

ducing overall coevolutionary trajectories quite different from those predicted by study of the component local interactions (15).

Some recent empirical studies have shown strong evidence for selection mosaics and coevolutionary hotspots (16, 17). For example, lodgepole pines differ geographically across northwestern North America in the sizes and shapes of their cones, which have been molded through a combination of adaptation to fire and two major seed predators, pine squirrels and red crossbills. In some regions the pines and crossbills show evidence of coevolution—large cones shaped in ways difficult for crossbills to handle, and crossbill morphology indicating local counteradaptation to these cones. In other parts of the Rockies, coevolution between the pines and crossbills is swamped by other selective forces, including predation on cones by abundant pine squirrels. Recent evidence suggests that the coevolutionary hotspots between crossbills and pines resulted from post-Pleistocene events, creating regions where pine squirrels were absent and crossbills were abundant and relatively sedentary. Similar selection mosaics are likely to be even more common in interactions involving parasitic species, which often have restricted potential for gene flow over large distances.

Coevolution and Human Welfare

As we learn more about the ongoing evolution of species interactions across Earth's complex landscapes, the effects seem to permeate every aspect of the health and welfare of human societies. A major spur to the development of the current ecological theory of evolving inter-

actions came from models of the evolution of virulence in parasites (18). Those models suggested that differences among populations in the transmission dynamics of parasites could shape differences in the evolution of virulence. Continuing refinement of the theory of the evolution of virulence will be crucial to understanding how we can minimize the evolution of new virulent strains of pathogens that attack us, our domesticated animals, our crop plants, and conserved species within reserves.

For human pathogens, we can confront the problem partially through surrogate coevolution by developing procedural methods that mimic the evolution of new defenses. Antibiotics are the equivalent of mutant resistant hosts, but rapid evolution of resistance to antibiotics has shown how difficult it is for species to develop durable defenses against enemies. The same problems apply to the development of resistant crop varieties and the management of biological control agents. To deploy surrogate coevolution effectively, we require successful case studies to use as models. Those case studies are the interactions that have been shaped over millennia within natural biological communities. No amount of research funding could ever replace these valuable natural experiments.

We live at a time when we are increasingly manipulating much of Earth's biodiversity for our own ends. Indeed, Janzen (19) has referred to us as "the most coevolutionary animal of them all." As we continue to manipulate biodiversity, our experience so far with the evolution of virulence in diseases, short-term effectiveness of resistant crop varieties, and rapid evolution of interactions within natural communities suggests that the

health and welfare of human societies will demand an increased understanding of the ongoing evolutionary dynamics of species interactions.

References and Notes

1. L. Margulis and R. Fester, Eds., *Symbiosis as a Source of Evolutionary Innovation: Speciation and Morphogenesis* (MIT Press, Cambridge, MA, 1991).
2. J. Maynard Smith and E. Szathmáry, *The Major Transitions in Evolution* (Freeman, San Francisco, 1995).
3. C. A. Kearns, D. W. Inouye, N. M. Waser, *Annu. Rev. Ecol. Syst.* **29**, 83 (1998).
4. O. Pellmyr et al., *Am. Nat.* **148**, 827 (1996).
5. D. Schluter, *Science* **266**, 798 (1994).
6. M. Berenbaum and A. Zangerl, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13743 (1998).
7. J. J. Burdon and P. H. Thrall, *Am. Nat.*, in press.
8. C. M. Lively, *ibid.*, in press.
9. N. B. Davies and M. de L. Brooke, *J. Anim. Ecol.* **58**, 225 (1989).
10. J. N. Thompson, *Trends Ecol. Evol.* **13**, 329 (1998).
11. G. Dwyer, S. A. Levin, L. Buttel, *Ecol. Monogr.* **60**, 423 (1990).
12. J. N. Thompson, *The Coevolutionary Process* (Univ. of Chicago Press, Chicago, 1994).
13. S. Gandon et al., *Proc. R. Soc. London Ser. B* **263**, 1003 (1996).
14. M. E. Hochberg and M. van Baalen, *Am. Nat.* **152**, 620 (1998).
15. S. L. Nuismer, J. N. Thompson, R. Gomulkiwicz, *Proc. R. Soc. London Ser. B* **266**, 605 (1999).
16. C. W. Benkman, *Am. Nat.*, in press.
17. M. A. Parker, *ibid.*, in press.
18. R. M. May and R. M. Anderson, *Proc. R. Soc. London Ser. B* **219**, 281 (1983).
19. D. H. Janzen, *Crawford Lectures R. Swed. Acad. Sci.* **3**, 1 (1984).
20. D. E. Soltis et al., *Syst. Bot.* **17**, 607 (1992).
21. J. M. Brown et al., *Mol. Ecol.* **6**, 215 (1997).
22. J. N. Thompson, unpublished data.
23. I thank C. Benkman, B. Cunningham, K. Merg, S. Nuismer, and T. Steury for helpful comments. Supported by NSF grant DEB-9707781 and by the National Center for Ecological Analysis and Synthesis (Coevolution Working Group), NSF grant DEB-9421535.

REVIEW

Chemical Etiology of Nucleic Acid Structure

Albert Eschenmoser

Systematic chemical studies indicate that the capability of Watson-Crick base-pairing is widespread among potentially natural nucleic acid alternatives taken from RNA's close structural neighborhood. A comparison of RNA and such alternatives with regard to chemical properties that are fundamental to the biological function of RNA provides chemical facts that may contain clues to RNA's origin.

