107. Chemistry and the Origin of Life[1])

by **Albert Eschenmoser**[a])* and **M. Volkan Kısakürek**[b])

[a]) Laboratorium für Organische Chemie, Eidgenössische Technische Hochschule, Universitätstrasse 16, CH-8092 Zürich

[b]) Organisch-chemisches Institut, Universität Zürich-Irchel, Winterthurerstrasse 190, CH-8057 Zürich

Dedicated to *Vlado Prelog* on the occasion of his 90th birthday

(24.V.96)

The natural genesis of life on Earth is a hypothesis of evolutionary science; it is the task of synthetic organic chemistry to test this hypothesis experimentally. The aim of an experimental aetiological chemistry is not primarily to delineate the pathways along which our ('natural') life on Earth *could* have originated, but to provide decisive experimental evidence, through the realization of model systems ('artificial chemical life'), that life *can* arise as a result of the organization of organic matter.

Chemist *Friedrich Wöhler*'s discovery of synthetic urea in 1828 is one of the roots of the development which, in our century, led to molecular biology. For the first time, it was observed that a product of animal (and human) metabolism could be made out of materials from the mineral world without the involvement of a living organism. This achievement was in flat contradiction to the then prevalent vitalist doctrine, according to which 'matters of life' could only be produced by living organisms. The 'synthesis' of urea was the first contribution of synthetic chemistry to the process of systematic demystification of the material aspects of the phenomenon 'life' by chemistry and biology, a development that reached its full impetus in our century. Within chemistry, *Wöhler*'s legacy is to be found today in the achievements of organic natural-product synthesis; in the last third of our century, these have given rise to the perception that chemists are now in a position where they essentially can synthesize any low-molecular-weight natural-product molecule, no matter how complicated, provided that they are sufficiently determined and willing to invest the necessary labor.

The main contributions to this process of demystification came, naturally, from biological chemistry and, above all, from molecular biology, the field that had set out to advance explosively in the wake of the discovery of the DNA double helix. Just to what extent 'life' as a material process has become demystified by this development is meanwhile made starkly clear to us by biotechnology, molecular biology's offspring. With biotechnology, molecular synthesis, previously a monopoly of chemistry among the natural sciences, has entered the realm of biology. As in chemistry, we experience again how our being in command of 'making' molecular structures of biological origin touches

[1]) This article is a slightly extended version in English of a chapter written in German by *A. E.* '*Zur Frage nach der Entstehung des Lebens*' to appear in '*Mensch und Natur*', Festschrift zur 250-Jahr-Feier der Naturforschenden Gesellschaft, Zürich (Ed. G. Benz), October 1996. *M. V. K.* has been responsible for the translation from German into English.

on our relation to nature in a way that goes beyond the rational. Physicist *Richard Feynman*'s dictum *'What I cannot create, I do not understand'* may indeed be overstated, but it gets to the heart of the matter with regard to that affective aspect in our attitude towards creative synthesis in general, and towards the problem of life's origin in particular.

The pathway from *Wöhler*'s discovery of artificial urea to the natural-product syntheses of this century carries on, when pursued consistently, to a radical challenge for synthetic organic chemistry in the first half of the coming century: the *creation of artificial chemical life*. Not so long ago, such an objective would have been dismissed as entirely fanciful. Today, however, it appears realistic, indispensable in fact. Regarding the thermodynamic and kinetic prerequisites, eminent physical chemists active in the field of self-organization theory of organic matter (*M. Eigen*, *I. Prigogine*, *H. Kuhn*, and others) have paved the way for the organic chemists, not to the least psychologically. Solutions to the problem of creating artificial chemical life would probably represent the contribution of farthest reaching philosophical impact that synthetic organic chemistry may have the opportunity to make to the everlasting process of enlightenment through science.

