipeptidic backbone, in which these structural elements were now joined in amidic fashion by way of the carboxyl function of the pyrimidine partner.[256] The influence of the carboxamide group at position 6 of the pyrimidine ring on the pK_a values of its 2,4-dioxo and 2,4-diamino derivatives is the reverse of the corresponding influence of an acylamino group at position 5. In fact, the two oligomers containing bases derived from orotic acid show a pairing behavior relative to DNA opposite to that of the pair derived from 5-aminopyrimidine, but analogous to that derived from *sym*-triazine.[256]

An interpretation of the relationship between strength of pairing for constitutionally complementary nucleobases and differences in their pK_a values is to be found in corresponding differences in hydration energies for unpaired versus paired bases: the closer the pK_a values for constitutionally analogous pairing partners to the neutral pH value of the aqueous medium, the more important should be the hydration of the unpaired partners relative to that of the paired partners, and thus the extent to which a difference in hydration energies will work to counter pairing.[257,258]

There is a rich palette of oligomer systems from which, based on qualitative criteria, one might suspect many could belong to the landscape of potential primordial information systems. Thus, apart from the systems just described, the pairing behavior of the hexamer of an oligomer system derived from D-glyceric acid bound to 2,4-dioxo-5-aminopyrimidine as recognition element was tested (Figure 82).[259] In view of the brevity of the base sequence, the observed pairing with DNA is remarkably strong. Extensive attempts were made to synthesize the corresponding 2,4-diamino oligomer, which would have offered the opportunity for a further test of the δ-pK_a correlation. Unfortunately, these attempts proved unsuccessful. One of the reasons for the failure is the tendency of the backbone of such oligomers to readily undergo β-elimination reactions.[259]

In retrospect, interpretation of the outcome of our studies on the pairing behavior of alternative nucleo-oligodipeptides bearing alternative recognition elements received an unanticipated emphasis: the pairing behavior we observed for the *non-canonical* bases directs attention emphatically toward a characteristic of the *canonical* nucleobases, one that distinguishes the latter ahead of other nucleobases and is highly relevant to the etiology of the nucleic acid structure: *Not only do the two canonical base pairs correspond in a most highly specific way constitutionally to their biological function, but their pairing partners also match, in a critical sense, optimum pK_a values for this function* (Figure 81).[253,254,256] Once again, it is an experimental comparison with potential alternative systems that makes it possible to recognize how a component of the natural nucleic acid structure determines the special suitability of the natural type of structure for its biological function. For viewing this coincidence between structure and function from a perspective appropriate to the four canonical nucleobases, we should recall one of the important results of

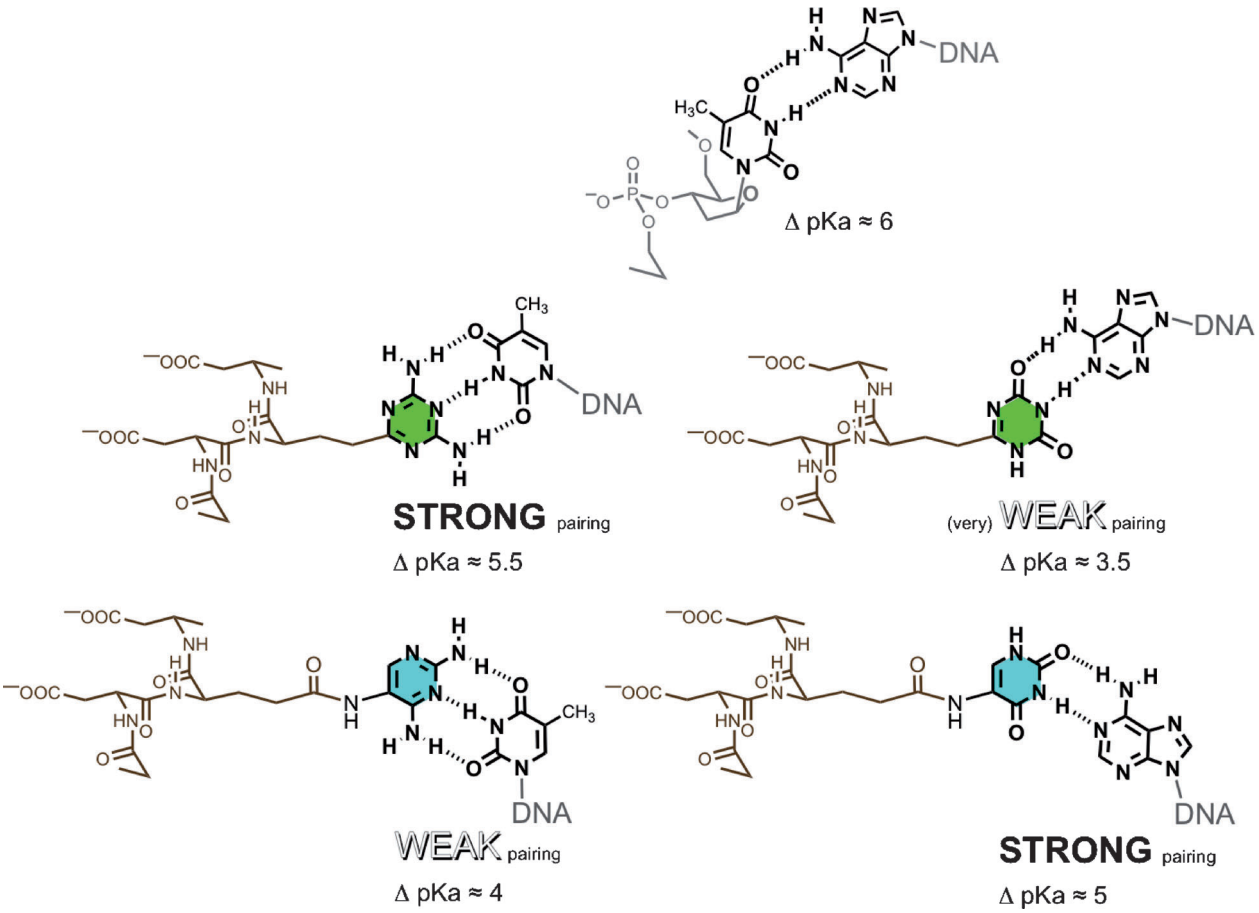


Figure 80. Summary of the observations reproduced in Figures 75–78: "Antisymmetry" in pairing behavior of triazines and 5-aminopyrimidines, and the dependence of pairing strength on differences in pK_a values of the pairing partners.[254]

classical prebiotic chemistry:[5] these bases are chemical descendants of HCN, an elementary *molecular* free energy reservoir, and perhaps one of the most ubiquitous organic molecules in interstellar space.

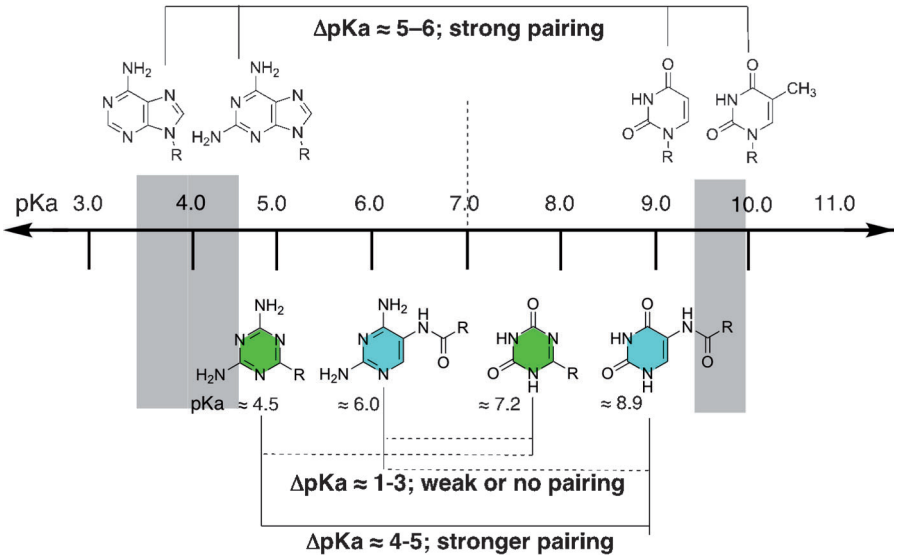


Figure 81. Correlation between base-pairing strength with ΔpK_a values of pairs of complementary bases in aqueous solution at neutral pH. The less the difference in pK_a values of the pairing partners, the weaker the pairing in aqueous medium. The canonical nucleobases correspond in this respect to an optimum.[254,256]

One would probably not want to grant to any of the base pairs utilized in the nucleo-oligodipeptide studies described above the potential to constitute a truly functionally proficient alternative base pair. However, it would be a mistake to impose on the search for potentially primordial informational oligomer systems a demand that they resemble present-day nucleic acids in constitutional robustness, as well as in both pairing efficiency and pairing selectivity. How far it is really necessary to go with such demands remains uncertain, because we are in the dark about the chemical prerequisites of primordial evolutionary steps on a molecular level (cf. below). Demonstrating experimentally that an oligomer system is capable of informational base pairing, and thus (at least in principle) able to reproduce itself, is

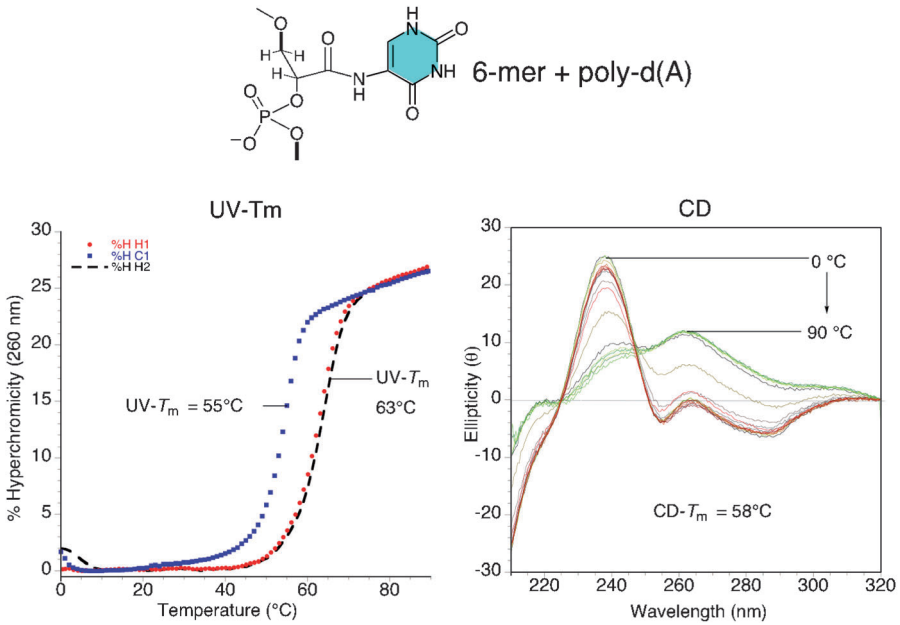


Figure 82. Notably strong cross-pairing between DNA and the homobasic hexamer sequence of an oligonucleotide derived from D-glyceric acid with 5-amino-2,4-dioxopyrimidine as recognition element (ca. 5+5 μM, 10 mM NaH_2PO_4, 0.1 mM Na_2EDTA, 1 M NaCl).[259]

from an overall standpoint only a first step on a very long road.

Every attempt as a chemist to picture in detail how many specific circumstances in a prebiotic environment would need to work together in order for an informational oligomer system of the oligophosphodiester or, for example, nucleo-oligodipeptide type of structure to be capable of constituting itself without the aid of enzymes must necessarily lead to a fundamental questioning of any such scenario. What in this context is to be scrutinized is not least the "robustness postulate" of classical prebiotic chemistry.

4.12. Constitutionally Labile ("Non-Robust") Base-Pairing Systems?

According to one of the implicit postulates of classical prebiotic chemistry, prebiotic molecules as well as reactions that might conceivably be assigned a role in the self-organization of matter must class as *robust*, otherwise they could not have been part of such a process.[62,216,217,222,260,261] The extent of our ignorance regarding the chemistry of biogenesis admonishes us to consider not only the robustness of starting materials, intermediates, and reactions, but also the possibility of the contrary. One would thereby distance oneself at the outset from ideas in which decisive processes may have occurred in an environment and at temperatures envisaged by the postulate of a biogenesis under hydrothermal conditions.[262] One would be led instead to pursue the notion of a "cold origin for life",[263] an alternative this author believes should from a chemical standpoint be preferred in any case.

The question is to be raised as to whether highly sensitive, non-robust structures and structure transformations, besides robust ones, might also have the potential to have played a decisive role in the critical early phases of prebiotic self-organization processes. Non-robust systems could respond more sensitively to accelerating and inhibiting factors, and—through contingent catalysis from outside and/or catalysts emerging within a system—be more easily assisted, successively guided, and finally controlled. In the vision of a "generational succession" of non-robust reaction patterns one would be dealing with half-life periods much shorter than would be the case with robust systems. However, one would require from such a scenario that systems nevertheless would ultimately achieve robustness, a requirement that represents an additional and, from a chemical point of view, especially serious challenge.

Examining the conventional demand for robustness in prebiotic chemistry is of direct concern in conjunction with the problems addressed in the preceding chapter: In a search for potential primordial informational oligomer systems, should not besides constitutionally robust types also non-robust ones be included? Whereas the former would be pictured as arising from activated monomers, and as copying themselves under conditions of kinetic reaction control, non-robust oligomer systems, by contrast, could assemble constitutionally under (partial) thermodynamic control. To be sure, the two pictures are not necessarily strict alternatives, since that of oligomer-assembly under (partial) thermodynamic control can in principle also apply to robust oligomer structures arising through dehydration steps; in such systems, depending on conditions, equilibrium can in principle always be established between formation and degradation, that is, between condensation and its reverse, through hydrolysis.[264] What is meant here is perhaps best illustrated by examples, for example, by formulas of oligomer systems such as those in Figure 83. These are strictly *hypothetical* examples of non-robust oligomer systems—exemplifying others and presumably also a host of unrecognized ones—whose formation would be conceivable without any special activation of the corresponding building blocks, and might occur under reaction conditions in which all the oligomerization steps would be reversible. Driving forces for their formation would necessarily have to come from environmental dehydration as well as from the pairing energy of selectively forming duplexes. In such a theoretical model, monomers and oligomers would be in a dynamic equilibrium with each other, and Watson–Crick base pairing would act as a selection factor, thereby causing selective formation of oligomer sequences stabilized by complementary sequences or templates.[265,266] To apply such a hypothesis of constitutionally labile self-assembling informational oligomers would make

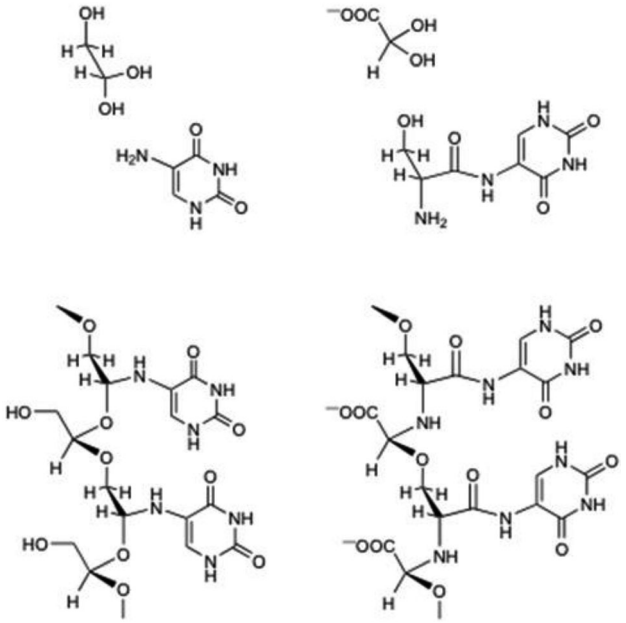


Figure 83. Hypothetical examples of non-robust (potential) pairing systems.