Wir wollen nicht nur wissen *wie* die Natur ist (und *wie* ihre Vorgänge ablaufen), sondern wir wollen auch nach Möglichkeit das vielleicht utopisch und anmassend erscheinende Ziel erreichen, zu wissen, warum die

Natur *so und nicht anders* ist. [Albert Einstein (1, p. 126)]

[We not only want to know *how* nature is (and *how* her transactions are carried through), but we also want to reach, if possible, a goal which may seem utopian and presumptuous, namely, to know why nature *is such and not otherwise*.]

Chemical etiology (2) of nucleic acid structure refers to systematic experimental studies

aimed at narrowing the diversity of possible answers to the question of why nature chose the structure type of ribofuranosyl nucleic acids, rather than some other family of molecular structures, as the molecular basis of life's genetic system. The quest is to uncover the criteria by which nature arrived at this choice; comprehending these criteria in chemical terms would constitute a central element of any theory on the origin of the particular kind of chemical life known today. The strategy is to conceive (through chemical reasoning) potentially natural alternatives to the nucleic acid structure, to synthesize such alternatives by chemical methods, and to compare them with the natural nucleic acids with respect to those chemical properties that are fundamental to the biological function of RNA and DNA. Basic to this research is the

The Skaggs Institute for Chemical Biology at The Scripps Research Institute (TSRI), 10550 North Torrey Pines Road, La Jolla, CA 92037, USA, and Laboratorium für Organische Chemie, Eidgenössische Technische Hochschule (ETH), Universitätstrasse 16, CH-8092 Zürich, Switzerland.

supposition that the RNA structure originated through a process that was combinatorial in nature with respect to the assembly and functional selection of an informational system within the domain of sugar-based oligonucleotides. The investigation can be viewed as an attempt to mimic the selectional part of such a hypothetical natural process by chemical means. In principle, the study has no bias with regard to the question of whether RNA first appeared in an abiotic or a biotic environment (3).

The etiological relevance of the strategy depends on the criteria by which the alternatives are chosen. The condition that a given candidate must fulfill in order to be selected for study is that one can deem it a potentially natural type of molecular structure. For this to be the case, one requires the system to be structurally derivable from a $(\text{CH}_2\text{O})_n$ sugar ($n = 4, 5, \text{ or } 6$) by the same type of potentially natural chemistry that allows the structure of RNA to be derived from ribose. Whatever the specific chemistry that may have actually resulted in RNA, I hypothesize that the same type of chemistry had the potential to produce not just a single oligonucleotide system, but rather a variety of them, provided that not only ribose, but also other members of the family of aldoses were part of the environment. That such should have been the case would be consonant with the $(\text{CH}_2\text{O})_n$ -sugars' exceptional chemical status of a family of molecular structures that, despite their apparent constitutional and configurational complexity, are all elementary in the sense that they can form from elementary starting materials.

The first chemical property by which a nucleic acid alternative is to be compared with RNA is the system's capacity for informational base pairing in the Watson-Crick mode. In more extended versions of the study, comparisons may also have to focus on constitutionally different kinds of base pairing, that is, on variants involving recognition partners distinct from the canonical purines and pyrimidines. Although it would seem that the constitutional diversity of potentially natural nucleobase alternatives is much smaller than that of possible backbones, a comprehensive chemical etiology of nucleic acid structure will have to screen nucleobase alternatives analogously to backbone alternatives. In principle, the same would hold for the third component of nucleic acid structure—the phosphodiester junction.

Screening alternative backbones with respect to base-pairing capacity may reveal that a given candidate is, in fact, not a base-pairing system. Such an "alternative" can be dropped from the list of potentially relevant evolutionary competitors of RNA. On the other hand, a system may possess a base-pairing capability comparable or even superior (with regard to pairing strength) to that of RNA. Then, it will have to be subjected to comparisons with re-

spect to chemical properties that are etiological-ly more demanding than base pairing alone. Heading the list of such properties, in reference to an abiotic scenario, is the capability for (non-enzymatic) self-replication under potentially natural conditions, one of at least two general conditions to be met by any informational molecular system that eventually would have to assume the role of a "genetic system" in a given environment. The second of these conditions is a system's capacity to express a chemical phenotype in that environment. The term "chemical phenotype" is intended to denote the whole of an informational system's sequence-specific, and thus "inheritable," reactivities and catalytic properties that are able to act as selection factors for the system's replication in a given environment. This notion points to one of the directions along which an experimental chemical etiology of nucleic acid structure may have

to proceed, should potentially natural alternative base-pairing systems be encountered that cannot be eliminated in early rounds of investigation as rivals of RNA and that do not reveal through their properties why the latter should have excelled in evolution. The structural and functional properties that may gain special relevance in such a study are largely unpredictable, and whether the natural system will ever disclose the reasons for its superiority over alternatives on the level of chemical properties is uncertain. It is conceivable that an alternative system may eventually require comparison with RNA on a level that would no longer be deemed "purely chemical," because the comparison would refer to the systems' evolvability. In an extreme case, this could amount to comparing and experimentally studying elements of an artificial alternative biology.