Artificial Chemical Life. – Being confronted with the term 'artificial chemical life', we first must rid ourselves of a 'defensive reflex' which we experience, because our comprehension of life is conditioned by those life forms which we (directly or indirectly) catch sight of in the empirical world. Even if we take the biologically most primitive organisms known to us, they still represent extremely complex systems, possessing a degree of organization of organic matter far exceeding that to be meant in this context. It would be a misconception to believe that experiments directed towards the creation of 'artificial chemical life' should centre around 'reproducing' the simplest of the life forms known to us today, *i.e.*, constructing their genomes and, not to forget, the entire palette of associated enzymes by so-called total synthesis. The objective is simpler in nature, but also directed at something more fundamental: *through an experimental search for chemical models of a transition between inanimate and living organic matter, we shall raise the issue of the necessary, as well as sufficient, structural and functional prerequisites of a process which we could agree upon as constituting 'life' in its most elementary form.*

If chemistry is to be engaged in the problem of the *origin of life*, then the objective cannot be to prove experimentally the actual pathway along which our biological life did commence about 3.8 billion years ago, nor, primarily, to demonstrate the way in which this life on Earth *could* have come into being; ultimately, the task is to demonstrate experimentally that what we can agree upon as constituting the essence of elementary life *can* arise out of inanimate organic matter. What we would expect to emerge from such studies is not *the* model of a genesis of 'life', but *several* models, which may differ widely. There will be those which deserve the rank of valid models for an origin of our biological life on Earth, in that they will fulfil, from a chemical and geological point of view, the requirements for assembly and function of their molecular structures under the (hypothetical) geochemical conditions of a primeval Earth. However, there will be other models for which this will clearly not be the case, the fulfilment of such requirements not having been intended in their design in the first place. The situation will reflect what has been characteristic for organic chemistry throughout its entire history: whereas its orginal task was to deal with the carbon compounds occurring in living nature, organic chemistry

invariably transcended this task and became involved in the creation of an ever growing world of artificial organic molecules.

There is, at present, no consensus about what *life in its most elementary form* is supposed to be. Among researchers interested in this question, three camps, essentially, can be made out: the 'geneticists', the 'metabolists', and the 'compartmentalists', depending on which aspect representatives assign pre-eminence in their definition of (minimal) chemical life, and stress most in their conception of life's origin, as well as in their strategies for an experimental realization of respective models. Such differences in emphasis are in no way detrimental to the cause, they reflect the complexity of the problem and channel efforts in the experimental design of models in three directions, of which all three – there is agreement on this point – are essential for *evolved* life, if not necessarily for *minimal* life.

Were a synthetic chemist to take up the challenge of designing an experimental chemical model of life's origin, he would – in the authors' opinion at least – consider the general view of the 'geneticists' as the most promising. This puts the constitutional self-assembly of a 'genetic system', a 'replicator', at the starting point. In more detail, albeit still formulated at a basic level, the task facing the chemist could be as follows: To experimentally delineate the pathway for a heterotrophic assembly of a *family* of molecular structures which can *carry* combinatorial structural information, *replicate* autocatalytically, and *vary* its information content by mutations. The *constitutional diversity* of the family must give rise to (and implicitly code for) a *conformational diversity* that has to generate a spectrum of structure-specific (therefore 'inheritable') reactivities and autocatalytic properties prone to act as selection factors in replication. Such a 'chemical phenotype' would have to provide the family with a potential to evolve – in interaction with an environment and supported by compartementalization – along gradients of increasing efficiency, diversity, and control of catalytic function, towards increasing metabolic independence of the system from the environment.

Disregarding the conspicuous vagueness on the important aspect of cellular compartmentalization, this formulation of the challenge for the experimental chemist is essentially an attempt to abstract the picture molecular biology is giving us of life as a chemical process, and to extrapolate it down to such a conceptual level, on which a chemist is able to design experimental models. On the other hand, we can also interpret it as an attempt to comprehend life in one of its most elementary forms, in keeping with the postulate referred to above, namely, that systematic experimentation towards such models is bound to produce results which would illustrate as well as illuminate, widen or, perhaps, confine, certainly influence and greatly stimulate the discussion on the problem of defining life on a physical level.

The Key Problem: the Coding of the Phenotype. – The key requirement of an experimental model system for chemical life is the system's potential to evolve. It is only on attempting a design, in detail, of artificial molecular structures which might meet such a requirement, that the enormous mechanistic complexity of natural-life processes, including those which are held to be the most primitive, is fully recognized. In such attempts one is forcefully reminded of what is, to be sure, a truism; however, one hardly ever becomes aware of its significance more radically than in this context: catalysis, that basic phenomenon of chemistry, is – as its variant '*autocatalysis*' – the alpha and omega of chemical life.