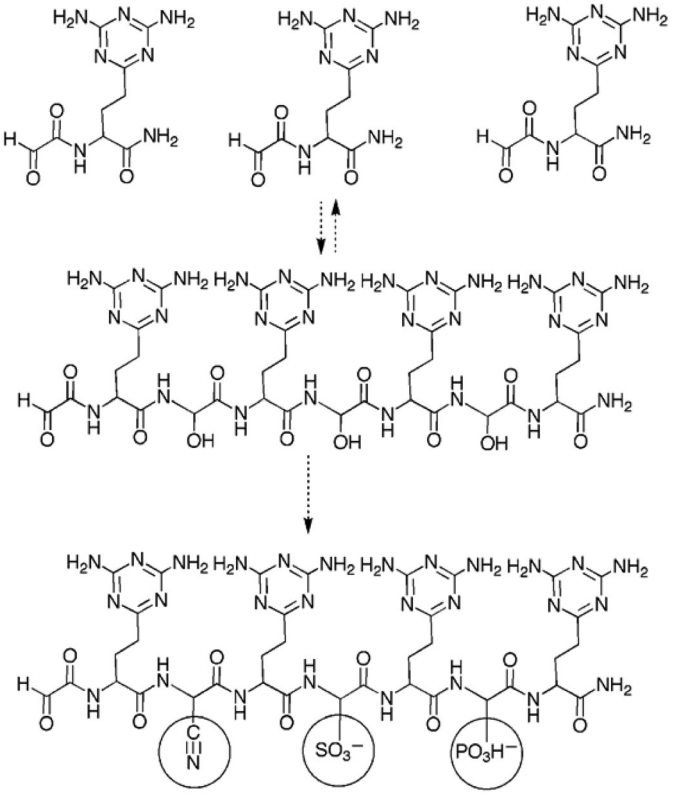


Figure 84. A non-robust (potential) pairing system that was studied in an exploratory way (cf. Figure 75).

sense only in combination with the ancillary concept that, in such a dynamic equilibrium system, there would have to emerge organocatalysts that would be in a position to exert influence over the formation of relevant components of the system. "Organocatalysis", the category into which phenotypic catalysis of informational oligomers of the type under consideration would fall, need not fundamentally require oligomer sequences of the length one would ordinarily assume necessary for biocatalysts. The astonishingly broad structural and functional diversity of organocatalysts, as has become apparent today through the recent renaissance of this field of research in organic chemistry, provides impetus for such lines of thought, and forces us to consider that the conceptual and experimental task of mapping the landscape of potentially primordial informational oligomers may need to include non-robust systems.

The ability to predict pairing potentials should be even more important for hypothetical non-robust oligomer systems than for robust systems, since to investigate experimentally their formation and properties will be significantly more difficult. Nevertheless, such pairing prognoses will be supported by the totality of experience from comparing predictions and corresponding experimental observations in robust systems. Within the framework of our studies of nucleo-oligodipeptides we have attempted to demonstrate experimentally, at least in an exploratory way, one example of oligomerization supposed to lead, under equilibrium conditions and assisted by a (robust!) template, to a non-robust oligomer from a non-activated monomer (Figure 84); but we were forced in the end to capitulate in the face of (not unexpected) analytical difficulties, and to leave open the interpretation of promising observations of a tentative nature (Figure 85).[267] Success has attended experiments recently carried out by Reza Ghadiri,[268] in which an informationally controlled assembly of oligodipeptide base sequences via reversible thioester group transfer was demonstrated under thermodynamic control using oligonucleotide templates. This model system apparently comes the closest to date to the scenario posed for the assembly of non-robust informational oligomer systems.

5. Attempt at a Forecast

A strategy for surveying the landscape of potential primordial informational oligomer systems is ultimately inspired by the "geneticist's" postulate, according to which emergence of a primitive genetic system—ultimately one corresponding to our present-day genetic oligomer systems—was the decisive step in biogenesis.[14a–e,198] This postulate contradicts the "metabolistic" vision,[10,14f,269,270] which sees the emergence of a genetic system of the type familiar today not as the beginning of evolution, but rather as an achievement of the evolution of an earlier "life". The latter in turn emanated from "proto-metabolic" reaction cycles supposed to have been autocatalytic, thus intrinsically replicative. Reference to this dichotomy of theoretical models is not intended here simply to assure formal balance in the allusion to contradictory pictures; the author believes rather that a comprehensive and purposeful experimental search for primordial informational oligomer and reaction systems, one that is forging ahead, is a way finally to reach chemical territory where "geneticists" and "metabolists" will be able to

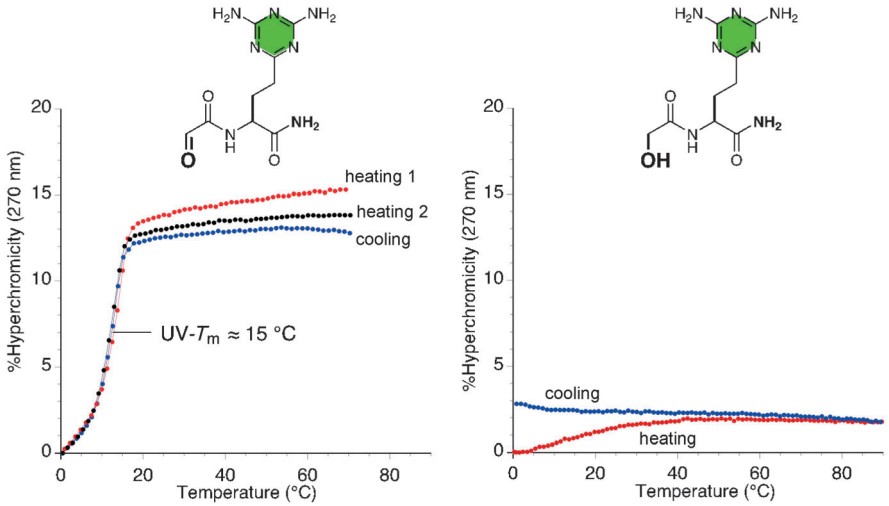


Figure 85. Observations from experiments toward demonstration of oligomerization of the monomer in Figure 84, glyoxyl-($T^{N,N}$)Glu, to a non-robust oligomer-complementary strand, hydroxyacetyl-Glu($T^{N,N}$), on an RNA template, r(U_{16}).[267] 10 mM Trisborate, pH 9.0, 1 M NaCl, 0.1 mM Na_2EDTA; RNA (10 μM), monomer (2 mM); measured 30 minutes after preparation of the sample; UV/T_m curves do not change within 6 days.

meet. To what extent informational replication systems that surface in the course of such a search will be distinguishable from ones we are familiar with today, both structurally and functionally, is just as difficult to anticipate as it is to predict how closely such systems will approach hypothetical autocatalytic cycles of metabolistic working hypotheses.

Were one to inquire about a *lowest common denominator* for geneticistic and metabolistic creeds, one such solution might be found in the demand for *reproducibility of catalysts* (not merely constituents), and indeed an obligation for ones that are functionally networked and, through their operation, capable of broadening their own structural and functional diversity.[271] The core of the difference in the two approaches is thus not to be seen in whether a system is or is not subject to "genetic" control, but rather in the *extent* of the constitutional and functional diversity of a (de facto "genetic") guidance system. With the goal of trying not so much to recognize what distinguishes the two models, but rather what they have in common, the term "genetic" as used here is meant to reflect *any form of system control through a carrier of catalytic functions that is self-reproducible autocatalytically, and which through its action is able to achieve diversification of catalyst structure and function, thereby conferring upon a system the potential to evolve*.[272] Even if a system were to acquire only to a limited extent the potential for self-diversification of structure and function of such catalysts, this potential would be consistent with the simplest and at the same time most austere "definition of life" at its most elementary level.[273] The central problem for a metabolistic scenario is the *diversification* of catalytic potential, in contrast to the (conventional) geneticistic model, where such diversity at least in principle is attainable essentially without limit. Primitive "life" in the sense of the geneticistic working hypothesis had an incomparably higher evolution potential than any type of "life" in a metabolistic model. Herein lies the root of the geneticistic creed.

If one were to inquire—in the sense of a "lowest common denominator" as referred to above—about equivalents of "genotype" and "phenotype" in a (conventional) metabolistic scenario, or about the molecular carriers of "genetic" information and the chemical means by which these in such a scenario might diversify, one would initially be tempted to write the question off as out of place. But it actually strikes at the heart of the difficulty associated with a metabolistic scenario. Appreciating this is almost a prerequisite to the drafting and potential experimental simulation of such scenarios. Whereas the notion of the interplay between genotypic and phenotypic in a geneticistic scenario involves questions of detail, but does no longer raise ones of a fundamental nature—the vision of the "RNA world" (see above) is evidence to the point—an attempt to locate equivalents of these two functions in a metabolistic scenario runs into a host of problems; it serves nevertheless for a review of similarities and differences between the two visions. Suited to formal exemplification of such a quest for a review is a scenario of the Morowitz–Wächterhäuser type,[10b,269c] because at its center is also to be found a "replicator". In the Morowitz proposal[269c] this would be a primitive equivalent of the reductive (CO_2-fixing) citric acid cycle that functions without enzymes. Here each constituent is a catalyst for its own replication, and simultaneously a catalyst for replication of all the other constituents of the cycle.[271] Diversification of the catalytic potential of such a system could come about in the simplest case in that extra-cyclic reaction pathways—environmentally induced, and proceeding from constituents of the cycle—would lead to reaction products among which there were (organo-) catalysts. If such a catalyst were in a position, directly or indirectly, to facilitate the generation or functioning of the parent cycle, this would imply emergence of a new catalytic cycle, interwoven with the parent cycle.

Although there are virtually no limits to *formal theoretical visions* of the emergence of diversifying autocatalytic reaction networks (cf., for example, ref. [270]), the "ifs and buts" from the chemist's standpoint are legion, and this is true for even the simplest variants on such scenarios, parent cycles included. Nevertheless, a word of caution is in order regarding judgments. In the realm of organic chemistry, a systematic search for potentially *primordial* autocatalytic cycles and networks has actually not even really begun.[274] On the other hand, organometallic chemistry already has at its disposal, in the justly famous Soai reaction,[275] a dramatic experimental illustration of the truly extraordinary selection potential that autocatalysis can unleash in chemical reality. Moreover, within pure organic chemistry, as already mentioned, one is on the brink of bringing about a renaissance in the exploration in concept, results, and insight of organocatalysis, which

sooner or later will fertilize in a lasting way our vision of the chemistry of biogenesis.[276]

Experimental efforts to simulate in a non-enzymatic way individual reaction steps in the reductive citric acid cycle have led to results that more nearly resemble refutation than support for the hypothesis underlying the experiments.[269b,277] The "formose reaction",[61] already referred to, has long been known to be a process that can operate autocatalytically.[278] This has for some time now been the focus of a series of remarkably consistent studies by Weber,[279] in which, under the heading "The Sugar Model", an attempt is being made to demonstrate experimentally the potential of sugars to act as a prebiotic source of biomolecules.[280,281] Work at the Scripps Research Institute is still in progress on the question of whether a chemistry starting with (thus far: formal) aquo-oligomers of carbon monoxide (glyoxylate and 2,3-dihydroxyfumarate)[282] might offer a complement to formose chemistry, and whether it conceals an as yet unknown family of autocatalytic cycles;[252] results to date support the first, but apparently not the second of the two propositions.[283] Since the early days of prebiotic chemistry there has been the notion that a chemistry of α-amino acids and oligopeptides brought about the onset of the pathway to biogenesis.[284] The high point to date of experimental studies prompted by this view is probably the work of Ghadiri on model networks of auto- and cross-catalytic self-replicating (synthetic) polypeptides.[16c,d,285] Of importance in this context is research on organocatalysis probing the catalytic potential of amino acids and oligopeptides, even if most of this work is planned and pursued with a purely chemical motivation.[286] The renaissance in the area of "organocatalysis" gained impetus primarily from the theoretical and also practical significance of the topic "enantioselectivity". Not least for this reason have ever more publications appeared recently dealing (or allegedly dealing) with the question of the origin of "homochirality" of biological building blocks and biopolymers.[287,288] Although as a consequence of this work our present-day view regarding possible sources of biological "homochirality" became enriched through important options, evaluation of their etiological relevance seems premature so long as we remain in the dark regarding the constitutional nature and properties of the primordial "take-off" replication system.[289]

The origin of homochirality is only one of the many central questions that are not dealt with here.[290] Another has to do with chemical and geochemical options for compartmentalization[14b] of chemical events along the road to biogenesis; spatial individualization of primordial functional units is generally regarded as a prerequisite to a system's potential to evolve. This includes questions of the role of mineral surfaces and layer minerals, or of intersection zones between lipidic and aqueous reaction spaces. All life on our earth is cellular; the "most biological" of attempts at defining the origin of life therefore equates the latter with emergence of the cell.[291] Central questions that arise within the concept of an autotrophic origin of life relate to site and nature of geochemical material and energy sources, and to chemical scenarios for assimilation of carbon, which in the end must entail an influx from reductive fixation of carbon dioxide.

What has apparently so far been overlooked is the possibility that the earth's crust itself could have originally been a reservoir of precursors of reactive carbon- and heteroatom-containing compounds, precursors that could have been created and conserved in a free-energy retaining form during the solidification process of the crust, and might then have been successively exposed to hydrolysis at "hot-spots" near the surface, releasing reactive organics such as HCN and cyanamide.[292]

What also cannot be discussed here are boundary conditions for biogenesis, aspects that extend far beyond chemical problems of the type discussed, reaching to questions dealing with the evolution of our planet and its geochemical environment, and even farther to "habitable zones" in extra-solar planetary systems. The scientific dimensions of the problem of life's origin are given expression in the overall spectrum of such questions, as are the dimensions of the challenges to that branch of science ultimately responsible for tracking down the central events that led to living matter: chemistry.