The research program delineated here has

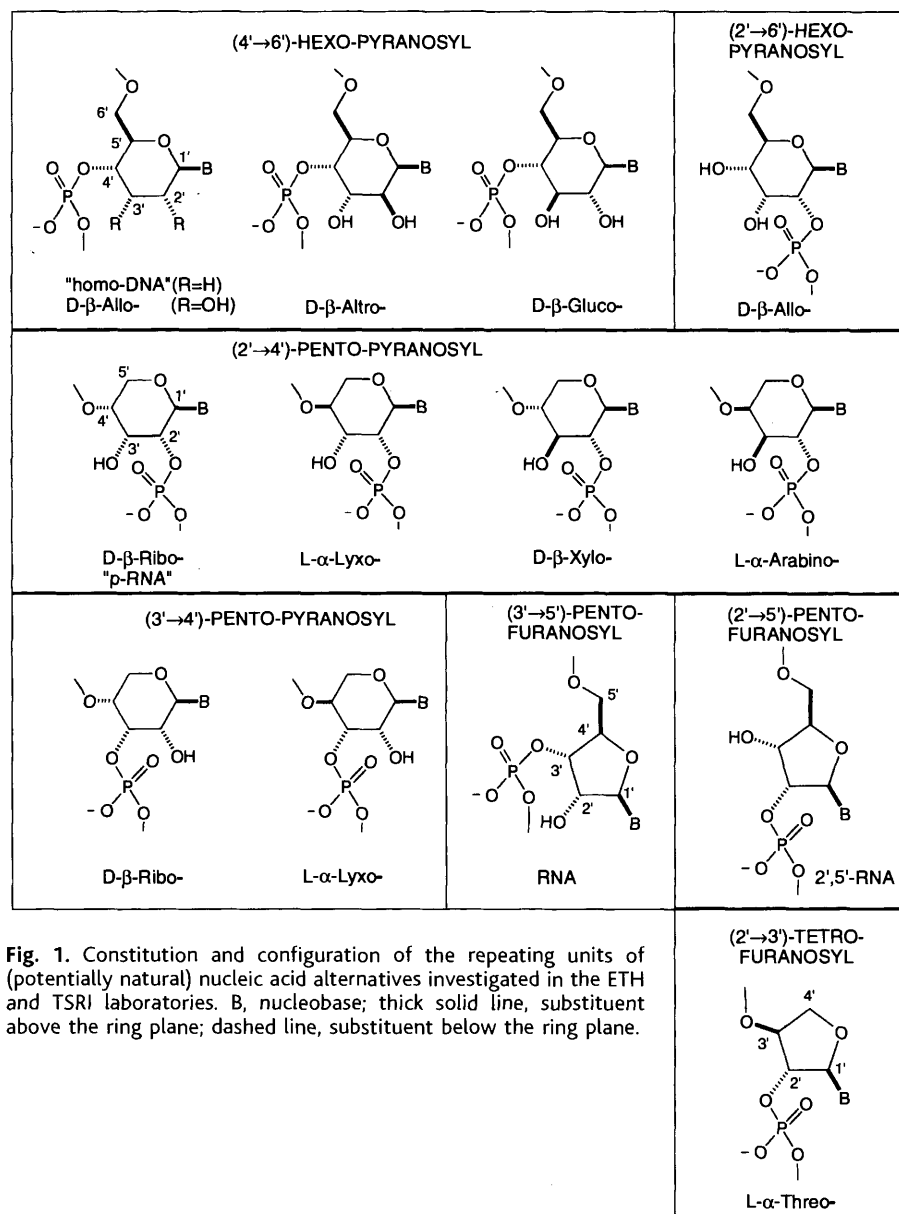


Fig. 1. Constitution and configuration of the repeating units of (potentially natural) nucleic acid alternatives investigated in the ETH and TSRI laboratories. B, nucleobase; thick solid line, substituent above the ring plane; dashed line, substituent below the ring plane.

its underpinnings in the general views held by the "geneticists" (4) school of thought on the etiology of life, pioneered by Eigen [(5); see also (6)], Orgel [(7); see also (8)], Kuhn (9), and others, and modified later in consequence of the discovery of catalytic RNA (10). The literature in the field of nucleic acid chemistry records contributions that either have adumbrated or complement the investigations surveyed here. A pioneering study by Usher [(11); see also (12–16)] dealt experimentally with the question of why RNA's phosphodiester junction is positioned between carbons 3' and 5' and not between 2' and 5'. Its conclusion pointed to the higher susceptibility to hydrolytic strand cleavage of duplexes containing a unit with a (2'→5')-phosphodiester junction. A markedly lower pairing strength of the (2'→5')-ribofuranosyl system has since been documented (17). Recently, the same question, but now asked of DNA (3), has been studied by Breslow (18) and Switzer (19). Alternative base pairs were the subject of investigations by Benner (20) and others [(21); see also (22)]. The advantages for RNA and DNA to have the sugar units held together by phosphodiester groups have been discussed in Westheimer's classic paper (23), and physicochemical aspects of the RNA structure's suitability for its role as a biological informational system have been summarized by Turner and Bevilacqua (24).

Hexopyranosyl-(4'→6') Oligonucleotide Systems

The ETH project of examining alternative nucleic acid systems was initiated in 1986 with the model study "homo-DNA" (25), an artificial oligonucleotide system differing from DNA in that the five-membered furanose ring is expanded to a six-membered pyranose ring by an additional methylene group (26), thus providing the opportunity to investigate whether five-memberedness of the sugar ring in an oligonucleotide system is a structural prerequisite for Watson-Crick base pairing (Fig. 1). Experimentally, homo-DNA was the first example (among the many now known) of a backbone-modified oligonucleotide system that shows much stronger Watson-Crick base pairing than DNA, and moreover, it was the first oligonucleotide system whose base pairing is orthogonal to that of the natural systems (25, 27). The reason for its greater base-pairing strength was postulated to be a consequence of the higher rigidity of pyranose as compared to that of furanose rings, resulting in a preorganization of the single strand's backbone conformation toward the double strand's pairing conformation (28, 29). The system also has a pronounced propensity for adenine and guanine self-pairing in the reverse-Hoogsteen mode.