Were it desired to stipulate a center of events within a living cell, then, from an aetiological point of view, the choice most probably would fall on the translation of the structural information of the genotype into the structure of the phenotype. This is where we find that chemically tremendously refined strategy of the living cell: maintaining the entire intracellular action under virtually total control through regulation by selective catalysis (and *anti*-catalysis), and, atop, through strictly obligatory coding for the synthesis of proteins, the protagonists of the phenotype. It is in this strategy, incorporating, above all, the automatism of genetic *pre*registration of phenotypic alterations where – in union with the genotype's ability for replication – the basic mechanistic prerequisite for the process of biological evolution is to be found. This 'enslavement' of the phenotype through coding its synthesis by the genotype of entirely different molecular structure is chemically so complex in its structural and functional requirements that chemists and biologists are unanimous in regarding the 'genetic code' and the mechanism of its functioning, as we know them today, as a highly matured achievement of biological evolution, which could not conceivably have been a part of the beginning.

Right at the other end of the spectrum of conceivable relationships between the molecular structures of genotype and phenotype is their being identical (*S. Spiegelman*). A family of molecular structures that constitute a genotype might, in principle, be capable of simultaneously playing the phenotype's role (and thereby be capable of evolving), if its constitutional diversity is accompanied by a conformational diversity (diversity of molecular shape) which – in analogy with the proteins – could fulfil (at least in principle) the requirements for the emergence of phenotypic catalytic functions. Selection and inheritance of such functions would implicitly be ensured, the genotype would still – so to say – 'code the synthesis of the phenotype', since the two are structural aspects of the same molecular species.

From the chemist's point of view, there is an important variant between the two extremes of the structural relationship between genotype and phenotype: the environmentally initiated post-replicative modification of the covalent structure of the genotype. Constitutional alterations of this kind may lead to additional (phenotypic) functions which can, if inheritable, be evolutionarily relevant. They will be inheritable if the underlying structural modification takes place in a structure-specific manner, which is equivalent to their emergence being coded for by the genotype. What thus arises is the prospect of a 'chemical *Darwin*ism', a fascinating and important area of research hitherto untouched in chemistry. *This then is the prospect awaiting experimental exploration by organic chemists: research into (chemo-, regio-, diastereo-, and enantio)selective reactivity and catalysis of organic molecules, directed towards an experimental chemical aetiology of life.*

Prebiotic Chemistry and the 'RNA World' Hypothesis. – What is the state of *experimental* aetiological research in chemistry today?

Rooted in the pioneering ideas of *A. I. Oparin* and *J. B. S. Haldane* in the thirties and initiated by *Urey-Miller*'s famed experiment in 1953 (formation of the simplest proteinogenic amino acids using electrical discharges in an atmosphere containing water, hydrogen, methane, and ammonia), a socalled 'prebiotic chemistry' has developed (*S. Miller, J. Oro, L. E. Orgel, J. Ferris*, and others), giving us today a fairly convincing picture of the type of chemical reactivity which may have been involved in the origin of the major

molecular building blocks of our biological world. The two principal chemical pillars of today's life, the proteins and the nucleic acids, are derived from building blocks (α-amino acids, carbohydrates, purines, and pyrimidines) which do indeed possess *elementary* molecular structures, elementary in the sense that representatives of these families of substances do demonstrably form under conditions of a kind that the possibility of their formation under prebiotic geochemical conditions appears to be a well founded (and hence today generally accepted) hypothesis. *Where* such a formation under 'geochemical conditions' might have occurred, or *where* the necessary reactive carbon-, nitrogen-, and oxygen-containing precursors (*e.g.* hydrocyanic acid, cyanamide, cyanoacetylene, formaldehyde, glycolaldehyde *etc.*) might have come *from* (questions concerning, *e.g.* energy sources, terrestrial *vs.* extraterrestrial origin of organic materials) are matters on which the views differ. However, there is a plethora of conceivable possible locations and provenances, and, while the subject presents an important field for further search, it would not seem to pose prohibiting difficulties from a chemical point of view. This assessment is based on the hypothesis of a heterotrophic (relying on already preformed, reactive organic substances) origin of life on Earth. We thereby pass over radically 'non-conservative' proposals, such as those which would, for example, view the beginning of evolution in autotrophic (so far unknown and speculative) 'metabolic' reaction cycles, or place it in the wealth of forms of the mineral world, or speculate wildly that it was of extraterrestrial origin.