What are the chances of finding lasting answers? These will depend on whether the hypothesis of a heterotrophic or that of an autotrophic origin lies closer to reality (which, of course, could be between the two extremes). Models of an autotrophic origin lean on the assumption that the start of a primitive metabolism would have been easier out of a chemically simple environment than out of a chemically complex one.[10] From a chemical point of view, conceptual deconstruction and experimental reconstruction of a constitutionally and spatially compact autotrophic origin—if there were to have been such a thing—would seem much more feasible than reconstruction of a heterotrophic beginning out of the legendary "prebiotic soup".

In a search for the chemistry of emergence of the type of life to which we belong, one needs to pay strict heed to molecular guides that all "*carry a message which it is our job to decipher*". Apart from nucleic acids and proteins, to such guides belong above all the archaic cofactors: vital to the metabolisms of all currently living things, these contain structural components that for us act as vitamins, with good grounds for believing that they or their precursors already played such a role in the critical phase of biogenesis, namely that of "vitamins" for the beginnings of life—*vitamins* in the true sense of the word. The idea of their pre-biological ancestry has impressive chemical support, especially in the case of riboflavin; so far—with the possible exception of the porphinoids—for none of the structurally complex cofactors other than riboflavin has a "chemomimetic" nature for biosynthesis of the core structure been so convincingly demonstrated through non-enzymatic simulation of chemically complex biosynthetic steps under potentially prebiotic conditions.[293,294] Clarifying by experimental synthetic analysis the question of whether the structures of the archaic cofactors are elementary to the extent of being potentially prebiotic could well provide a breakthrough at the front line of research into the problem of the chemistry of biogenesis.

"We'll never be able to know" is a truism that leads to resignation with respect to any experimental effort to search for the chemistry of life's origin. But such resignation runs

radically counter to the challenge imposed upon chemistry as a natural science. Notwithstanding the prognosis according to which the shortest path to understanding the metamorphosis of the chemical into the biological is by way of experimental modeling of "artificial chemical life", the scientific search for the route nature adopted in creating the life we know will arguably never truly end. It is, after all, part of the search for our own origin.

Epilogue and Acknowledgement

The work described in the preceding retrospective summary on an etiology of vitamin B_{12} and nucleic acid structure has extended over a period of more than three decades. I wish to acknowledge and thank all my former doctoral candidates and postdoctoral associates who, before and after my retirement at ETH, achieved and were co-responsible for the research results described, whether in the Laboratory of Organic Chemistry at the ETH, the Institute for Pharmacy at the University of Frankfurt, or the Scripps Research Institute in La Jolla. I have refrained from listing all my former co-workers due simply to their sheer number, and for this I beg their forgiveness; this applies especially to those whose names do not appear in any of the cited literature, even though their contributions involved no less dedication or commitment. Whenever we failed to achieve a goal in our teaching and learning community—the way I would like to think of our research group —it was ultimately not the fault of the learners, but of the one responsible for setting the goals. With each passing year it has become clearer to me what a privilege it has been to be able to dedicate myself with such a community to pure scientific effort, and to pursue the questions that are the subject of this retrospective.

The studies carried out prior to my retirement in 1992 were supported both organizationally and financially by the administration of the ETH Zürich, the Swiss National Science Foundation, the (former) CIBA Corporation in Basel, as well as Firmenich & Co. in Geneva. After 1992, continued efforts by a small research group at the ETH for more than a decade were made possible thanks to organizational accommodation on the part of my colleagues in the Laboratory of Organic Chemistry, and financial support from the administration of the ETH, the CIBA and Firmenich companies, as well as the newly established Novartis AG. In the years 1994–1997 a part of my research group enjoyed guest privileges at the Pharmaceutical Institute of the Johann Wolfgang-Goethe-Universität in Frankfurt. After 1996 the focal point of my research activities shifted to a research group managed jointly with Ram Krishnamurthy in the Skaggs Institute for Chemical Biology at the Scripps Research Institute (TSRI) in La Jolla (California). Our work there was supported financially by the Skaggs Foundation and the administration of TSRI, as well as by NASA. In 2009, I retired again—this time voluntarily.

Behind this course of events stands a number of people whom I feel a special obligation to thank. I start with my colleagues Duilio Arigoni and Andrea Vasella, both of whom in the meantime are also retired, in whose former research groups in the organic chemistry laboratory at the ETH after 1992 I was administratively a guest. The magisterial-unconventional help provided to me in those years in large measure by Ralf Hütter, at the time Vice President for Research at the ETH Zürich, I shall not forget. In the late eighties, when the transition from B_{12} to homo-DNA chemistry demanded from our group a reorientation in working techniques, we may not have succeeded without the help of Christian Leumann.I owe him my special acknowledgment. To two friends, Gerhard Quinkert and Christian Noe, I am indebted for a wonderful period of mutual activity with a group of postdoctoral fellows in the newly-constructed science building at the University of Frankfurt. My very special thanks go out to the President of TSRI, Richard Lerner; my wife Elisabeth and I are deeply grateful for the decision on his part that provided us with an opportunity to experience and enjoy the precious and unforgettable years in California. I wish to include in my thanks the friends and colleagues among the Scripps faculty, with whom I was in close personal and scientific contact over all those years, a contact that to me was precious, reassuring and rejuvenating.

Very special appreciation and lasting gratitude are due to my younger colleague at TRSI, Ramanarayanan Krishnamurthy. "Ram" arrived in 1992, after completing doctoral studies at Ohio State University (Prof. David Hart), as a postdoctoral fellow in my newly-formed "post-emeritus" research group at the ETH. He subsequently was a research visitor at SIO in La Jolla as a scientific associate of Gustaf Arrhenius. In 1996 he became the first member, cofounder, and co-organizer of my research group at TSRI. He advanced from research associate to associate professor at TSRI, and during the course of thirteen years whenever I was in Switzerland rather than in California, Ram looked after my La Jolla postdoctoral group both scientifically and administratively.

Ram Krishnamurthy

To have been offered in the autumn of my professional life the collaboration and loyalty of such a humanly endearing, professionally competent, and intellectually brilliant young researcher in the person of Ram Krishnamurthy is an extraordinary gift. Everything that our research group at TSRI accomplished over the years bears the imprint of his thoughts and participation, his organizing ability, and his active and constant concern for everything required for high quality experimental work by a research group in a chemical laboratory. In the course of our collaboration, Ram moved more and more from the role of learner to that of teacher, becoming a source of skill and knowledge upon which the other co-workers were able to draw. Since 2009 Ram has been working with his own independent research group at TSRI.

Finally, I would like to express my gratitude to Dr. William E. Russey, who translated the German manuscript of this paper into English. His empathy for the author's intentions, his linguistic as well as chemical competence, and his patience with an author who cannot stop changing a manuscript, made the collaboration with him a fine and memorable experience.

Received: May 30, 2011
Published online: December 7, 2011

Translated by Dr. William Russey, Huntingdon

[1] V. Prelog, *Naturwiss. Rundsch.* **1985**, *38*, 259 – 2662.
[2] Webster, Unabridged Dictionary, 2nd ed., **1979**.
[3] "Aitiology" in *Oxford Dictionary*, 2nd ed., Oxford University Press, Oxford, **2002**.
[4] "Toward a Chemical Etiology of the Natural Nucleic Acid's Structure": A. Eschenmoser in *Proceedings of the Robert A. Welch Foundation 37th Conference on Chemical Research "40 Years of the DNA Double Helix"*, Houston, **1993**, p. 201.
[5] a) S. L. Miller, *Science* **1953**, *117*, 528; S. L. Miller, *J. Am. Chem. Soc.* **1955**, *77*, 2351; b) J. Oró, *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407; J. Oró, A. P. Kimball, *Arch. Biochem. Biophys.* **1961**, *94*, 217.
[6] Molière, "Le Bourgeois Gentilhomme", Act II, scence IV.
[7] S. L. Miller, L. E. Orgel, *The Origins of Life on Earth*, Prentice-Hall, Englewood Cliffs, **1974**; C. Will, J. Bada, *The Spark of Life*, Perseus, Cambridge, **2000**.
[8] a) A. I. Oparin, *Origin of Life*, *2nd ed.*, Dover, New York, **1952**, p. 153; see also the English version (published in ref. [9], pp. 199 – 234) of the original first Russian edition that appeared in 1924; b) J. B. S. Haldane, *Ration. Annu.* **1929**, *148*, 3; reproduced in Ref. [9] (pp. 242 – 249).
[9] a) J. D. Bernal, *The Origin of Life*, Weidenfeld & Nicolson, London, **1967;** b) Chapter "Postscript" in ref. [9a], p. 194.
[10] a) G. Wächtershäuser, *Microbiol. Rev.* **1988**, 452; b) G. Wächtershäuser, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 200; c) G. Wächtershäuser, *Philos. Trans. R. Soc. London Ser. B* **2006**, *361*, 1787; d) H. Hartman, *J. Mol. Evol.* **1975**, *4*, 359.
[11] Y. Wolman, W. J. Haverland, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1972**, *69*, 809; J. R. Cronin, S. Pizzarello, *Science* **1997**, *275*, 951; M. P. Callahan, K. E. Smith, H. J. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House, J. P. Dworkin, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13995.
[12] W. M. Irvine, *Origins Life Evol. Biosphere* **1998**, *28*, 365; M. C. McCarthy, P. Thaddeus, *Chem. Soc. Rev.* **2001**, *30*, 177.
[13] a) C. De Duve, *Vital Dust: Life as a Cosmic Imperativ*, Basic Books, Harper Collins, New York, **1995**; b) C. De Duve, *Blueprint for a Cell: The Nature and Origin of Life*, Patterson, Burlington, **1991**.
[14] a) M. Eigen, *Naturwissenschaften* **1971**, *58*, 465; b) H. Kuhn, *Angew. Chem.* **1972**, *84*, 838; *Angew. Chem. Int. Ed. Engl.* **1972**, *11*, 798; c) M. Eigen, M. Schuster, *Naturwissenschaften* **1978**, *65*, 341; d) H. Kuhn, J. Waser, *Angew. Chem.* **1981**, *93*, 495; *Angew. Chem. Int. Ed. Engl.* **1981**, *20*, 500 – 520; e) S. Lifson, *J. Mol. Evol.* **1997**, *44*, 1; f) S. A. Kauffmann, *J. Theor. Biol.* **1986**, *119*, 1; g) D. Segré, D. Ben-Eli, D. Deamer, D. Lancet, *Origins Life Evol. Biosphere* **2001**, *31*, 119; h) P. Decker, *Ann. N. Y. Acad. Sci.* **2006**, *316*, 236; i) E. Szathmary, *Philos. Trans. R. Soc. London Ser. B* **2006**, *361*, 1761; j) H. Morowitz, E. Smith, *Complexity* **2007**, *13*, 51; k) A. Pross, *J. Phys. Org. Chem.* **2008**, *21*, 724.
[15] A. Eschenmoser, *Origins Life Evol. Biosphere* **1994**, *24*, 389.
[16] a) M. Kindermann, I. Stahl, M. Reimold, W. M. Pankau, G. von Kiedrowski, *Angew. Chem.* **2005**, *117*, 6908; *Angew. Chem. Int. Ed.* **2005**, *44*, 6750; b) E. J. Hayden, G. von Kiedrowski, N. Lehman, *Angew. Chem.* **2008**, *120*, 8552; *Angew. Chem. Int. Ed.* **2008**, *47*, 8424; c) D. H. Lee, K. Severin, Y. Yokobayashi, M. R. Ghadiri, *Nature* **1997**, *390*, 591; d) G. Ashkenasy, R. Jagasia, M. Yadav, M. R. Ghadiri, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 10872.
[17] A. Eschenmoser, *Angew. Chem.* **1988**, *100*, 5; *Angew. Chem. Int. Ed. Engl.* **1988**, *27*, 5.
[18] Y. Yamada, D. Miljikovic, P. Wehrli, B. Golding, P. Löliger, R. Keese, K. Müller, A. Eschenmoser, *Angew. Chem.* **1969**, *81*, 301; *Angew. Chem. Int. Ed. Engl.* **1969**, *8*, 343; A. Eschenmoser, *Pure Appl. Chem.* **1969**, *20*, 1. For X-ray structure determinations of synthetic A/D-seco-corrin and corrin complexes in both series of corrin model syntheses, see J. D. Dunitz, E. F. Meyer, Jr., *Proc. R. Soc. A* **1965**, *288*, 324; J. D. Dunitz, E. F. Meyer, Jr., *Helv. Chim. Acta* **1971**, *54*, 77; M. Dobler, J. D. Dunitz, *Helv. Chim. Acta* **1971**, *54*, 90; M. Currie, J. D. Dunitz, *Helv. Chim. Acta* **1971**, *54*, 98.
[19] a) "Roads to Corrins", A. Eschenmoser, *Quart. Revs.* **1970**, *24*, 366; b) "Studies on Organic Synthesis" A. Eschenmoser in *23rd Int. Congress of Pure and Applied Chemistry, IUPAC, Supplement Vol. 2*, Butterworths, London, **1971**, p. 69; c) "Chemical Synthesis of Corrinoids: Current Problems and Recent Advances": A. Eschenmoser in *Vitamin B_{12}* (Hrsg.: B. Zagalak, W. Friedrich), Walter de Gruyter, Berlin, **1979**, pp. 89 – 117; d) "Über organische Naturstoffsynthese: Von der Synthese des Vitamins B_{12} zur Frage nach dem Ursprung der Corrinstruktur": A. Eschenmoser, *Nova Acta Leopold.* **1982**, *55*, 5.
[20] a) W. Fuhrer, P. Schneider, W. Schilling, H. Wild, J. Schreiber, A. Eschenmoser, *Chimia* **1972**, *26*, 320; H. Maag, N. Obata, A. Holmes, P. Schneider, W. Schilling, J. Schreiber, A. Eschenmoser, *Chimia* **1972**, *26*, 320; b) A. Eschenmoser, *Naturwissenschaften* **1974**, *61*, 513; c) A. Eschenmoser, C. Wintner, *Science* **1977**, *196*, 1410. The last mentioned publication provides in its Figure 16 the formula scheme for the photochemical variant of the vitamin B_{12}-synthesis, and provides a complete list of the ETH theses that describe in full detail the experiments carried out by the ETH group in the context of the vitamin B_{12} project. All these theses are accessible with their full text (in German) in the Internet under the names of the respective authors (http://www.e-collection.ethbib.ethz.ch). The authors and their theses are as follows: Jost Wild (Diss. ETH Nr. 3492; **1964**), Urs Locher (Nr. 3611; **1964**), Alexander Wick (Nr. 3617; **1964**), Peter Löliger (Nr. 4074; 1**968**), René Wiederkehr (Nr. 4239; **1968**), Paul Dubs (Nr. 4297; **1969**), Peter Schneider (Nr. 4819; **1972**), Hans Maag (Nr. 5173; **1973**), Walter Fuhrer (Nr. 5158; **1973**) and Werner Schilling (Nr. 5352; **1974**). In the experimental parts of the theses one finds detailed descriptions of the reactions steps formulated in Figure 16 of ref. [20c]; incorporated in these reports are the experimental contributions of Dr. Jakob Schreiber and the following postdoctoral collaborators who also were involved in this part of the B_{12} project: Dr. Bernard Golding, Dr. Andrew Holmes, Prof. Dr. Naruyoshi Obata, Dr. Naoto Hashimoto, Dr. Heinz Gschwend, Dr. Fritz Karrer, Dr. Dan Becker, Dr. Walter Hunkeler, Dr. John Gleason, and Dr. Brian Place.
[21] W. Fuhrer, "Totalsynthese von Vitamin B_{12}: Der photochemische Weg", Diss. ETH Zürich, Nr. 5158, **1973**, (312 pp.)
[22] a) R. Neier, "Ueber den Mechanismus der photochemischen A/D-Secocorrin-Corrin-Cycloisomerisierung", Diss. ETH Zürich, Nr. 6178, **1978**, (179 pp.); b) H.-J. Wild, "Die Synthese von Corrin-Komplexen durch photochemische A/D-Cycloisomerisierung", Diss. ETH Zürich, Nr. 4848, **1972**, (142 pp.).
[23] M. Gardiner, A. J. Thomson, *J. Chem. Soc. Dalton* **1974**, 820.
[24] A. Eschenmoser, *Chem. Soc. Rev.* **1976**, *5*, 377.
[25] a) A. Pfaltz, B. Hardegger, P. M. Müller, S. Farooq, B. Kräutler, A. Eschenmoser, *Helv. Chim. Acta* **1975**, *58*, 1444; b) B. Kräutler, A. Pfaltz, R. Nordmann, K. O. Hodgson, J. D. Dunitz, A. Eschenmoser, *Helv. Chim. Acta* **1976**, *59*, 924; c) A. Pfaltz, N. Bühler, R. Neier, K. Hirai, A. Eschenmoser, *Helv. Chim. Acta* **1977**, *60*, 2653.
[26] a) V. Rasetti, A. Pfaltz, C. Kratky, A. Eschenmoser, *Proc. Natl. Acad. Sci. USA* **1981**, *78*, 16; b) V. Rasetti, K. Hilpert, A. Fässler, A. Pfaltz, A. Eschenmoser, *Angew. Chem.* **1981**, *93*, 1108; *Angew. Chem. Int. Ed. Engl.* **1981**, *20*, 1058.