Although homo-DNA was rewarding from a chemical point of view, it was never sup-

posed to be a potentially natural nucleic acid alternative (30), but it was considered as a model system for the family of hexopyranosyl-(4'→6') oligonucleotides, whose structures differ from that of homo-DNA by two additional hydroxy groups per repetitive unit and relate to the natural hexoses in the same way that RNA relates to ribose. Three systems within this family have been studied: β -allo-, β -altro-, and β -glucopyranosyl (31) (Fig. 1). None of these exhibit discernible Watson-Crick base pairing between adenine and uracil up to dodecamer strands under standard conditions. Instead, purine-purine self-pairing in the reverse-Hoogsteen mode occurs in the β -allo- and β -altropyranosyl series (but not in the β -glucopyranosyl series), although again much less efficiently than in homo-DNA. Guanine-cytosine pairing in the β -allo and β -altro systems is far weaker than in RNA; moreover, it is strongly sequence dependent with regard to the pairing mode. The data

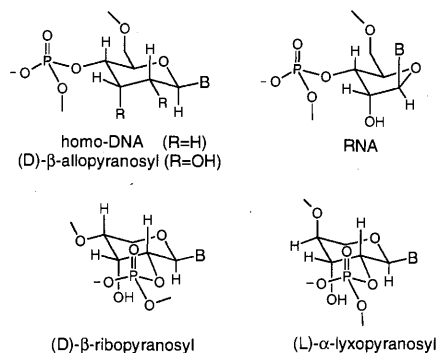


Fig. 2. Idealized pairing conformations of sugar-phosphodiester units in hexopyranosyl-(4'→6') and pentopyranosyl-(2'→4') oligonucleotides as well as in RNA duplexes.

SELF-PAIRING	$A_8 \cdot A_8$	-	-	15	<5	-	
$T_8 \cdot T_8$	RNA	r	pr	pl	pa	px	A_8
-	RNA	r	23	-	(-)	-	
-	pr	-	46	31	45	46	
<5	pl	-	43	51	67	42	
27	pa	(-)	59	68	79	55	
-	px	-	45	27	43	47	
T_8							

SELF-PAIRING		-	15	23	48	<10	
		RNA r	pr	pl	pa	px	4' TAAAAATA 2'
	RNA r	17	(-)	(-)	(-)	(-)	
22	pr	(-)	46	30	45	46	
20	pl	(-)	40	46	61	37	
49	pa	(-)	54	49	75	53	
23	px	(-)	46	26	46	44	
4' TATTTTAA 2'							

Fig. 3. Pairing, cross-pairing, and self-pairing in the pentopyranosyl-(2'→4') oligonucleotide family. Melting temperature (T_m) (in degrees Celsius) of (left) $A_8 \cdot T_8$ duplexes and (right) $d(4' \text{ TATTTTAA } 2') \cdot (4' \text{ TAAAAATA } 2')$ duplexes. r, (D)- β -ribofuranosyl (RNA); pr, (D)- β -ribofuranosyl; pl, (L)- α -lyxopyranosyl; pa, (L)- α -arabinopyranosyl; and px, (D)- β -xylopyranosyl. Peripheral rows show self-pairing of partner strands; shaded diagonal rows show pairing between partner strands of the same system; other rows show cross-pairing between partner strands of different systems, -, no pairing observed; (-), cross-pairing is not observable because of self-pairing of one of the partner strands. Conditions are as follows: 5+5 μ M in 1.0 M NaCl, 10 mM NaH_2PO_4 , and 0.1 mM Na_2 EDTA (pH 7.0). Data from (36) and (38); see also (49).

given in Table 1 demonstrate the incompetence of the β -allo- and β -altropyranosyl systems for informational base pairing.

The contrast between the base-pairing behavior of homo-DNA and its fully hydroxylated counterparts was interpreted as the consequence of intrastrand steric hindrance in the pairing conformation of the bulkier fully hydroxylated systems (Fig. 2, top left). Models, based on a nuclear magnetic resonance (NMR) structure analysis of the homo-DNA duplex ($6' \text{ -A}_5\text{T}_5 \text{ -4'}_2$) (32), indicated that the insertion of an additional hydroxy group at the equatorial 2' position of the pyranose ring in homo-DNA results in a steric clash between this hydroxy group and the neighboring nucleobase. In monodeoxy- β -altropyranosyl model systems, studies of the temperature at which ~50% of duplex molecules are dissociated into single strands (T_m) corroborated the hypothesis that adenine self-pairing in the 2'-deoxy series is comparable in

Table 1. Base-pairing in the hexopyranosyl-(4'→6') oligonucleotide series. T_m values of duplexes of self-complementary sequences 6' (or 5')-CGCGAATTCGCG-4' (or 3') were compared in neutral and acidic solution. T_m values refer to an oligomer concentration of 10 μ M in 0.15 M NaCl. Data for allo and altro sequences (referring to duplexes containing U instead of T as a nucleobase) point to (weak) reverse-Hoogsteen pairing at pH 4.3 (37).