The heterotrophic 'replicator' of the geneticists' scenario must, in all probability, be a family of molecules which are polymers. Low-molecular-weight systems cannot (or, more cautiously, can hardly) be envisaged to fulfil the central requirement of storage and autocatalytic replication of *combinatorial* structural information. There is the further requirement of extensive structural regularity (if not identity) of the monomeric building blocks, since, to the chemist, the chances for autocatalytic replication of irregular polymeric information carriers would appear to be drastically reduced on mechanistic grounds. What all that means can be exemplified by RNA, the 'official' candidate for the role of the first replicator in the mind of many biologists. The carbohydrate building blocks bearing the informational elements (the nucleobases) in a combinatorial arrangement are both constitutionally and configurationally identical, as well as with regard to their sense of chirality, and they are connected together in the same way *via* phosphodiester groups. It was the discovery of the ribozymes (RNA strands that can catalyze phosphodiester transesterifications without proteins) by *T. Cech* and *S. Altmann*, that led to the proposal of what we know today as the 'RNA world' hypothesis (*W. Gilbert*). In this it is considered that RNA was the original replicator, that it assembled under natural conditions in the absence of (then non-existing) enzymes, that it replicated autocatalytically and developed catalytic functions to promote its assembly and reproduction by virtue of its (implicitly sequence-coded) shape diversity, and, eventually, that it evolved to the RNA-DNA-protein world of the simplest biological organisms.

The concept of an RNA world merits the status of an aetiological scientific theory, at least as far as the first part of its scenario is concerned. With respect to one of the two determinants of evolutionary processes – the (physical) *laws* and the (chance) *events* – the claims of the hypothesis can be subjected to experimental testing, namely, the alleged chemical properties of RNA, its potential for self-assembly, non-enzymatic self-replication, and first steps in evolving.

And how do things look in this respect? Fairly bad so far. Whereas the question of availability of starting materials (racemic ribose, ribose phosphates, nucleobases) may not pose crucially serious difficulties, with respect to the details of a constitutional self-organization of building blocks into functional homochiral oligo-ribonucleosides, there are difficulties and questionable points in droves. Experiments carried out so far still leave it an open question whether a constitutional self-assembly under prebiological conditions could be assumed for RNA. *Orgel*'s work, over two decades, on non-enzymatic template-directed synthesis of RNA strands has (so far) failed to demonstrate that RNA could really have replicated under natural conditions. Not least because of this failure, but also as a challenging consequence of the discovery of ribozymes, research on (enzymatic!) *in vitro* evolution on RNA 'sequence libraries' has been initiated (*J. Szostak*, *G. Joyce*) aiming at the discovery of specific RNA sequences which could act as general 'polymerases' in a replication of RNA. These efforts are stimulated by the daring idea that such 'polymerases' might have formed by pure chance through self-assembly under pre-biological conditions.

Chemical Aetiology of Nucleic-Acid Structure. – An alternative approach to the problem of life's origin, one, which is rigorously experimental and largely free of speculation, is to focus on the quest for a chemical rationalization of the structure type of our present-day natural nucleic acids through a systematic investigation of the chemistry of nucleic-acid *alternatives* (*Fig. 1*). Such alternatives are structural variants, which, according to their potential for constitutional self-assembly as assessed by chemical reasoning,