[27] R. B. Woodward, *Pure Appl. Chem.* **1968**, *17*, 519; R. B. Woodward, *Pure Appl. Chem.* **1971**, *25*, 283; R. B. Woodward, *Pure Appl. Chem.* **1973**, *33*, 145; "Synthetic Vitamin B_{12}": R. B. Woodward in *Vitamin B_{12}* (Eds.: B. Zagalak, W. Friedrich), Walter de Gruyter, Berlin, **1979**, pp. 37–87.

[28] G. H. Cookson, C. Rimington, *Biochem. J.* **1954**, *57*, 476; D. Mauzerall, *J. Am. Chem. Soc.* **1960**, *82*, 2601.

[29] G. Ksander, G. Bold, R. Lattmann, C. Lehmann, T. Früh, Y.-B. Xiang, K. Inomata, H. P. Buser, J. Schreiber, E. Zass, A. Eschenmoser, *Helv. Chim. Acta* **1987**, *70*, 1115.

[30] J. E. Johansen, V. Piermattie, C. Angst, E. Diener, C. Kratky, A. Eschenmoser, *Angew. Chem.* **1981**, *93*, 273; *Angew. Chem. Int. Ed. Engl.* **1981**, *20*, 261; R. Waditschatka, A. Eschenmoser, *Angew. Chem.* **1983**, *95*, 639; *Angew. Chem. Int. Ed. Engl.* **1983**, *22*, 630.

[31] "Chemistry of Corphinoids": A. Eschenmoser, *Ann. N. Y. Acad. Sci.* **1986**, *471*, 108.

[32] A. P. Johnson, P. Wehrli, R. Fletcher, A. Eschenmoser, *Angew. Chem.* **1968**, *80*, 622; *Angew. Chem. Int. Ed. Engl.* **1968**, *7*, 623; P. M. Müller, S. Farooq, B. Hardegger, W. S. Salmond, A. Eschenmoser, *Angew. Chem.* **1973**, *85*, 914; *Angew. Chem. Int. Ed.* **1973**, *85*, 954.

[33] C. Kratky, R. Waditschatka, C. Angst, J. E. Johansen, J. C. Plaquevent, J. Schreiber, A. Eschenmoser, *Helv. Chim. Acta* **1985**, *68*, 1312.

[34] F. Kreppelt, "Regioselektive Rekonstituierung von Vitamin B_{12} durch Nukleotidierung von Cobyrinsäure-heptakis(cyanmethyl)ester", Diss. ETH Zürich, Nr. 9458, **1991**. In this thesis are incorporated the experimental contributions of Günter Bartels (postdoctorate 1980–1981) and Rolf Nussberger (PhD student, deceased 21.7.1985).

[35] In the wake of these investigations, Bernhard Kräutler (at that time cand. habil. at ETH) prepared by regioselective partial syntheses all three structures with the alternative nucleotide loops, and determined the corresponding base-on/base-off equilibria experimentally (unpublished work; private communication Prof. B. Kräutler, Innsbruck). A recent discovery by Kräutler that is also relevant for the etiology of the B_{12}-nucleotide loop is that of the "nor-pseudovitamin B_{12}" from *Sulfurospirillum multivorans*, which, besides containing a different nucleobase, lacks the C-methyl group at the loop. Investigations carried out in this context on the influence of that methyl group on the base-on/base-off equilibrium in vitamin B_{12} are described in: (P. Butler, M.-O. Ebert, A. Lyskowski, K. Gruber, C. Kratky, B. Kräutler, *Angew. Chem.* **2006**, *118*, 1004; *Angew. Chem. Int. Ed.* **2006**, *45*, 989). See also the work of Felix Zelder on B_{12}-Mimetika (K. Zhou, F. Felder, *Angew. Chem.* **2010**, *122*, 5305; *Angew. Chem. Int. Ed.* **2010**, *49*, 5178), as well as Kräutler's observations on his recently synthesized "iso-vitamin B_{12}" (S. Murtaza, M. Ruetz, K. Gruber, B. Kräutler, *Chem. Eur. J.* **2010**, *16*, 10984).

[36] A. R. Battersby, *Science* **1994**, *264*, 1551; F. Blanche, B. Cameron, J. Crouzet, L. Debussche, D. Thibaut, M. Vuilhorgne, F. J. Leeer, A. R. Battersby, *Angew. Chem.* **1995**, *107*, 421; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 383.

[37] A. I. Scott, *Acc. Chem. Res.* **1990**, *23*, 308; A. I. Scott, *Angew. Chem.* **1993**, *105*, 1281; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1223; A. I. Scott, *J. Org. Chem.* **2003**, *68*, 2529.

[38] A. Fässler, "Synthese und Eigenschaften von Modellkomplexen des Coenzyms F-430", Diss. ETH Zürich, Nr. 7799, **1985** (pp. 81–84; 201–207).

[39] K. Hilpert, "Synthetische Untersuchungen zur Biosynthese des Corrinsystems", Diss. ETH Zürich, Nr. 7390, **1983**.

[40] G. L. Holliday, J. M. Thorton, A. Marquet, A. G. Smith, F. Rébeillé, R. Mendel, H. L. Schubert, A. D. Lawrence, M. J. Warren, *Nat. Prod. Rep.* **2007**, *24*, 972.

[41] R. Waditschatka, E. Diener, A. Eschenmoser, *Angew. Chem.* **1983**, *95*, 641; *Angew. Chem. Int. Ed. Engl.* **1983**, *22*, 631; C. Leumann, K. Hilpert, J. Schreiber, A. Eschenmoser, *J. Chem. Soc. Chem. Commun.* **1983**, 1401.

[42] R. E. Eakin, *Proc. Natl. Acad. Sci. USA* **1963**, *49*, 360; "Polynucleotide Replication and the Origin of Life": L. E. Orgel, J. E. Sulston in *Prebiotic and Biochemical Evolution* (Eds.: A. P. Kimball, J. Oro), North-Holland, Amsterdam, **1971**, pp. 89–94; B. White, *J. Mol. Evol.* **1976**, *7*, 101.

[43] H. von Dobeneck, B. Hansen, E. Vollmann, *Z. Naturforsch. B* **1972**, *27*, 922.

[44] J. J. Scott, *Biochem. J.* **1956**, *62*, 6P; S. Granick, D. Mauzerall, *J. Biol. Chem.* **1958**, *232*, 1119; A. I. Scott, A. Townsend, K. Okada, M. Kajiwara, *Trans. N. Y. Acad. Sci.* **1973**, *35*, 72; B. Franck, H. Stratmann, *Heterocycles* **1981**, *15*, 919; A. R. Butler, S. D. George, *Tetrahedron* **1992**, *48*, 7879; R. Neier, *J. Heterocycl. Chem.* **2000**, *37*, 487; J. S. Lindsey, V. Chandrashaker, M. Taniguchi, M. Ptasek, *New J. Chem.* **2011**, *35*, 65.

[45] S. I. Beale, M. H. Gold, S. Granick, *Phytochemistry* **1979**, *18*, 441.

[46] T. Früh, "Zur Chemie der Uroporphyrinogen-octanitrile", Diss. ETH Zürich, Nr. 8054, **1986**.

[47] C. Lehmann, "Untersuchungen über Uroporphyrinogen-octanitrile", Diss. ETH Zürich, Nr. 8215, **1987**; C. Lehmann, B. Schweizer, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1997**, *80*, 1412.

[48] T. Oberhauser, "Untersuchungen über Pyrrocorphine", Diss. ETH Zürich, Nr. 9741, **1992**.

[49] P. M. Müller, "Corphin: ein neuartiges porphinoid-corrinoides Ligandsystem", Diss. ETH Zürich, Nr. 5135, **1973**.

[50] K. Zimmermann, "Vergleichende Untersuchungen an Cobaltcorphin- und Cobaltcorrinkomplexen", Diss. ETH Zürich, Nr. 9038, **1989**.

[51] J. Halpern, *Science* **1985**, *227*, 869; also cf. Halpern's work "Why does Nature not use the Porphyrin Ligand in Vitamin B_{12}?" (M. K. Geno, J. Halpern, *J. Am. Chem. Soc.* **1987**, *109*, 1238).

[52] B. P. Hay, R. G. Finke, *J. Am. Chem. Soc.* **1986**, *108*, 4820; B. P. Hay, R. G. Finke, *J. Am. Chem. Soc.* **1987**, *109*, 8012.

[53] R. P. Gunsalus, R. S. Wolfe, *FEMS Microbiol. Lett.* **1978**, *3*, 191; P. Schönheit, J. Moll, R. K. Thauer, *Arch. Microbiol.* **1979**, *123*, 105; G. Diekert, B. Klee, R. K. Thauer, *Arch. Microbiol.* **1980**, *124*, 103.

[54] A. Pfaltz, B. Jaun, A. Fässler, A. Eschenmoser, R. Jaenchen, H. H. Gilles, G. Diekert, R. K. Thauer, *Helv. Chim. Acta* **1982**, *65*, 828; D. A. Livingston, A. Pfaltz, J. Schreiber, A. Eschenmoser, D. Ankel-Fuchs, J. Moll, R. Jaenchen, R. K. Thauer, *Helv. Chim. Acta* **1984**, *67*, 334; A. Fässler, A. Kobelt, A. Pfaltz, A. Eschenmoser, C. Bladon, A. R. Battersby, R. K. Thauer, *Helv. Chim. Acta* **1985**, *68*, 2287.

[55] G. Färber, W. Keller, C. Kratky, B. Jaun, S. Spinner, A. Kobelt, A. Eschenmoser, *Helv. Chim. Acta* **1991**, *74*, 697.

[56] A. Fässler, A. Pfaltz, B. Kräutler, A. Eschenmoser, *J. Chem. Soc. Chem. Commun.* **1984**, 1365; C. Kratky, A. Fässler, A. Pfaltz, B. Kräutler, B. Jaun, A. Eschenmoser, *J. Chem. Soc. Chem. Commun.* **1984**, 1368.

[57] S. Scheller, M. Goenrich, R. Boecher, R. K. Thauer, B. Jaun, *Nature* **2010**, *465*, 606; S. Scheller, M. Goenrich, S. Mayr, R. K. Thauer, B. Jaun, *Angew. Chem.* **2010**, *122*, 8289; *Angew. Chem. Int. Ed.* **2010**, *49*, 8112; B. Jaun, R. K. Thauer, *Met. Ions Life Sci.* **2007**, *2*, 323; B. Jaun, R. K. Thauer, *Met. Ions Life Sci.* **2009**, *4*, 115.

[58] A. Lazcano, R. Guerrero, L. Margulis, J. Oro, *J. Mol. Evol.* **1988**, *27*, 283.

[59] "Prospects for Understanding the Origin of the RNA World" G. F. Joyce, L. E. Orgel in *The RNA World*, *2nd ed.* (Eds.: R. F.