6'-CGCGAA-TTCGCG-4'	T_m (°C)	
	pH 7.0	pH 4.3
DNA	58	44
Homo-DNA	86	75
Allopyranosyl	<3	20
Altropyranosyl	<0	13

strength to that of homo-DNA, but in the 3'-deoxy system, it is as weak as in the β -allopyranosyl series (33).

The outcome of these studies has led to the conclusion that, for functional reasons, the three hexopyranosyl-(4'→6') oligonucleotide systems investigated (34) could not have acted as viable competitors of RNA in the emergence of nature's genetic system. Because steric bulk of the hexopyranosyl sugar units ("too many atoms") was deemed responsible, studies were refocused on the sterically less bulky pentose-derived nucleic acid alternatives, no longer inquiring "Why pentose and not hexose nucleic acids?"—the question that had led to the start of the ETH project in the first place—but rather "Why ribose and not another pentose?" and "If ribose, why ribofuranose and not ribopyranose?"

Pentopyranosyl-(2'→4') Oligonucleotide Systems

The primary object of our studies has been the pyranosyl isomer of RNA (p-RNA), a system

derived from the pyranose form of ribose with the nucleobases placed equatorially on the pyranose chairs and with phosphodiester junctions between (equatorial) positions 2' and 4' (Figs. 1 and 2). Conformational analysis on the level of idealized conformations (35) predicted p-RNA to be a Watson-Crick base-pairing system that would form duplexes that will be quasi-linear or at least much less helical than RNA (Fig. 2). Experimentally, p-RNA was not only a much stronger Watson-Crick pairing system than RNA, but also a more selective one with respect to pairing modes (no self-pairing of guanine-rich sequences in the Hoogsteen or reverse-Hoogsteen mode). After an extensive investigation of the system's structural and chemical properties, the study was extended to the whole family of diastereoisomeric pentopyranosyl-(2'→4') oligonucleotide systems containing β -ribo-, β -xylo-, α -lyxo-, and α -arabinopyranosyl units as building blocks, always with the nucleobases equatorially positioned on the pyranosyl chairs. Surprisingly, all members of this family of RNA isomers were

stronger Watson-Crick base-pairing systems than RNA itself (36); the α -arabinopyranosyl isomer ranks among the strongest oligonucleotide-type base-pairing systems encountered (37). Promiscuous cross-pairing between all members of the pentopyranosyl family was observed (Fig. 3) (36, 38), suggesting a remarkable capacity of their diastereoisomeric backbones for adopting a common type of pairing conformation. Nevertheless, differences are discernible: Cross-pairing is smoother between members (between ribo and xylo and between lyxo and arabino) that have phosphodiester bridges attached in the same conformation (4'-equatorial versus 4'-axial; see Fig. 2). The pentopyranosyl (2'→4') oligonucleotides do not cross-pair with RNA. However, a somewhat capriciously sequence-dependent cross-pairing capability with DNA was observed in the case of the (L)- α -lyxopyranosyl-(3'→4') system (see Fig. 1, bottom left), which, remarkably, was found to constitute a base-pairing system by itself, in contrast to the corresponding