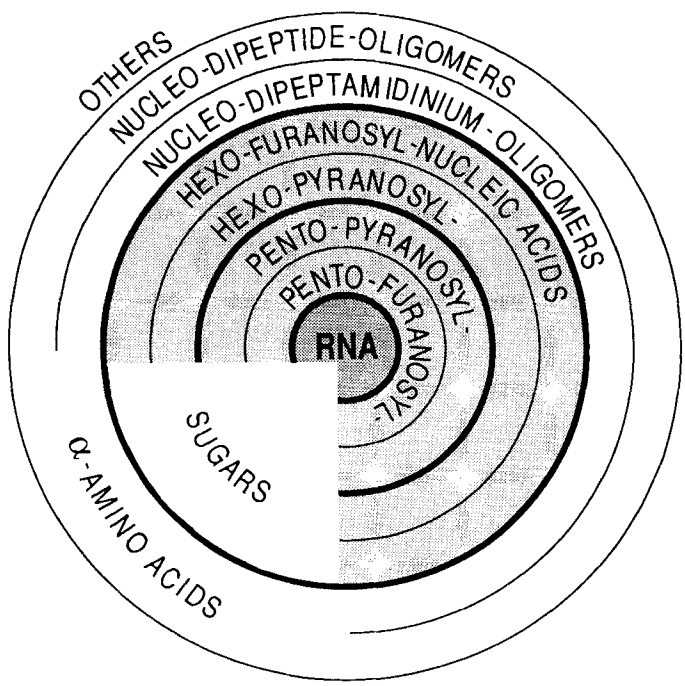

Fig. 1. *Chemical aetiology of nucleic-acid structure: Nature chose RNA (and, later, DNA) as genetic system(s) out of a diversity of potential structure alternatives. Why RNA and* not *one of these alternatives?* (Drawing by *M. Bolli*)

could have, but did not (or *may* have only transiently) become Nature's genetic system. Insight into the reasons for the selection of our present-day nucleic acids may be gained through the chemical synthesis of such alternatives, experimental ascertainment of those of their chemical properties which are relevant in context (base pairing, self-replication, phenotypic characteristics), and comparison of these properties with those of the natural nucleic acids. Systematic comparisons of this kind will lead to an understanding, at the chemical level, of the presumed functional superiority of the natural structure type over its potential alternatives. Were we to succeed in tieing an integrated net of comparison encompassing a complete set of relevant alternative stuctures over the natural structure type, we would come close to an experimentally supported, chemical rationalization of the emergence of the nucleic-acid structure type in evolution. It might, however, be the case that alternative systems thereby encountered would not conform to the *Darwin*ian extremist's conjecture, according to which the natural structure type should prove to functionally excel all alternatives. This would demand our very special attention, since functional superiority at the *chemical* level might be different from, and more important for the emergence of a primitive replicator in evolution than, functional superiority at the evolved *biological* level. Should such a situation arise, it would stipulate as extensive a chemical investigation as possible of the alternative system, and this in two directions; one would be the synthetic chemistry related to the system's potential for constitutional self-assembly under natural conditions and, the other, the chemistry at a functional level, related to the system's potential to evolve. Research towards the latter objective might even itself 'evolve' into the quest as well as into the opportunity for exploring the alternative system's own 'biology'. Such an 'alternative biology' would have to be systematically compared with (corresponding elements of) the biology we know. To understand through comparison is what the chemist in his research can strive for.

Previous observations on the prebiotic chemistry of carbohydrates teach us that C_5 and C_6 sugars, aldopentoses and aldohexoses, are to be assessed a comparable self-assembly potential. Why, then, has Nature chosen a pentose and not a hexose for its nucleic acids? And, given a pentose, why then ribose out of the four diastereoisomeric pentoses? And, finally, why ribofuranose and not ribopyranose? *To pose this escalade of questions and to pursue the questions experimentally is to mimic natural variation and selection among nucleic-acid alternatives.* If the configurational dichotomy of the nucleosidic bonds in the carbohydrate building blocks is taken into account, and the constitutional diversity of phosphodiester junctions as well, then the result is so many (formally) possible alternatives (*Fig. 2*) enter the scene that more refined selection criteria for fixing priorities for experimentation are required. Chief among these is function prognosis: prediction, derived from either organic conformational analysis or computer-supported molecular modelling, of which variants stand any chance at all of proving themselves as functioning base-pairing systems. Considering that the escalade of 'why questions' above can go beyond the realm of carbohydrates, that the question 'why carbohydrates and not α-amino acids as backbone building blocks' arises, and, realizing furthermore, that there are also alternatives for the nucleic-acid bases, the purines and pyrimidines (*Fig. 3*), as well as for the phosphodiester bridge, then it becomes clear that a comprehensive experimental commitment by organic chemists, directed by effective priority assessment, would be needed. *Without such a commitment we can hardly expect ever to comprehend why our world is an DNA-RNA-protein world.*