Gesteland, T. R. Cech, J. F. Atkins), Cold Spring Harbor Laboratory Press, Plainview, **1999**, pp. 49.
[60] D. Müller, S. Pitsch, A. Kittaka, E. Wagner, C. E. Wintner, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1410.
[61] A. Butlerow, *C. R. Hebd. Seances Acad. Sci.* **1861**, *53*, 145; A. Butlerow, *Justus Liebigs Ann. Chem.* **1861**, *120*, 295; cf. a summary of literature referring to the "formose reaction" in ref. [60].
[62] R. Shapiro, *Origins Life Evol. Biosphere* **1988**, *18*, 71.
[63] P. Decker, H. Schweer, R. Pohlmann, *J. Chromatogr.* **1982**, *244*, 281.
[64] J. D. Bernal, *The Physical Basis of Life*, Routledge and Kegan Paul, London, **1951**.
[65] Analysis of aldolization product mixtures: desalting by ion exchange, followed by enzymatic dephosphorylation with alkaline phosphatase, reduction with excess $NaBH_4$ at room temperature, peracetylation of sugar alcohol mixtures with acetic anhydride in pyridine at 100 °C, and gas chromatography of the mixture of sugar alcohol acetates (for details see ref. [60]).
[66] S. L. Miller, *Cold Spring Harbor Symp. Quant. Biol.* **1987**, *52*, 17.
[67] G. Zubay, *Origins Life Evol. Biosphere* **1998**, *28*, 13.
[68] A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science* **2004**, *303*, 165; Q. Li, A. Ricardo, S. A. Benner, J. D. Winefordner D. H. Powell, *Anal. Chem.* **2005**, *77*, 4503. Vgl. auch B. E. Prieur, *C. R. Acad. Sci. Ser. IIc* **2001**, 667; A. F. Amaral, M. M. Marques, J. A. L. da Silva, J. J. F. Frausto da Silva, *New J. Chem.* **2008**, *32*, 2043; E. S. Grew, J. L. Bada, R. M. Hazen, *Origins Life Evol. Biosphere* **2011**, DOI: 10.1007s11084-010-9233-y.
[69] G. Quinkert, E. Egert, C. Griesinger, *Apekte der Organischen Chemie – Struktur*, Verlag Helvetica Chimica Acta, Basel, **1995**, p. 102; G. Quinkert, E. Egert, C. Griesinger, *Apects in Organic Chemistry—Structure*, Verlag Helvetica Chimica Acta, Basel, **1996**, p. 101.
[70] For the formation of the allose derivative as main product of the aldolization of a glycolaldehyde derivative, the hydroxy group of which is protected as an acetal, cf. ref. [116].
[71] TSRI = The Scripps Research Institute; SIO = Scripps Institute of Oceanography.
[72] K. Kuma, B. Gedulin, B. Paplawski, G. Arrhenius, *Origins Life Evol. Biosphere* **1989**, *19*, 573.
[73] S. Pitsch, A. Eschenmoser, H. Gedulin, S. Hui, G. Arrhenius, *Origins Life Evol. Biosphere* **1995**, *25*, 297.
[74] R. Krishnamurthy, S. Pitsch, G. Arrhenius, *Origins Life Evol. Biosphere* **1999**, *29*, 139.
[75] For the preparation of glycoladehyde phosphate by ozonolysis of allylphosphate cf. ref. [60], and references therein.
[76] S. Drenkard, J. Ferris, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1373; E. Wagner, Y.-B. Xiang, K. Baumann, J. Gück, A. Eschenmoser, *Helv. Chim. Acta* **1990**, *73*, 1391; Y.-B. Xiang, S. Drenkard, K. Baumann, D. Hickey, A. Eschenmoser, *Helv. Chim. Acta* **1994**, *77*, 2209; J. E. Dickens, W. M. Irvine, M. Ohishi, G. Arrhenius, S. Pitsch, A. Bauder, F. Müller, A. Eschenmoser, *Origins Life Evol. Biosphere* **1996**, *26*, 97.
[77] S. Pitsch, E. Pombo-Villar, A. Eschenmoser, *Helv. Chim. Acta* **1994**, *77*, 2251; vgl. auch S. Pitsch, A. Eschenmoser, *Pol. J. Chem.* **1994**, *68*, 2383.
[78] R. Krishnamurthy, S. Pitsch, unpublished work at SIO.[71]
[79] Cf. the literature survey in ref. [81].
[80] In the context of attempts made at TSO to achieve a direct phosphorylation of glycolaldehyde with cyclo-triphosphate, R. Krishnamurthy had made a literature search on phosphorylations by this reagent and thereby encountered the work of Quimby and Feldmann (see below) on the ammonolysis of cyclotriphosphate to amidotriphosphate. Later at TSRO, it was our awareness of the existence of this work that inspired recognition and subsequent experimental testing of the concept of intramolecular phosphorylation as depicted in Figure 20. O. T. Quimby, T. J. Flautt, *Z. Anorg. Allg. Chem.* **1958**, *296*, 220; V. W. Feldmann, E. Z. Thilo, *Z. Anorg. Chem.* **1964**, *328*, 113.
[81] R. Krishnamurthy, G. Arrhenius, A. Eschenmoser, *Origins Life Evol. Biosphere* **1999**, *29*, 333.
[82] R. Krishnamurthy, S. Guntha, A. Eschenmoser, *Angew. Chem.* **2000**, *112*, 2369; *Angew. Chem. Int. Ed.* **2000**, *39*, 2281.
[83] F. Westheimer, *Science* **1987**, *235*, 1173.
[84] A. Eschenmoser, *Science* **1999**, *284*, 2118.
[85] In this paper a molecular system is designated "informational" if it has the potential to occur in a variety of functionally equivalent constitutions and (in principle) is capable to direct the faithful replication of the constitutional variants.
[86] H.-J. Roth, "Synthese von Pyranose-Nucleosiden aus Pyrimidin-Basen", Diplomarbeit ETH Zürich, **1986**; J. Hunziker, "Synthese von Purin-Hexose-Nucleosiden", Diplomarbeit ETH Zürich, **1986**.
[87] The first observation on the pairing behavior of homo-DNA goes back to 1988 and consisted in the (surprising and, at the time, irritating) finding that the homobasic adenine hexamer by itself showed the same UV melting curve as its 1:1-mixture with the homobasic uracil hexamer (self-pairing of homo-dA_6: $T_m \approx 46$ °C, ca. 6 μM; 0.15 M NaCl, pH 7.0; UV 260 nm; hyperchromicity ca. 6%). Before their publication in 1992/93[90,98,99] the synthesis and pairing properties of homo-DNA oligonucleotides were reported in various lectures, for the first time at the IUPAC Congress on Natural Products in Kyoto (June 2, 1988; observation of the self-pairing of homo-dA_6) at the Bürgenstock Conference on Stereochemistry (May 4, 1989) and the fall meeting of the Swiss Chemical Society in Berne (Oct. 20, 1989). Three lectures delivered during the years 1990 and 1991 also appeared in print.[88]
[88] "Kon-Tiki-Experimente zur Frage nach dem Ursprung von Biomolekülen" A. Eschenmoser in *Materie und Prozesse. Vom Elementaren zum Komplexen* (Eds.: W. Gerok et al.), Wiss. Verlagsges., Stuttgart, **1991**, pp. 135, Verh. Ges. Dtsch. Naturforsch. Aerzte, 116. Vers. Berlin, **1990**; "Warum Pentose- und nicht Hexose-Nukleinsäuren?": A. Eschenmoser, *Nachr. Chem. Tech. Lab.* **1991**, *39*, 795 (the date of the lecture mentioned in the editor's preface is a misprint, and should read Nov. 9, 1990); "Chemie potentiell präbiologischer Naturstoffe": A. Eschenmoser, *Nova Acta Leopoldina* **1992**, *67*, 201 – 233 (English version, ref. [126]).
[89] This statement is based on the notion that, from an etiological point of view, the structure of a 2,3-dideoxyhexose, though outwardly simpler, is more "complex" than that of a corresponding natural hexose. 2,3-Dideoxyhexoses are not products of a (potentially prebiotic) formose reaction.
[90] A. Eschenmoser, M. Dobler, *Helv. Chim. Acta* **1992**, *75*, 218; see also the supplement in ref. [98], footnote 17 on p. 1448, as well as in ref. [99], Scheme 19 on p. 340.
[91] In this context, see also the theoretical models published at the outset of Herdewijn's involvement in the chemistry of the pyranose forms of nucleic acids: H. L. De Winter, N. M. Blaton, O. M. Peeters, C. J. De Ranter, A. Van Aerschot, P. Herdewijn, *Acta Crystallogr. Sect. C* **1991**, *47*, 835, 838, 1693, 2245.
[92] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 738.
[93] "Fifty Years with Double-Stranded RNA": A. Rich, *Scientist* **2006**, 34.
[94] S. K. Singh, P. Nielsen, A. A. Koshkin, J. Wengel, *Chem. Commun.* **1998**, 455; A. A. Koshkin, P. Nielsen, M. Meldgard, V. K. Rajwanshi, S. K. Singh, J. Wengel, *J. Am. Chem. Soc.* **1998**, *120*, 13252; vgl. auch S. Obika, D. Nanbu, Y. Hari, K. Morio, Y. In, T. Ishida, T. Imanishi, *Tetrahedron Lett.* **1997**, *38*,

8735; S. Obika, D. Nanbu, Y. Hari, J. Andoh, K. Morio, T. Doi, T. Imanishi, *Tetrahedron Lett.* **1998**, *39*, 5401.

[95] A. Van Aerschot, O. Verheggen, C. Hendrix, P. Herdewijn, *Angew. Chem.* **1995**, *107*, 1483; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1338; A. Augustyns, F. Vandendriessche, A. Van Aerschot, R. Busson, C. Urbanke, P. Herdewijn, *Nucleic Acids Res.* **1992**, *20*, 4711; R. Declercq, A. Van Aerschot, R. J. Read, P. Herdewijn, L. Van Meervelt, *J. Am. Chem. Soc.* **2002**, *124*, 928.

[96] G. Otting, M. Billeter, K. Wüthrich, H.-J. Roth, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 2701.

[97] M. Egli, P. S. Pallan, R. Pattanayek, C. J. Wilds, P. Lubini, G. Misanov, M. Dobler, C. Leumann, A. Eschenmoser, *J. Am. Chem. Soc.* **2006**, *128*, 10847.

[98] M. Böhringer, H.-J. Roth, J. Hunziker, M. Göbel, R. Krishnan, A. Giger, B. Schweizer, J. Schreiber, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1992**, *75*, 1416.

[99] J. Hunziker, H.-J. -Roth, M. Böhringer, A. Giger, U. Diederichsen, M. Göbel, R. Krishnan, B. Jaun, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 259.

[100] K. Groebke, J. Hunziker, W. Fraser, L. Peng, U. Diederichsen, K. Zimmermann, A. Holzner, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1998**, *81*, 375.

[101] Additional experimental data referring to synthesis and properties of homo-DNA oligonucleotides are to be found in the following theses (accessible in the internet under the respective authors' names, http://www.e-collection.ethbib.ethz.ch): H.-H. Roth, "Homo-DNS: Herstellung, Paarunseigenschaften und Struktur von Adenin/Thymin-haltigen Oligonukleotiden", Diss. ETH Zürich, Nr. 9591, **1991**; M. Böhringer, "Synthese und strukturelle Charakterisierung von Adenin-Oligonukleotiden mit 2′,3′-Dideoxy-D- und L-Gluco-Pyranosen als Zuckereinheiten", Diss. ETH Zürich, Nr. 9377, **1991**; J. Hunziker, "Homo-DNA: Herstellung und Paarungseigenschaften von Guanin-haltigen Oligonukleotiden", Diss. ETH Zürich, Nr. 9814, **1992**; Ling Peng, "1. Synthese und Eigenschaften von Oligo-2′-Deoxy-1′,2′,-seco-D-Ribonukleotiden. – 2. Purin-Purin-gepaarte Oligonukleotide der homo-DNA Reihe", Diss. ETH Zürich, Nr. 10035, **1993**. See also ref. [121] and [102].

[102] K. Groebke, "Ueber Purin-Purin-Paarungen bei Hexopyranose-Nukleisäuren", Diss. ETH Zürich, Nr. 10149, **1993**. For a recent publication dealing with Watson–Crick purine–purine pairing in the DNA series cf. B. D. Heuberger, C. Switzer, *ChemBioChem* **2008**, *9*, 2779.

[103] M. Tarköy, M. Bolli, B. Schweizer, C. Leumann, *Helv. Chim. Acta* **1993**, *76*, 481; M. Tarköy, C. Leumann, *Angew. Chem.* **1993**, *105*, 1516; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1432.

[104] C. Leumann, *Bioorg. Med. Chem.* **2002**, *10*, 841 (Review).

[105] In one of his reviews (ref. [107]), Herdewijn referred to "conformationally constrained oligonucleotides that were first studied by M. Ikehara" (M. Ikehara, S. Uesugi, T. Shida, *Chem. Pharm. Bull.* **1980**, *28*, 189). However, those studies did not relate to pre-organization of an oligonucleotide backbone.

[106] K. Augustyns, A. Van Aerschot, C. Urbanke, P. Herdewijn, *Bull. Soc. Chim. Belg.* **1992**, *101*, 119.

[107] P. Herdewijn, *Ann. Chem.* **1996**, 1337 (review).

[108] P. Herdewijn, *Chem. Biodiversity* **2010**, *7*, 1 (Review).

[109] M. Froeyen, E. Lescrenier, L. Kerremans, H. Rosemeyer, F. Seela, B. Verbeure, I. Lagoja, J. Rozenski, A. Van Aerschot, R. Busson, P. Herdewijn, *Chem. Eur. J.* **2001**, *7*, 5183.

[110] The nucleobase in the originally described "homo-DNA" is attached to the pyranose ring in the anomeric β-position. Herdewijn[109] renamed it as "β-homo-DNA" to distinguish it from its anomer, "α-homo-DNA".

[111] E. Lescrinier, M. Froeyen, P. Herdewijn, *Nucleic Acids Res.* **2003**, *31*, 2975 (review); K. Nauwelaerts, E. Lescrenier, P. Herdewijn, *Chem. Eur. J.* **2007**, *13*, 90.

[112] R. P. Hammer (see ref. [122]) in our laboratory had also prepared oligonucleotides of the α-homo-DNA series containing adenine and thymine bases; he observed base pairing much weaker than in the β-series. His observations are described in the thesis of A. Giger.[121]

[113] C. Leumann, *Chimia* **2001**, *55*, 295 (review).

[114] T. Imanishi, S. Obika, *Chem. Commun.* **2002**, 1653.

[115] J. Wengel, *Acc. Chem. Res.* **1999**, *32*, 301.

[116] C. R. Noe, M. Knollmüller, P. Ettmayer, *Liebigs Ann. Chem.* **1991**, 417; C. R. Noe, M. Knollmüller, P. Ettmayer, J. Freissmuth, *Liebigs Ann. Chem.* **1994**, 611.

[117] R. W. Fischer, A. G. Helg, K. Groebke, U. Diederichsen, A. Giger, R. Hammer, R. Krishnamurthy, C. Miculka, P. Seiler, B. Jaun, S. Pitsch, C. Leumann, A. Eschenmoser, unpublished.

[118] R. W. Fischer, "Allopyranosyl-Nukleinsäure: Synthese, Paarungseigenschaften und Struktur von Adenin/Uracil-haltigen Oligonukleotiden", Diss. ETH Zürich, Nr. 9971, **1992**.

[119] A. G. Helg, "Allopyranosyl-Nukleinsäure: Synthese, Paarungseigenschaften und Struktur von Guanin/Cytosin-enthaltenden Oligonukleotiden", Diss. ETH Zürich, Nr. 10464, **1994**.

[120] U. Diederichsen, "A. Hypoxanthin-Basenpaarungen in homo-DNA Oligonukleotiden. – B. Zur Frage des Paarungsverhaltens von Glucopyranosyl-Oligonukleotiden", Diss. ETH Zürich, Nr. 10122**, 1993**.