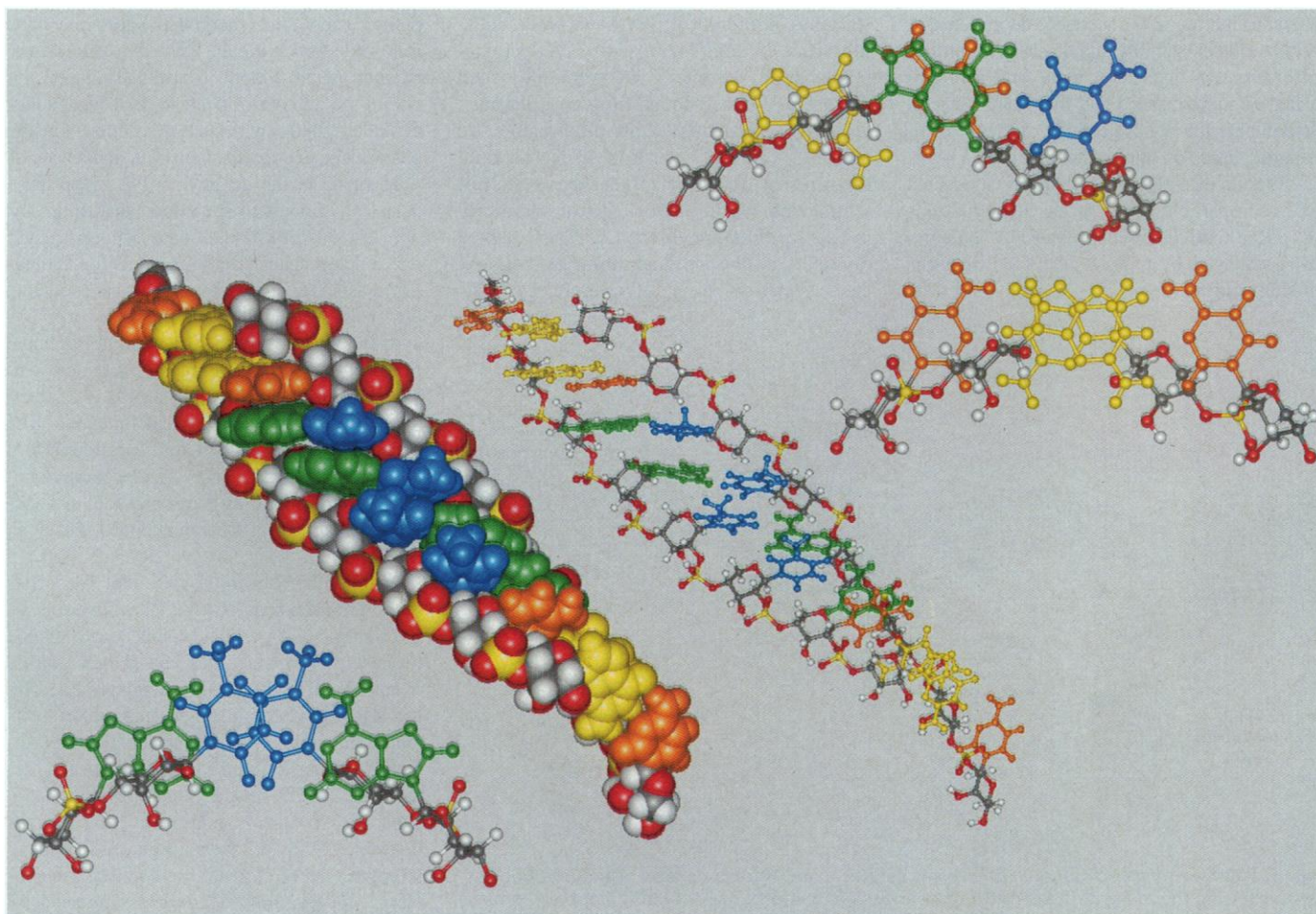


Fig. 4. Time-averaged and refined structure of an octamer duplex derived from self-complementary pyranosyl-RNA sequence 4'-CGAATTCG-2', based on a 1000-ps Molecular Dynamics calculation (Amber program) and an NMR structure analysis (40). The ball-and-stick representation and the excerpts of purine-pyrimidine and purine-purine stacking (top

right) and pyrimidine-pyrimidine "nonstacking" (bottom left) illustrate the interstrand base-stacking that is characteristic of pentopyranosyl-(2'→4') oligonucleotide duplexes. White, hydrogen; red, oxygen; gray, carbon; orange, cytosine; yellow, guanine; green, adenine; blue, thymine. Courtesy of R. M. Wolf, Novartis Pharma AG, Basel, Switzerland.

(3'→4') isomer of the ribopyranosyl series (39). The α -lyxopyranosyl-(3'→4') system is the first duplex-forming oligonucleotide system that has only five (instead of the usual six) covalent bonds in the repetitive unit of its backbone.

Pyranosyl Isomer of RNA

An NMR structure analysis by Schlönvogt *et al.* (40) of the p-RNA duplex derived from the self-complementary base sequence (4'-CGAATTCG-2') confirmed the type of pairing conformation predicted by conformational analysis, and modeling (40) based on molecular dynamics indicates that its duplex ladder structure has a weak left-handed twist (duplex derived from D-ribose) (Fig. 4). Base pairing in p-RNA is strictly antiparallel, and base stacking is overwhelmingly interstrand as opposed to intrastrand, the latter being the stacking mode common to the natural systems and most clearly displayed in a B-form DNA helix. Interstrand stacking correlates conformationally with the pronounced inclination of the p-RNA backbone toward the base-pair axes; it is also interpreted as the major controlling factor (other than base-pair constitution) in the sequence dependence of p-RNA duplex stability and the relative rates of both template-directed ligations and self-templated oligomerizations and the major factor in determining the regioselectivity of the effect of dangling bases on duplex stability (41).

Strands of p-RNA form hairpin structures with comparable ease, as do RNA strands (41). Replicative copying of p-RNA sequences is achieved by template-directed ligations of short oligomers with 2',3'-cyclophosphate

derivatives, possibly the simplest mode of phosphate activation (35, 42). Ligation is regioselective; it exclusively affords the correct (2'→4')-phosphodiester bridges, whereas the analogous cyclophosphate-mediated ligation in the RNA series fails to produce the natural (3'→5') connection, resulting instead in the isomeric (2'→5') junction (42, 43). Copying guanine-rich template sequences in the p-RNA series is not hampered by guanine-guanine self-pairing. Replication experiments, including ones with templates that form hairpin structures, have had marked success in terms of sequence copying but have failed to demonstrate (under the conditions tested) template-catalysis turnover numbers greater than one (44).

The 2',3'-cyclophosphate derivatives of p-RNA tetramers containing hemi-self-complementary base sequences such as (4'-ATCG-2') or (4'-GCCG-2') were shown to undergo self-templating oligomerization and co-oligomerization to duplexes of higher oligomers (45). The process is highly chiroselective and offers a pathway for the constitutional self-assembly of complex libraries of (largely) homochiral p-RNA strands starting from D- and L-ribose-derived diastereoisomeric tetramers. In principle, such chiroselective self-templating co-oligomerizations of activated short oligonucleotide oligomers, when starting from racemic mixtures of diastereoisomers, have the potential to break molecular mirror symmetry by deracemization, provided the sequence diversity of the product libraries will exceed a critical level of complexity (45).

This capacity, intrinsic to any informational polymer family that can chiroselectively self-assemble to form libraries of critical constitutional diversity, deserves special attention in the context of the problem of the origin of biomolecular homochirality. Homochiral p-RNA strands with an opposite sense of chirality also can pair with each other but much less strongly so and by pairing rules that are the inverse to the canonical ones (for example, cytosine pairs with isoguanine, and guanine pairs with isocytosine) (22).