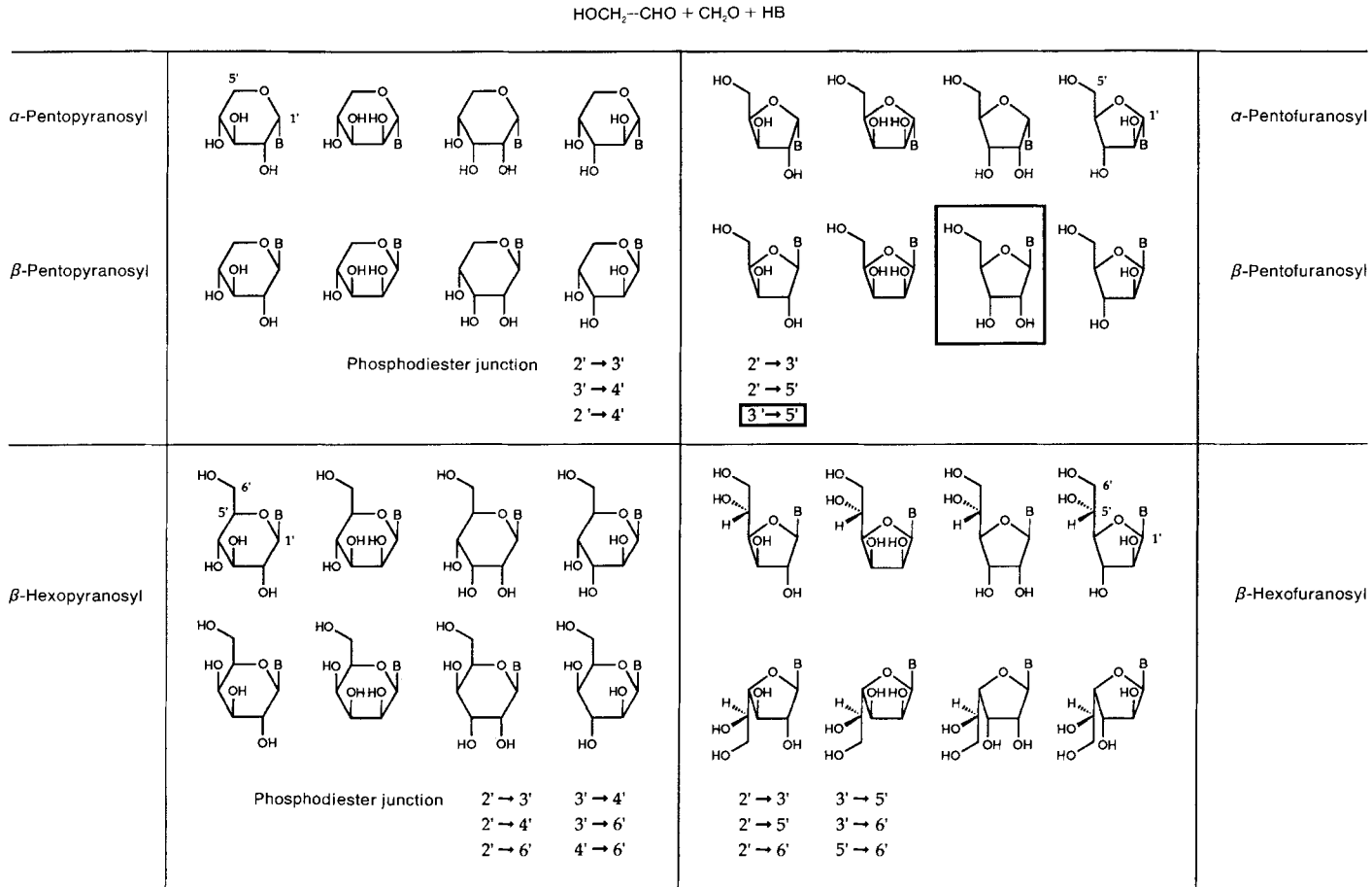

Fig. 2. *Chemical aetiology of nucleic-acid structure: structural diversity of possible alternative sugar building blocks related to the natural* D*-ribofuranosyl unit of RNA*