[121] A. Giger, "Untersuchungen über Oligonukleotide mit 2-Deoxy-D-ribose, 2,3-Dideox-D-Glucopyranose und D-Allopyranose als Zuckerbausteine", Diss. ETH Zürich, Nr. 9975, **1992**.

[122] Robert Hammer, Postdoktorat ETH Zürich, **1990–1992**.

[123] Ramanarayanan Krishnamurthy, Postdoktorat ETH Zürich, **1992–1994**.

[124] Christian Miculka, Postdoktorat, Institut für Pharmazie (Direktion Prof. Christian Noe), Johann-Wolfgang-Goethe Universität Frankfurt/M., **1993–1995**.

[125] Dissertations accessible as full texts (German) in the internet http://www.e-collection.ethbib.ethz.ch.

[126] "Chemistry of Potentially Prebiological Natural Products": A. Eschenmoser, E. Loewenthal, *Chem. Soc. Rev.* **1992**, *21*, 1; A. Eschenmoser, *Nova Acta Leopold.* **1992**, *67*, 201.

[127] "Hexose Nucleic Acids": A. Eschenmoser, *Pure Appl. Chem.* **1993**, *65*, 1179.

[128] Towards a Chemical Etiology of the Natural Nucleic Acids Structure": A. Eschenmoser in *Chemical Synthesis: Gnosis to Prognosis* (Eds.: C. Chatgilialoglu, V. Snieckus), Kluwer, Dordrecht, **1996**, 293 (*NATO ASI Ser. Ser. E, Vol. 320*).

[129] Experimental details on the synthesis of mono- and oligonucleotides of the allose series are to be found in: ref. [118], pp. 208–288, ref. [119], pp. 257–347, ref. [121], pp. 215–252; of the altrose series: ref. [102], pp. 351–394; of the glucose series: ref. [120], pp. 212–282.

[130] To deduce from the observed properties of allo-, altro-, and gluco-variants that hexopyranose oligonucleotide systems could not have served as genetic systems is a conclusion to be extrapolated only with caution to the entire family of hexopyranose alternatives. This proviso seems indicated as a consequence of observations made later in the series of pentopyranose oligonucleotide systems (cf. Section 4.6).

[131] "Too many notes": Emperor Josef II to Wolfgang Amadeus Mozart after attending the first performance of the opera "Die Hochzeit des Figaro"; stage play and movie "Amadeus" (Peter Shaffer 1979, Milos Forman 1984).

[132] H. Moser, A. Fliri, A. Steiger, G. Costello, J. Schreiber, A. Eschenmoser, *Helv. Chim. Acta* **1986**, *69*, 1224.

[133] S. Pitsch, S. Wendeborn, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1993**, *76*, 2161.

[134] S. Pitsch, R. Krishnamurthy, M. Bolli, S. Wendeborn, A. Holzner, M. Minton, C. Lesueur, I. Schlönvogt, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **1995**, *78*, 1621.

[135] R. Krishnamurthy, S. Pitsch, M. Minton, C. Miculka, N. Windhab, A. Eschenmoser, *Angew. Chem.* **1996**, *108*, 1619; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 1537.
[136] I. Schlönvogt, S. Pitsch, C. Lesueur, A. Eschenmoser, B. Jaun, R. M. Wolf, *Helv. Chim. Acta* **1996**, *79*, 1249.
[137] R. Micura, M. Bolli, N. Windhab, A. Eschenmoser, Angew. Chem. **1997**, *109, 899; Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 870; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 870.
[138] M. Bolli, R. Micura, A. Eschenmoser, *Chem. Biol.* **1997**, *4*, 309.
[139] M. Bolli, R. Micura, S. Pitsch, A. Eschenmoser, *Helv. Chim. Acta* **1997**, *80*, 1901.
[140] R. Micura, R. Kudick, S. Pitsch, A. Eschenmoser, *Angew. Chem.* **1999**, *111*, 715; *Angew. Chem. Int. Ed.* **1999**, *38*, 680.
[141] S. Pitsch, S. Wendeborn, R. Krishnamurthy, A. Holzner, M. Minton, M. Bolli, C. Miculka, N. Windhab, R. Micura, M. Stanek, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **2003**, *86*, 4270.
[142] S. Ilin, I. Schlönvogt, M.-O. Ebert, B. Jaun, H. Schwalbe, *ChemBioChem* **2002**, *3*, 93.
[143] R. Naylor, P. T. Gilham, *Biochemistry* **1966**, *5*, 2722.
[144] L. E. Orgel, *Nature* **1992**, *358*, 203; A. R. Hill, L. E. Orgel, T. Wu, *Origins Life Evol. Biosphere* **1993**, *23*, 285.
[145] The model depicted in Figure 45 originates from calculations by Dr. Romain Wolf, at that time with Ciba-Geigy AG, Basel.
[146] D. A. Usher, A. H. Hale, *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 1149.
[147] In sharp contrast to activation via the 2′,3′-cyclophosphate, no ligation whatsoever was observed under otherwise identical reaction condition when activation was attempted as the isomeric 3′,4′-cyclophosphate.[139]
[148] These are the diastereomers 4′-pr DDDD, 4′-pr-LDDD, 4′-pr-DLDD, 4′-pr-DDLD, 4′-pr-DDDL, 4′-pr-DDLL, 4′-pr-DLDL, 4′-pr-LDDL; cf. ref. [149].
[149] pr = "pyrano-ribo"; L = L-Chirality, D = D-Chirality of the ribopyranose unit.
[150] This corresponds to the "chiroselectivity gap" that had been observed in the ligations of tetramer-2′,3′-cyclophosphates on octamer templates.
[151] For an example of attempted characterization by HPL chromatography of a combinatorial oligomerization (4′-DTCG-cp + 4′-GCCG-cp, D = 2-amino-adenine) cf. Figure 5 in ref. [138]
[152] a) M. M. Green, B. A. Garetz, *Tetrahedron Lett.* **1984**, *25*, 2831; b) J. Siegel, *Chirality* **1998**, *10*, 24.
[153] M. Renz, R. Lohrman, L. E. Orgel, *Biochim. Biophys. Acta Nucleic Acids Protein Synth.* **1971**, *240*, 463.
[154] K. L. Agarwal, H. Büchi, M. H. Caruthers, N. Gupta, H. G. Khorana, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. Rajbhandary, J. H. Van de Sande, V. Sgaramella, H. Weber, T. Yamada, *Nature* **1970**, *227*, 27.
[155] G. Karig, A. Fuchs, A. Büsing, T. Brandstetter, S. Scherer, J. W. Bats, A. Eschenmoser, G. Quinkert, *Helv. Chim. Acta* **2000**, *83*, 1049; H. Schwalbe, J. Wermuth, C. Richter, S. Szalma, A. Eschenmoser, G. Quinkert, *Helv. Chim. Acta* **2000**, *83*, 1079.
[156] a) 2′-Deoxy-xylofuranose-NA ("Xylo-DNA"): H. Rosemeyer, M. Krecmerova, F- Seela, *Helv. Chim. Acta* **1991**, *74*, 2054; A. Schöppe, H.-J. Hinz, H. Rosemeyer, *Eur. J. Biochem.* **1996**, *239*, 33; F. Seela, M. Heckel, H. Rosemeyer, *Helv. Chim. Acta* **1996**, *79*, 1451); cf. also N. E. Poopeiko, M. Juhl, B. Vester, M. D. Sorensen, J. Wengel, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2285; A. Ramaswamy, M. Froeyen, P. Herdewijn, A. Ceulemans, *J. Am. Chem. Soc.* **2010**, *132*, 587; b) arabinofuranose-NA: P. A. Giannaris, M. J. Damha, *Can. J. Chem.* **1994**, *72*, 909; A. Y. Denisov, A. M. Noronha, D. J. Wilds, J.-F. Trempe, R. T. Pon, K. Gehring, M. J. Damha, *Nucleic Acids Res.* **2001**, *29*, 4284; c) (5′→2′)-RNA: D. A. Usher, A. H. McHale, *Proc. Natl. Acad. Sci. USA* **1976**, *73*, 1149; R. Kierzek, L. He, D. H. Turner, *Nucleic Acids Res.* **1992**, *20*, 1685; K.-E. Jung, C. Switzer, *J. Am. Chem. Soc.* **1994**, *116*, 6059; B. J. Premraj, S. Raja, N. S. Bhavesh, K. Shi, R. V. Hosur, M. Sundaralingam, N. Ybathindra, *Eur. J. Biochem.* **2004**, *271*, 2956; d) α-anomer of RNA: F. Debart, G. Tosquellas, B. Rayner, J.-L. Imbach, *Bioorg. Med. Chem. Lett.* **1994**, *4*, 1041; M. J. Damha, P. A. Giannaris, P. Marfey, *Biochemistry* **1994**, *33*, 7877; e) (5′→2′)-DNA: J. P. Dougherty, C. J. Rizzo, R. C. Breslow, *J. Am. Chem. Soc.* **1992**, *114*, 6254; T. L. Sheppard, R. C. Breslow, *J. Am. Chem. Soc.* **1996**, *118*, 9810.
[157] M. Beier, F. Reck, T. Wagner, R. Krishnamurthy, A. Eschenmoser, *Science* **1999**, *283*, 699.
[158] M.-O. Ebert, A. Luther, H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, B. Jaun, *Helv. Chim. Acta* **2002**, *85*, 4055.
[159] O. Jungmann, M. Beier, A. Luther, H. K. Huynh, M.-O. Ebert, B. Jaun, R. Krishnamurthy, A. Eschenmoser, *Helv. Chim. Acta* **2003**, *86*, 1259.
[160] O. Jungmann, H. Wippo, M. Stanek, H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **1999**, *1*, 1527. The relatively strong self-pairing of the non-self-complementary sequences 4′-TATTTTAA and 4′-TTAAAATA observed in all four pentapyranosyl systems shown in Figure 57 is presumably the result of a shift of the reading frame: an "upstream frame-shift" by one unit in both sequences leads to four stabilizing A/T-pairs, three destabilizing T/T- and three A/A-mismatches, as well as one stabilizing base-stacking of a dangling A- and T-base, respectively, at the 2′-ends.
[161] H. Wippo, F. Reck, R. Kudick, M. Ramaseshan, G. Ceulemens, M. Bolli, R. Krishnamurthy, A. Eschenmoser, *Bioorg. Med. Chem.* **2001**, *9*, 2411.
[162] F. Reck, H. Wippo, R. Kudick, M. Bolli, G. Ceulemenas, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **1999**, *1*, 1531.
[163] F. Reck, H. Wippo, R. Kudick, R. Krishnamurthy, A. Eschenmoser, *Helv. Chim. Acta* **2001**, *84*, 1778.
[164] T. Wagner, H. K. Huynh, R. Krishnamurthy, A. Eschenmoser, *Helv. Chim. Acta* **2002**, *85*, 399.
[165] K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347.
[166] K.-U. Schöning, P. Scholz, X. Wu, S. Guntha, G. Delgado, R. Kirshnamurthy, A. Eschenmoser, *Helv. Chim. Acta* **2002**, *85*, 4111.
[167] A. Eschenmoser, *Origins Life Evol. Biosphere* **2004**, *34*, 277.
[168] M.-O. Ebert, C. Mang, R. Krishnamurthy, A. Eschenmoser, B. Jaun, *J. Am. Chem. Soc.* **2008**, *130*, 15105.
[169] M.-O. Ebert, B. Jaun, *Chem. Biodiversity* **2010**, *7*, 2103.
[170] C. J. Wilds, Z. Wawrzak, R. Krishnamurthy, A. Eschenmoser, M. Egli, *J. Am. Chem. Soc.* **2002**, *124*, 13716; P. S. Pallan, C. J. Wilds, Z. Wawrzkak, R. Krishnamurthy, A. Eschenmoser, M. Egli, *Angew. Chem.* **2003**, *115*, 6073; *Angew. Chem. Int. Ed.* **2003**, *42*, 5893.
[171] The base-pairing strength in cross-pairing between sequences of the (4′→3′)-lyxopyranosyl series and antiparallel-complementary sequences of DNA or RNA shows—in contrast to the cross-pairing of the latter with TNA—pronounced sequence dependence (cf. ref. [161]).
[172] L. E. Orgel, *Science* **2000**, *290*, 1306; P. Herdewijn, *Angew. Chem.* **2001**, *113*, 2309; *Angew. Chem. Int. Ed.* **2001**, *40*, 2249.
[173] J. G. Schmidt, P. E. Nielsen, L. E. Orgel, *Nucleic Acids Res.* **1997**, *25*, 4797; I. A. Kozlov, B. de Bouevere, A. Aeschot, P. Herdewijn, L. E. Orgel, *J. Am. Chem. Soc.* **1999**, *121*, 5856.
[174] A. Eschenmoser, *Origins Life Evol. Biosphere* **1997**, *27*, 535.
[175] X. Wu, G. Delgado, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **2002**, *4*, 1283.
[176] X. Wu, S. Guntha, M. Ferencic, R. Krishnamurthy, A. Eschenmoser, *Org. Lett.* **2002**, *4*, 1279.
[177] F. B. Howard, J. Frazier, H. T. Miles, *J. Biol. Chem.* **1966**, *241*, 4293; C. Hartel, M. W. Göbel, *Helv. Chim. Acta* **2000**, *83*, 2541.