Discussion

The base-pairing strength landscape of Fig. 5 surveys relative base-pairing capabilities of potentially natural nucleic acid alternatives, giving melting temperatures (T_m values) of $A_8 \cdot T_8$ and $A_{12} \cdot T_{12}$ duplexes (46) determined under standard conditions (31, 36, 38, 39, 47–49). The profile of this landscape in the pentopyranosyl region is confirmed by T_m data of duplexes with irregular AT sequences (Fig. 3) as well as with sequences containing GC base pairs. These findings falsify one of the possible hypotheses about the criteria that nature may have followed in selecting the RNA structure, namely, that maximization of base-pairing strength within the domain of pentose-derived oligonucleotide systems was such a criterion. On the other hand, in conjunction with the observations made in the hexopyranosyl series, these data lend support to the notion that optimization, not maximization, of base-pairing strength was a determinant of RNA's selection. Biological reasoning would emphasize that moderate base-pairing strength, as encountered in RNA and resulting from the high conformational flexibility of the ribofuranose backbone, was essential for the evolution of a rich diversity of nucleic acid-related biological functions. The observed base-pairing strength in the family of the conformationally more rigid pentopyranosyl oligonucleotide systems provides a background of chemical facts in support of such an argument.

Also, other observed chemical properties related to high base-pairing strength could be of etiological relevance. One is overtolerance to base-pair mismatches. Such behavior is exemplified in the pentopyranosyl series by the amazingly strong self-pairing of non-self-complementary base sequences such as (4')-TATTTTAA-(2') and (4')-TTAAAATA-(2') (see Fig. 3). The central issue, however, is the capability of nonenzymatic autocatalytic replication (50). Whereas high base-pairing strength can be expected to facilitate the selective recognition of template sequences by activated ligands and to accelerate sequence copying, it will concomitantly strengthen product inhibition of template sequences. Such inhibition has been observed to dominate strand replication in the p-RNA series.

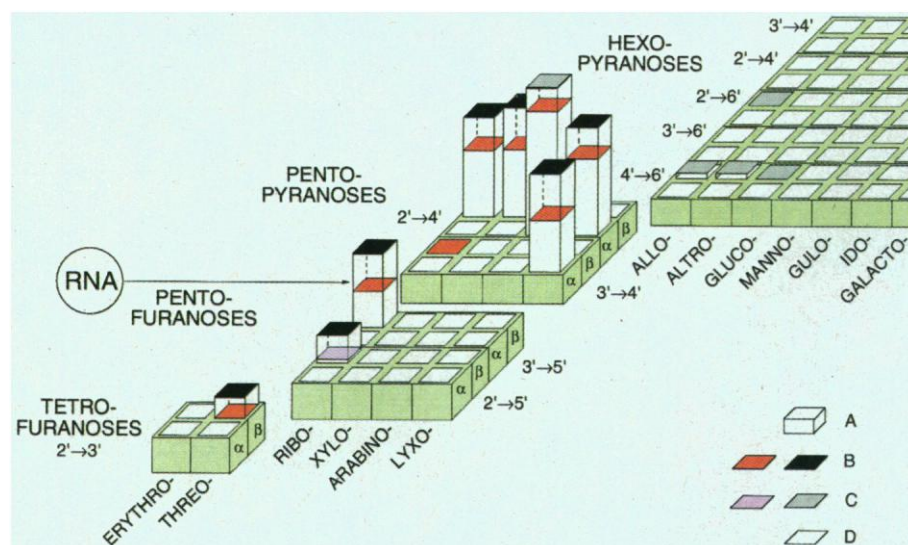


Fig. 5. Pairing-strength landscape of hexose-, pentose-, and tetrose-derived oligonucleotide systems, showing the range of the constitutional and configurational diversity of (potentially natural) alternatives of the RNA structure and giving T_m values of duplexes $A_8 \cdot T_8$ (red level) and $A_{12} \cdot T_{12}$ (black level) investigated in the ETH and TSRI laboratories. For conditions, see the caption of Fig. 3. (A) Column height unit is 10°C above 0°C (see also Fig. 3); (B) observed T_m values; (C) estimated T_m values, purple for $A_8 \cdot T_8$ and gray for $A_{12} \cdot T_{12}$; and (D) not investigated. Data are as follows: hexopyranosyl series (37), pentopyranosyl series (36, 38, 39), pentofuranosyl series (48), and tetrofuranosyl series (47).

The issue is not specific to strongly pairing nucleic acid alternatives, however. Although great strides have been made in achieving template-directed copying of base sequences in the RNA (50–52), 2',5'-RNA (15, 53), DNA (54), and p-RNA (42, 44) series and genuinely auto- and cross-catalytic replication has been accomplished for constitutionally modified, short deoxyribonucleotide sequences (55), it has not been demonstrated that any oligonucleotide system possesses the capacity for efficient and reliable nonenzymatic replication under potentially natural conditions (56). Whether, and under what conditions, such a capacity can be demonstrated experimentally remains one of the crucial criteria for judging, from a chemical point of view, whether RNA (or, for that matter, any nucleic acid alternative) could have initiated an evolutionary process. Ongoing research probing RNA's chemical phenotype by running enzyme-assisted *in vitro* evolution of RNA sequences deserves to be mentioned specifically in this context, particularly the demanding attempts to discover (in this way) ribozymes that would be able to catalyze RNA replication (57).