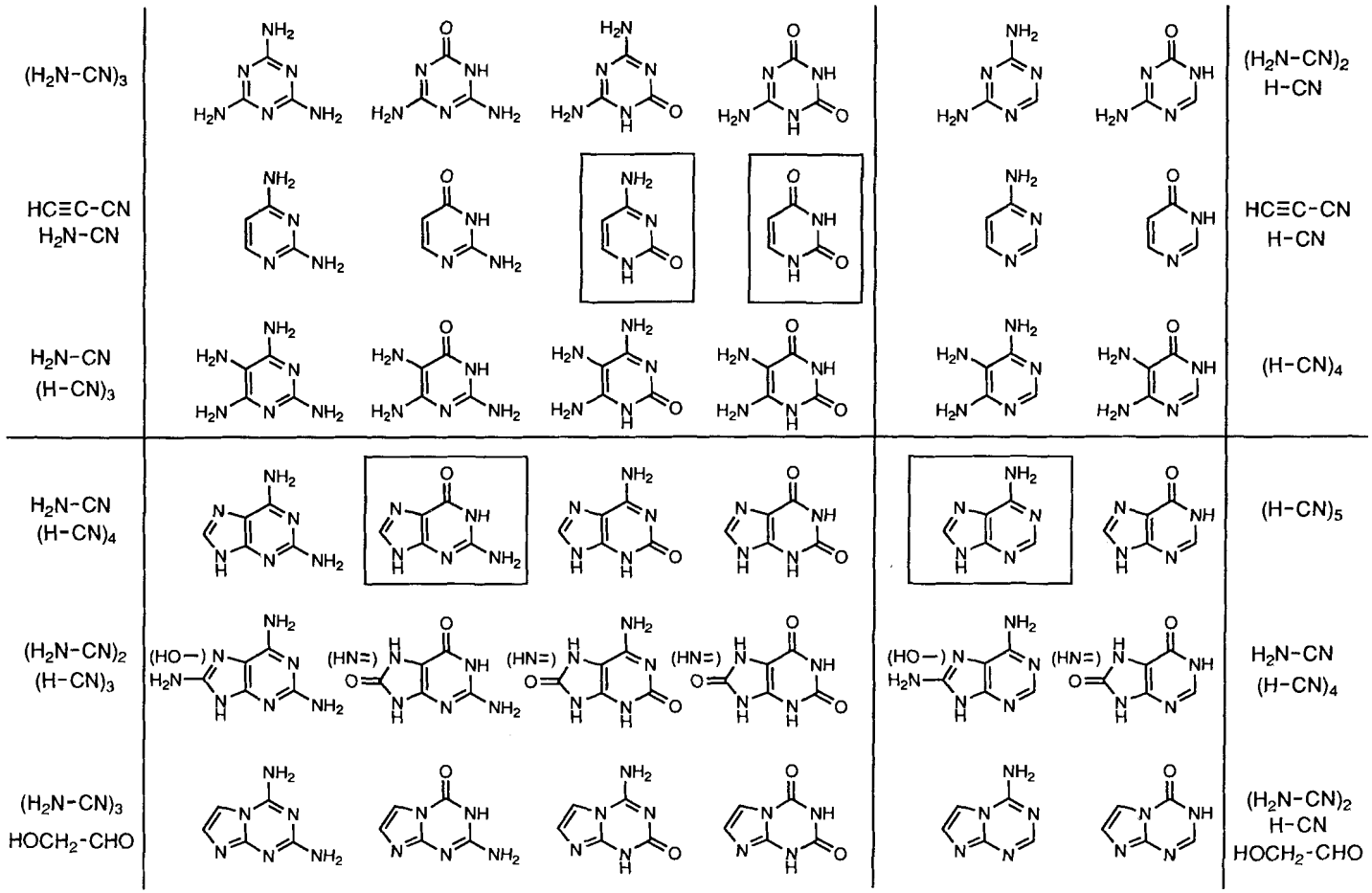

Fig. 3. *Chemical aetiology of nucleic-acid structure: structural diversity of nucleobase alternatives whose potential for constitutional self-assembly is considered to be comparable to that of the canonical bases*