[178] R. Lohrman, L. E. Orgel, *Nature* **1976**, *261*, 342; W. S. Zielinski, L. E. Orgel, *Nature* **1987**, *327*, 346.
[179] G. von Kiedrowski, B. Wlotzka, J. Helbing, M. Matzen, S. Jordan, *Angew. Chem.* **1991**, *103*, 456; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 423.
[180] N. G. Dolinnaya, N. I. Sokolova, D. T. Ashirbekova, D. T. Shabarova, *Nucleic Acids Res.* **1991**, *19*, 3067.
[181] J. C. Chaput, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 9274.
[182] J. C. Chaput, J. K. Ichida, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 856.
[183] J. K. Ichida, K. Zou, A. Horhota, B. Yu, L. W. McLaughlin, J.-W. Szostak, *J. Am. Chem. Soc.* **2005**, *127*, 2802.
[184] V. Kempeneers, K. Vastmans, J. Rozenski, P. Herdewijn, *Nucleic Acids Res.* **2003**, *31*, 6221.
[185] B. D. Heuberger, C. Switzer, *Org. Lett.* **2006**, *8*, 5809.
[186] L. Zhang, A. Peritz, E. Meggers, *J. Am. Chem. Soc.* **2005**, *127*, 4175.
[187] E. Meggers, L. Zhang, *Acc. Chem. Res.* **2010**, *43*, 1092.
[188] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497.
[189] P. E. Nielsen, *Acc. Chem. Res.* **1999**, *32*, 624.
[190] P. E. Nielsen, *Origins Life Evol. Biosphere* **1993**, *23*, 323.
[191] B. Hyrup, P. E. Nielsen, *Bioorg. Med. Chem.* **1996**, *4*, 5.
[192] For literature on earlier attempts to synthesize base-pairing systems containing acyclic backbones see the reviews of Nielsen and Meggers.[187] For pairing properties of acyclic phosphodiester oligomers derived from glycerol derivatives see for example, J. C. Chaput, C. Switzer, *J. Mol. Evol.* **2000**, *51*, 464; and for recently reported nucleic acid analogues derived from 2-amino-1,3-dihydroxymonophosphate see H. Kashida, K. Murayama, T. Toda, H. Asanuma, *Angew. Chem.* **2011**, *123*, 1321 – 1324; *Angew. Chem. Int. Ed.* **2011**, *50*, 1285 – 1288.
[193] The following incident exemplifies the assessment: At the Bürgenstock Conference (May 1989) the author mentioned in his lecture that homo-DNA oligonucleotides show stronger pairing than corresponding DNA oligonucleotides. In a subsequent discussion a biochemist expressed serious doubts about the correctness of the interpretation of the experimental observations. "It is inconceivable" so it was argued, that the natural system DNA should not be "the best" of all possible pairing systems.
[194] Cf. for example, A. F. Amaral, M. M. Marques, J. A. L. da Silva, J. J. R. Frausto da Silva, *New J. Chem.* **2008**, *32*, 2043; roughly two thirds of the 2,3-cyclophosphate of ribose exists as furanose derivative (S. Pitsch, C. Spinner, K. Atsumi, P. Ermert, *Chimia* **1999**, *53*, 291); a propos "why ribose?" cf. also for example, R. Bielski, M. Tencer, *Origins Life Evol. Biosphere* **2007**, *37*, 167
[195] For a discussion in an etiological context of thermodynamic data of RNA's pairing structure, see D. H. Turner, P. C. Bevilacqua, "Thermodynamic Considerations for Evolution by RNA": D. H. Turner, P. C. Bevilacqua in *The RNA World* (Eds.: R. F. Gesteland, J. F. Atkins), Cold Spring Harbor Laboratory Press, Cold Spring, **1993**, chap. 17, p. 447.
[196] W. Gilbert, *Nature* **1986**, *319*, 618.
[197] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147; C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849.
[198] a) C. Woese, *The Genetic Code, the Molekular Basis for Genetic Expression*, Harper and Row, New York, **1967**; F. H. C. Crick, *J. Mol. Biol.* **1968**, *38*, 367; L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381.
[199] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99.
[200] T. A. Steitz, P. B. Moore, *Trends Biochem. Sci.* **2003**, *28*, 411.
[201] T. A. Lincoln, G. F. Joyce, *Science* **2009**, *323*, 1229; see also A. Wochner, J. Attwater, A. Coulson, P. Holliger, *Science* **2011**, *332*, 209.
[202] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818.
[203] A. A. Beaudry, G. F. Joyce, *Science* **1992**, *257*, 625.
[204] S. Fusz, A. Eisenführ, S. G. Srivatsan, A. Heckel, M. Famulok, *Chem. Biol.* **2005**, *12*, 941.
[205] G. F. Joyce, *Angew. Chem.* **2007**, *119*, 6540; *Angew. Chem. Int. Ed.* **2007**, *46*, 6420 (review).
[206] A. G. Cairns-Smith, *Genetic Takeover and the Mineral Origins of Life*, Cambridge University Press, Cambridge, **1982**.
[207] G. J. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 4398.
[208] "Was RNA the first Genetic Polymer?": L. E. Orgel in *Evolutionary Tinkering in Gene Expression* (Eds.: M. Grunberg-Manago, B. F. C. Clark, H. G. Zachau), Plenum, New York, **1989**.
[209] In distant analogy to developments that may have led to a succession from the RNA world to an RNA protein world.
[210] M. Koppitz, P. E. Nielsen, L. E. Orgel, *J. Am. Chem. Soc.* **1998**, *120*, 4563.
[211] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol.* **1972**, *1*, 249.
[212] R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1970**, *47*, 531.
[213] R. Lohrmann, L. E. Orgel, *Science* **1968**, *161*, 65; R. Lohrmann, L. E. Orgel, *Nature* **1973**, *244*, 418.
[214] G. von Kiedrowski, *Angew. Chem.* **1986**, *98*, 932; *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 932; D. Sievers, G. von Kiedrowski, *Nature* **1994**, *369*, 221.
[215] L. E. Orgel, *Nature* **1992**, *358*, 203.
[216] V. M. Kolb, J. P. Dworkin, S. L. Miller, *J. Mol. Evol.* **1994**, *38*, 549.
[217] M. Levy, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7933.
[218] J. P. Ferris, G. Ertem, *J. Am. Chem. Soc.* **1993**, *115*, 12270; G. Ertem, J. P. Ferris, *Nature* **1996**, *379*, 238; G. Ertem, J. P. Ferris, *J. Am. Chem. Soc.* **1997**, *119*, 7197; J. P. Ferris, A. R. Hill, R. Liu, L. E. Orgel, *Nature* **1996**, *381*, 59.
[219] W. Huang, J. P. Ferris, *Chem. Commun.* **2003**, 1458.
[220] C. Anastasi, F. F. Buchet, M. A. Crowe, A. L. Parkes, M. W. Powner, J. M. Smith, J. D. Sutherland, *Chem. Biodiversity* **2007**, *4*, 721.
[221] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239; A. Choudhary, K. J. Kamer, M. W. Powner, J. D. Sutherland, R. T. Raines, *ACS Chem. Biol.* **2010**, *5*, 655; M. W. Powner, J. D. Sutherland, J. W. Szostak, *J. Am. Chem. Soc.* **2010**, *132*, 16677.
[222] R. Shapiro, *Origins—A Skeptic's Guide to the Creation of Life on Earth*, Summit Books, New York, **1986**.
[223] S. Zahng, C. Switzer, J. C. Chaput, *Chem. Biodiversity* **2010**, *7*, 24.
[224] J. J. Chen, X. Cai, J. W. Szostak, *J. Am. Chem. Soc.* **2009**, *131*, 2119; A. T. Horhota, J. W. Szostak, L. W. McLaughlin, *Org. Lett.* **2006**, *8*, 5345; Y.-W. Yang, S. Zhang, E. O. McCollum, J.-C. Chaput, *J. Mol. Evol.* **2007**, *65*, 289.
[225] Addition, substitution, and elimination reactions leading to, and proceeding at, multiple bonds are the major carbon skeleton-forming reactions in prebiotic chemistry; S_N2 reactions seem not to among them, at least according to present thinking. Even in biology, enzymatically catalyzed S_N2 reactions at saturated carbon (with the exception of methyl or allylic carbon) are conspicuously rare, c.f., e.g., ref. [226].
[226] D. O'Hagen, J. W. Schmidberger, *Nat. Prod. Rep.* **2010**, *27*, 900.
[227] In the laboratory of Orgel[208] it was shown that activation of the 6′-phosphate of 2′-deoxyglucopyranosyl-adenine leads to cyclization much more easily than activation of the 5′-phosphate of 2′-deoxyadenosine. The *trans*-ring junction in the two cyclization products is supposed to be responsible for this difference: the *trans*-junction of a 5- with a 6-membered ring is strained, in contrast to the *trans*-junction between two 6-membered rings.
[228] P.-C. Zamecnik, M. L. Stephenson, *Proc. Natl. Acad. Sci. USA* **1978**, *75*, 280.

[229] F. Kuschel, H. Ringsdorf, *Nachr. Chem.* **2010**, *58*, 759.
[230] S. Hoffmann, *Z. Chem.* **1979**, *19*, 241 – 249.
[231] "Synthetic Analogues of Nucleic acids—a Review": A. S. Jones, *Int. J. Biol. Macromol.* **1979**, *1*, 194.
[232] J. D. Buttrey, A. S. Jones, R. T. Walker, *Tetrahedron* **1975**, *31*, 73.
[233] In the chapter "Polynucleotide Models" of ref. [230], Siegfried Hoffmann distinguishes between "near and distant analogations". The former is meant to refer to chemical syntheses of oligonucleotide sequences, whereas the latter attempts to achieve a simulation of the natural nucleic acid's base-pairing through attaching nucleobases to polymer backbones familiar in the polymer chemistry of that time, or by converting monomeric units bearing a nucleobase into corresponding oligomers by polymerisations or polycondensations (cf. ref. [232] for early work in this field, and ref. [230], p. 246 for a survey of such attempts). Hoffmann's comment referring to the "distant analogations": "*Der Bogen spannt sich zurück zu fernen Struktur- und Funktions-Simulationen, die mit all ihren Unzulänglichkeiten und stereoelektronischen Unvollkommenheiten doch den Versuch des gegenwärtigen Finalprodukts der Evolution darstellen, sechs Milliarden Jahre in der Spanne einer Generation abstrahierend nachzuvollziehen*".
[234] H. De Koning, U. K. Pandit, *Recl. Trav. Chim. Pays-Bas* **1971**, *90*, 1069.
[235] E. P. Stirchak, J. E. Summerton, D. D. Weller, *Nucleic Acids Res.* **1989**, *17*, 6129; J. Summerton, D. Weller, *Antisense Nucleic Acids Drug Dev.* **1997**, *7*, 187.
[236] P. E. Nielsen in *Pseudo-peptides in Drug Discovery* (Ed.: P. E. Nielsen), Wiley-VCH, Weinheim, **2004**, p. 153.
[237] V. A. Kumar, K. N. Ganesh, *Acc. Chem. Res.* **2005**, *38*, 404.
[238] Y. Huang, S. Dey, X. Zhang, F. Sönnichsen, P. Garner, *J. Am. Chem. Soc.* **2004**, *126*, 4626.
[239] For references of such work see ref. [252] (in there ref. [54]). See also footnote 67 in ref. [252], in which the author indulges in reminiscences of his early work on replacing the phosphodiester backbone of DNA by a peptide backbone.
[240] U. Diederichsen, *Angew. Chem.* **1996**, *108*, 458; *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 445; U. Diederichsen, *Angew. Chem.* **1997**, *109*, 1966; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1886; M. F. H. Hoffmann, A. M. Brückner, T. Hupp, B. Engels, U. Diederichsen, *Helv. Chim. Acta* **2000**, *83*, 2580.
[241] U. Diederichsen, H. W. Schmitt, *Angew. Chem.* **1998**, *110*, 312; *Angew. Chem. Int. Ed.* **1998**, *37*, 302; A. M. Brückner, H. W. Schmitt, U. Diederichsen, *Helv. Chim. Acta* **2002**, *85*, 3855; A. Weiss, U. Diederichsen, *Eur. J. Org. Chem.* **2007**, 5531.
[242] K. E. Nelson, M. Levy, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3868.
[243] Prebiotic building blocks of nucleo-oligodipeptides would have been α-amino acids and canonical or non-canonical nucleobases. The problems with nucleo-oligodipeptides are such referring to structural assembly, not to availability of building blocks. For PNA, both would be problematic, especially the availability of hydrazine as building block (cf. the comments in the preceding section on the question of a prebiotic formation of the RNA structure).
[244] For recent investigations on this question see L. J. Leman, L. E. Orgel, M. R. Ghadiri, *Science* **2004**, *306*, 283; L. J. Leman, L. E. Orgel, M. R. Ghadiri, *J. Am. Chem. Soc.* **2006**, *128*, 20.
[245] H. Moser, A. Fliri, A. Steiger, G. Costello, J. Schreiber, A. Eschenmoser, *Helv. Chim. Acta* **1986**, *69*, 1224.
[246] A. Eschenmoser, *Chimia* **2005**, *59*, 836.
[247] This expectation is based on the presumption that the *gauche* effect will stabilize the pairing conformation of the oligomer chain. Support for this is provided by an observation according to which a PNA-analog lacking the carbonyl of the amide group within the backbone chain still shows pairing (M. C. Myers, J. K. Pokorski, D. H. Appella, *Org. Lett.* **2004**, *6*, 4699), as do analogs of the structure type depicted in Figure 79, in which the amide group within the backbone chain is replaced by (CH_2-O) (K. H. Altmann, C. Schmit Chiesi, G. Garcia-Escheverria, *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1119; C. García-Escheverría, D. Hüsken, C. Schmit Chiesi, K. H. Altmann, *Bioorg. Med. Chem. Lett.* **1997**, *7*, 1123; M. Kuwahara, M. Arimitsu, M. Sisido, *J. Am. Chem. Soc.* **1999**, *121*, 256).
[248] G. Zubay, T. Mui, *Origins Life Evol. Biosphere* **2001**, *31*, 87.
[249] For earlier studies on alternative nucleobases that were not, however, carried out in an etiological context, see, e.g., J. A. Piccirilli, T. Krauch, S. A. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33; F. Seela, R. Kröschel, *Org. Biomol. Chem.* **2003**, *1*, 3900; E. T. Kool, *Acc. Chem. Res.* **2002**, *35*, 936.
[250] S. Miyakawa, H. Cleaves, S. L. Miller, *Origins Life Evol. Biosphere* **2002**, *32*, 209.
[251] 5-Aminouracil and the HCN tetramer are at the same "oxidation level" (formation of the former from the latter requiring neither an oxidation nor a reduction step). In contrast, formation of uracil from the HCN-tetramer requires a reduction step (for a more detailed discussion see ref. [252], p. 12836).
[252] A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821.
[253] G. K. Mittapalli, K. R. Reddy, H. Xiong, O. Munoz, B. Han, F. De Riccardis, R. Krishnamurthy, A. Eschenmoser, *Angew. Chem.* **2007**, *119*, 2522; *Angew. Chem. Int. Ed.* **2007**, *46*, 2470.
[254] G. K. Mittapalli, Y. M. Osornio, M. A. Guerrero, K. R. Reddy, R. Krishnamurthy, A. Eschenmoser, *Angew. Chem.* **2007**, *119*, 2530; *Angew. Chem. Int. Ed.* **2007**, *46*, 2478.
[255] Whether the two formally complementary 2-amino-6-oxo and 2-oxo-6-amino isomers of the 5-aminopyrimidine family (Figure 74) would pair with each other was tested in an exploratory experiment in the (Glu-Asp)-nucleo-oligodipeptide series using a dodecamer sequence that contained the two bases in alternating fashion. The self-complementary sequence showed moderately strong pairing, which was markedly dependent, however, on time. The cooling curve (UV) also showed extensive hysteresis (see Figure 3 in ref. [254]). This is interpreted to indicate that the pairing process requires a tautomerization in at least one of the bases.
[256] X. Zhang, R. Krishnamurthy, *Angew. Chem.* **2009**, *121*, 8268; *Angew. Chem. Int. Ed.* **2009**, *48*, 8124; R. Krishnamurthy, *J. Mex. Chem. Soc.* **2009**, *53*(1), 23.
[257] Theoretical calculations recently published by Sponer et al.[258] failed to reveal a correlation between pairing strength and pK_a difference between the pairing partners. However, the relationship between pK_a values and hydration of base partners in non-paired versus paired state—a relationship on which our qualitative interpretation of the observed correlation is based—has not been taken into account in those calculations.
[258] J. E. Šponer, A. Vásquez-Mayagoitia, B. G. Sumpter, J. Leszczynski, J. Šponer, M. Otyepka, P. Banáš, M. Fuentes-Cabrera, *Chem. Eur. J.* **2010**, *16*, 3057.
[259] H.-R. Marcos, M. Hernandez-Rodriguez, J. Xie, Y. M. Osornio, R. Krishnamurthy, *Chem. Asian J.* **2011**, *6*, 1252.
[260] R. Larralde, M. P. Robertson, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 8158.
[261] S. L. Miller, A. Lazcano in *Life's Origin: The Beginnings of Biological Evolution* (Ed.: W. Schopf), University of California Press, Berkeley, **2002**, pp. 78 – 112.
[262] *Marine Hydrothermal Systems and the Origin of Life* (Ed.: N. G. Holm), Kluwer, Dordrecht, **1992**.
[263] S. L. Miller, A. Lazcano, *J. Mol. Evol.* **1995**, *41*, 689; M. Levy, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 7933; J. L. Bada, A. Lazcano, *Science* **2002**, *296*, 1982; cf. also ref. [250].