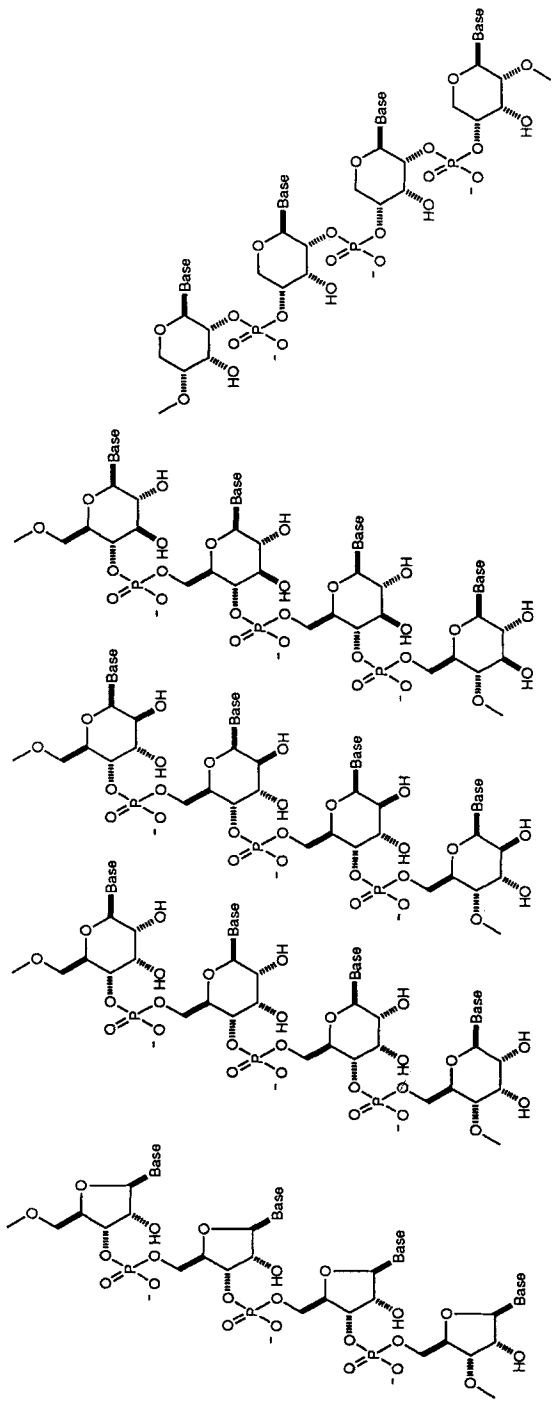

Fig. 4. *Constitution and configuration of nucleic-acid alternatives studied experimentally in the ETH laboratories*

In conclusion, a glimpse at some of the results from the ETH laboratories may demonstrate how experiments directed towards such a chemical aetiology of nucleic-acid structure can indeed be rewarding. *Fig. 4* gives the structural formulae of the nucleic-acid alternatives, which have been chemically synthesized, and whose pairing capabilities have been studied. In three of them, the ribofuranose is replaced by a hexapyranose sugar with otherwise unchanged structure type. While (6′→4′)-oligonucleotides incorporating *gluco*pyranosyl building blocks have been found not to exhibit any regular base pairing at all, *allo*pyranosyl and *altro*pyranosyl (6′→4′)-oligonucleotides do show base pairing which, however, is far inferior to that of natural RNA with regard to strength, selectivity, and regularity of pairing. Complementary investigations on model systems clearly indicate that it is the steric hindrance of the pairing conformation by specific hexopyranose hydroxy groups that is responsible for the weakness of the base pairing in these systems. This interpretation is also to be extrapolated to those remaining hexopyranosyl (6′→4′)-oligonucleotides which have not been examined directly. Hexopyranosyl-RNA alternatives of this type, therefore, stood no chance in the evolution of the nuleic acids in competition with RNA, and this for *functional* reasons.

Entirely different results were obtained with the fourth, recently studied nucleic-acid alternative, pyranosyl-RNA. This system is *isomeric* with RNA, composed of the same building blocks but differing in that ribose is incorporated in the pyranosyl, rather than the furanosyl, form (*Fig. 4*). Experimentally, p-RNA, in comparison with natural RNA, exhibits not only *stronger,* but also (with reference to the constitutional mode of base pairing) more *selective* base pairing. The most recent experiments also show p-RNA to be superior to RNA, at the *chemical* level, in its potential for (non-enzymatic!) replication: p-RNA base sequences can be copied replicatively through template-controlled ligation of small subsequences under *potentially natural* conditions; for RNA, corresponding experimental evidence for such a capability has hitherto been lacking. These chemical features of the (presumably) thermodynamically more stable and (presumably) more readily self-assembling p-RNA demand attention; have we hit upon a former evolutionary *competitor* or even *precursor* of our present-day nucleic acids? Fortunately, there is neither a need, nor the time for indulging in speculations and making media-effective 'proposals'; needed is a comprehensive chemical investigation into p-RNA as an informational molecular system and an incisive experimental inquiry on its potential for constitutional self-assembly in comparison to RNA. The properties of p-RNA to be uncovered by such studies will have to speak for themselves.