[264] See in this context John Cornforth's notion on the role of hydrolysis in the origin of life (J. Cornforth, *Interdiscip. Sci. Rev.* **1984**, *9*, 107).

[265] For a proposal of such a scenario see footnote 26 in ref. [165].

[266] For the concept of a "dynamic combinatorial library" cf.: P. A. Brady, R. P. Bonar Law, S. I. Rowan, C. J. Suckling, J. Brady, K. M. Sanders, *Chem. Commun.* **1996**, 319; I. Huc, J.-M. Lehn, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2106; cf. also C. Miculka, N. Windhab, G. Quinkert, A. Eschenmoser, "Neue Substanzbibliothek und daraus hergestellte supramolekulare Komplexe", Patent DE 19619373A 1; Anmeldetag 14.5.1996; Offenlegungstag 20.11.1997.

[267] G. K. Mittapalli, R. Krishnamurthy, A. Eschenmoser, unpublished.

[268] Y. Ura, J. M. Beierle, L. J. Leman, L. S. Orgel, M. R. Ghadiri, *Science* **2009**, *325*, 73. In this context see also the work of *David Lynn* (Z.-Y. J. Zhan, D. G. Lynn, *J. Am. Chem. Soc.* **1997**, *119*, 12420; X. Li, Z.-Y. J. Zhan, R. Knipe, D. G. Lynn, *J. Am. Chem. Soc.* **2002**, *124*, 746; X. Li, A. F. Hernandez, M. A. Grover, N. V. Hud, D. G. Lynn, *Heterocycles* **2011**, *82*, 1477) and David Liu (D. M. Rosenbaum, D. R. Liu, *J. Am. Chem. Soc.* **2003**, *125*, 13924).

[269] a) H. J. Morowitz, *Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis*, Yale University Press, New Haven, **1992**; b) H. J. Morowitz, J. D. Kostelnik, J. Yang, G. D. Cody, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7704; c) E. Smith, H. J. Morowitz, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13168; see in this context Orgel's critical remarks: L. E. Orgel, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12503; L. E. Orgel, *PLOS Biol.* **2008**, *6*, 5.

[270] S. A. Kaufmann, *The Origins of Order. Self-Organization and Selection in Evolution*, Oxford University Press, Oxford, **1993**; cf. also ref. [14f].

[271] It was pointed out in this context that two types of catalysis should be distinguished: "template catalysis" versus "network catalysis",[269c] an alternative proposal being "reaction catalysis" versus "circuit catalysis".[272] Template control of the formation of a complementary oligomer strand (with rapid strand separation of the intermediate duplex) is a special case of catalysis of the first type; the property of each constituent of an autocatalytic cycle to act as a catalyst for each reactions step within the cycle would be referred as catalysis of the second type. However, analyzed on a reaction-topological level, the two apparently so disparate types of catalysis become identical.

[272] A. Eschenmoser, *Chem. Biodiversity* **2007**, *4*, 554.

[273] "Life is a self-sustained chemical system capable of undergoing Darwinian Evolution"; from the definition employed by NASA; cf. G. J. Joyce in the preface to *Origins of Life* (Hrsg.: E. W. Deamer, G. R. Fleischacker), Jones and Bartlett, Boston, **1994**.

[274] On "Replicators in Chemistry and Biology" see E. Szathmary, *Phil. Trans. R. Soc. London B* **2000**, *355*, 1669. A strategy for searching experimentally for autocatalytic reactions cycles (including potentially primordial ones) would be: target an organocatalyst of known catalytic capacity and design for the structure of this organocatalyst a "one-pot synthesis" in such a way that at least one of the reaction steps would be expected to be catalyzed by that organocatalyst. If the latter was supposed to be a primordial catalyst (e.g., one of the "archaic" cofactors), the designed synthesis would need to operate within the constraints of prebiotic chemistry.

[275] K. Soai, T. Shibata, H. Marioka, K. Choji, *Nature* **1995**, *378*, 767; T. Kawasaki, K. Suzuki, Y. Hakoda, K. Soai, *Angew. Chem.* **2008**, *120*, 506; *Angew. Chem. Int. Ed.* **2008**, *47*, 496; T. Kawasaki, T. Tutsumi, K. Suzuki, M. Ito, K. Soai, *Science* **2009**, *324*, 492.

[276] B. List, R. A. Lerner, C. F. Barbas, *J. Am. Chem. Soc.* **2000**, *122*, 2395; B. List, *Tetrahedron* **2002**, *58*, 5573; B. List, *Chem. Rev.* **2007**, *107*, 5413; C. F. Barbas, *Angew. Chem.* **2008**, *120*, 44; *Angew. Chem. Int. Ed.* **2008**, *47*, 42; E. N. Jacobsen, D. W. C. MacMillan, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20618; R. R. Knowles, E. N. Jacobsen, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 20678; A. Berkessel, H. Gröger, *Asymmetric Organocatalysis—from Biomimetic Concepts to Applications in Asymmetric Synthesis*, Wiley-VCH, Weinheim, **2005**.

[277] G. D. Cody, N. Z. Boctor, R. M. Hazen, J. A. Brandes, H. J. Morowitz, H. S. Yoder, Jr., *Geochim. Cosmochim. Acta* **2001**, *65*, 3557; X. V. Zhang, S. T. Martin, *J. Am. Chem. Soc.* **2006**, *128*, 16032.

[278] H. Schmalfuss, M. Congehl, *Biochem. Z.* **1927**, *185*, 70; W. Langenbeck, *Naturwissenschaften* **1942**, *30*, 30; R. Breslow, *Tetrahedron Lett.* **1959**, *21*, 22; H. J. Kim, A. Ricardo, H. I. Illangkoon, M. J. Kim, M. A. Carrigan, F. Frye, S. A. Benner, *J. Am. Chem. Soc.* **2011**, DOI: 10.1021/ja201769f.

[279] a) A. L. Weber, *Origins Life Evol. Biosphere* **1998**, *28*, 259; b) A. Weber, *Origins Life Evol. Biosphere* **2001**, *31*, 71; c) A. L. Weber, *Origins Life Evol. Biosphere* **2004**, *34*, 473; d) A. Weber, *Origins Life Evol. Biosphere* **2007**, *37*, 195.

[280] See also the theoretical and experimental studies on the formose process by Peter Decker in the context of his "Formol-Bioid" concept (ref. [281] and [63]).

[281] P. Decker, A. Speidel, *Z. Naturforsch. B* **1972**, *27*, 257; P. Decker, H. Schweer, *Origins Life Evol. Biosphere* **1984**, *14*, 335.

[282] Glyoxylic acid and dihydroxyfumaric acid are hydrated forms of potential oligomers of carbon monoxide.

[283] V. N. Sagi, F. Hu, P. Karri, V. Punna, G. Meher, R. Krishnamurthy; unpublished work at TSRI.

[284] S. W. Fox, *Science* **1960**, *132*, 200; A. Weber, S. L. Miller, *J. Mol. Evol.* **1981**, *17*, 273; F. Dyson, *Origins of Life*, *2nd ed.*, Cambridge University Press, Cambridge, **1999**; R. Pascal, L. Boiteau, A. Commeyras, *Top. Curr. Chem.* **2005**, *259*, 69.

[285] D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin, M. R. Ghadiri, *Nature* **1996**, *382*, 525; vgl auch S. Yao, S. Gosh, R. Zutshi, J. Chmielewski, *Nature* **1998**, *396*, 447.

[286] Z. G. Hajosh, D. R. Parish, *J. Org. Chem.* **1974**, *39*, 1615; U. Eder, G. Sauer, R. Wiechert, *Angew. Chem.* **1971**, *83*, 492; *Angew. Chem. Int. Ed. Engl.* **1971**, *10*, 496; E. R. Jarvo, S. J. Miller, *Tetrahedron* **2002**, *58*, 2481; S. Miller, *Acc. Chem. Res.* **2004**, *37*, 601; P. Krattiger, R. Kovasy, J. D. Revell, S. Ivan, H. Wennemers, *Org. Lett.* **2005**, *7*, 1101; M. Raj, V. K. Singh, *J. Org. Chem.* **2009**, *74*, 4289.

[287] W. A. Bonner, *Origins Life Evol. Biosphere* **1991**, *21*, 59; B. L. Feringa, R. A. van Delden, *Angew. Chem.* **1999**, *111*, 3624; *Angew. Chem. Int. Ed.* **1999**, *38*, 3418; S. Pizzarello, M. Lahav, *Origins Life Evol. Biosphere* **2010**, *40*, 1 (Special Issue: Emergence of Biochemical Homochirality); T. Kawasaki, K. Suzuki, Y. Hakoda, K. Soai, *Angew. Chem.* **2008**, *120*, 506; *Angew. Chem. Int. Ed.* **2008**, *47*, 496; R. Breslow, Z.-L. Cheng, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5723.

[288] The (chemical) question to be asked on the origin of biological homochirality (why are families of biomolecules homochiral?) is to be distinguished from the question (referring to physics) on the origin of the *sense* of the biomolecules's chirality. (cf. S. Mason, *Chem. Soc. Rev.* **1988**, *17*, 347).

[289] Behind this reasoning is the expectation that, in principle, each autocatalytically replicating system has the potential to "purify itself" to homochirality by virtue of intrinsic enantio- and diastereoselectivity of its constitutive reaction steps.

[290] Apart from the discussion that relates to Figure 52.

[291] P. L. Luisi, *Origins Life Evol. Biosphere* **1998**, *28*, 613; J. W. Szostak, D. P. Bartel, P. L. Luisi, *Nature* **2001**, *409*, 387; D. W. Deamer, *Microbiol. Mol. Biol. Rev.* **1997**, *61*, 230; C. R. Woese, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8742.

[292] The quest envisages matter that would have contained permetallated derivatives of carbon/oxygen- and carbon/nitrogen-containing organics, in which positions normally bearing a hydrogen atom would have been substituted by metal atoms. Eventual contact with water could thus have liberated reactive organic molecules such as HCN and cyanamide, as well as molecules corresponding to hydrated (formal) oligomers of CO and HCN, among others. Known chemical processes that might be relevant to point to in this context are, for instance, the formation of calcium cyanamide ($CaCN_2$) from calcium carbide and molecular nitrogen at temperatures over 1000 °C, or the Lassaigne test that the older generation of chemists may still remember (fusion of organic material with metallic sodium, plunging everything into water, and then testing for cyanide). The question of whether the Earth's crust could have been a source of organics relevant to the origin of life is addressed to the earth sciences, and only secondarily to chemistry. If such a scenario could not be positively excluded, the question posed would be of considerable etiological interest. The nearest counterargument, namely that no geochemical traces of such "reservoirs of organic material" are found today, would certainly not—after billions of years of a "hydrolyzing" earth history—be a valid reason for negating the question. (for previous comments on this subject cf. Figure 36 in *Origins Life Evol. Biosphere* **2004**, *34*, 277; and ref. [252]).

[293] The statement refers to the condensation of 6,7-dimethyl-8-D-ribityl-lumazine with itself to produce riboflavin with concomitant re-formation of one equivalent of 5-amino-6-D-ribityl-aminouracil, the molecule that happens to be the precursor of that very lumazine. This uracil derivative produces the lumazine by condensation with either 2,3-dioxobutane or 3,4-dihydroxy-2-butanone-4-phosphate, or even ribulose-1,5-diphosphate. The uracil derivative, the lumazine derivative and the dihydroxy-2-butanone derivative are intermediates in the biosynthesis of riboflavin; the two overall reaction steps that connect them have been shown also to occur non-enzymatically, the riboflavin-producing step (simple heating of the lumazine derivative in aqueous solution) displaying marvelously complex chemistry [cf. the work of H. C. S. Wood (T. Patterson, H. C. S. Wood, *J. Chem. Soc. Chem. Commun.* **1969**, 290; R. L. Beach, G. W. E. Plaut, *Biochemistry* **1970**, *9*, 760), of *A. Bacher* (K. Kis, K. Kugelbrey, A. Bacher, *J. Org. Chem.* **2001**, *66*, 2555), and the dissertation of C. Strupp (thesis. ETH, Nr. 9832, **1992**); for structural formulas see Figure 9 of ref. [126].

[294] Whereas the biosynthetic pathway to riboflavin is—while mechanistically complex—clearly "chemomimetic" and, therefore, ultimately "simple", the pathways leading to thiamin uncovered thus far (see, e.g., I. D. Spenser, R. L. White, *Angew. Chem.* **1997**, *109*, 1096; *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 1032; T. P. Begley, *Nat. Prod. Rep.* **2006**, *23*, 15) and to pyridoxal (see, e.g., T. B. Fitzpatrick, C. Moccand, C. Roux, *ChemBioChem* **2010**, *11*, 1185) are—in part—of extraordinary chemical complexity. Such a degree of complexity in a biosynthetic pathway to a potentially primordial biomolecule may be a hint at a scenario in which the molecule (or one of its type) would have fulfilled a catalytic role before its biosynthesis became established, which (in the extreme case) means that it may have been formed pre-biotically. Extraordinary complexity of biosynthetic pathway might reflect extraordinary evolutionary pressure that acted upon the emergence of biosynthetic access to a primordially pre-existing molecule (or its structural type), the functional potential of which was essential to the system's evolution (cf. for example, ref. [17]). As regards the structural type and/or function of thiamine and pyridoxal, demonstrating the existence of potentially prebiotic pathways to these molecules remains an open challenge.