Not only RNA's (or an alternative's) potential for nonenzymatic replication, but also its chances for formation in an abiotic natural environment remain open to question. Whereas there is a consensus on the notion that the building blocks of RNA (sugars, purines, and pyrimidines) potentially are of prebiotic origin (51, 58) and whereas the broad chemical contours of an assembly of the RNA structure from such building blocks seem clear, convincing experimental evidence that such a process can in fact occur under potentially natural conditions is still lacking (3, 59); this is particularly true with regard to such crucial steps as nucleotide formation and phosphate activation (60). Considering the etiological importance of the issue together with the chemical experience of a vast diversity of possible paths and conditions in the generation of almost any complex organic molecule, much more experimental work on this problem is needed before a chemically reliable assessment of RNA's (or an alternative's) chances for an abiotic origin can or should be made.

In principle, a chemical etiology of nucleic acid structure has to reckon with the possibility that the RNA structure might have originated as a consequence of synthetic contingency, not as a result of synthetic variation and functional selection. It is conceivable that circumstances could have favored the selective formation of the RNA structure in preference to alternatives, be it in an abiotic or a biotic environment (61). This would imply a synthetic rather than a functional selection as the primary determinant in RNA's emergence. If this were the case, then the chemical study of nucleic acid alternatives and the discovery of efficient informational

base-pairing systems would acquire an altered importance, namely, that it would provide chemical facts in questioning the uniqueness of RNA and, implicitly, of the kind of life that is known.

Asking the central question regarding the criteria for RNA's natural selection and extending the inquiry to whether its emergence was dominated by combinatorial generation and functional selection or by synthetic contingency could mean to embark on a program of much more comprehensive chemical screening of potentially natural nucleic acid alternatives. In such a program, the selection of candidates would not be limited to RNA's structural neighborhood (62) and, moreover, would include the systematic exploration of the candidates' potential for being generated under natural conditions. Eventually, the course and the outcome of such studies would emphasize the point that the aim of an experimental etiological chemistry must be not primarily to show how life on Earth could have originated, but to provide decisive experimental evidence—through the realization of model systems in the laboratory (“artificial chemical life”)—that life can arise as a result of the organization of organic matter.

References and Notes

1. A. Einstein, in *Festschrift für Aurel Stodola*, E. Honegger, Ed. (Orell Füssli Verlag, Zürich, 1929), p. 126.
2. Etiology is the science of causes or origins (*Webster's New 20th Century Dictionary of the English Language Unabridged*, ed. 2, s.v. “etiology”).
3. RNA, not DNA, is to serve as a reference structure for choosing alternatives, because biological, as well as chemical, reasoning indicates that RNA preceded DNA in biological evolution [see, for example, G. F. Joyce and L. E. Orgel, in *The RNA World*, R. F. Gesteland, T. R. Cech, J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Plainview, NY, ed. 2, 1999), p. 49].
4. For an alternative view, see, for example, S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford Univ. Press, Oxford, 1993).
5. M. Eigen, *Naturwissenschaften* **58**, 465 (1971); *Stufen zum Leben* (Piper Verlag, München, 1987) [P. Wooley, Transl., *Steps Toward Life* (Oxford Univ. Press, Oxford, 1992)].
6. S. Lifson, *J. Mol. Evol.* **44**, 1 (1997).
7. L. E. Orgel, *J. Mol. Biol.* **38**, 381 (1968).
8. F. H. C. Crick, *ibid.*, p. 367; C. R. Woese, in *The Genetic Code*, H. O. Halverson, H. L. Roman, E. Bell, Eds. (Harper & Row, New York, 1967), p. 179; D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 217 (1967).
9. H. Kuhn, *Angew. Chem. Int. Ed. Engl.* **11**, 798 (1972); H. Kuhn and J. Waser, *ibid.* **20**, 500 (1981).
10. K. Kruger et al., *Cell* **31**, 147 (1982); C. Guerrier-Takada et al., *ibid.* **35**, 849 (1983); see also N. R. Pace and T. L. Marsh, *Origins Life Evol. Biosphere* **16**, 97 (1985); R. Lewin, *Science* **231**, 545 (1986).
11. D. A. Usher, *Nature New Biol.* **235**, 207 (1972).
12. ——— and A. H. McHale, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1149 (1976).
13. D. A. Usher, *Science* **196**, 311 (1977).
14. M. M. Dhingra and R. H. Sarma, *Nature* **272**, 798 (1978); H. Sawai, J. Seki, H. Ozaki, *J. Biomol. Struct. Dyn.* **13**, 1043 (1996).
15. T. P. Prakash, C. Roberts, C. Switzer, *Angew. Chem. Int. Ed. Engl.* **36**, 1522 (1997).
16. In a pioneering study, U. Sequin [*Experientia* **29**, 1059 (1973)] proposed that DNA strands composed of α nucleosides should also be able to undergo base pairing.
17. R. Kierzek, L. He, D. H. Turner, *Nucleic Acids Res.* **20**, 1685 (1992).
18. J. P. Dougherty, C. J. Rizzo, R. Breslow, *J. Am. Chem. Soc.* **114**, 6254 (1992); T. L. Sheppard and R. Breslow, *ibid.* **118**, 9810 (1996).
19. H. Hashimoto and C. Switzer, *ibid.* **114**, 6255 (1992); T. P. Prakash, K.-E. Jung, C. Switzer, *Chem. Commun.* **1996**, 1793 (1996).
20. J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **343**, 33 (1990); C. Y. Switzer, S. E. Moroney, S. A. Benner, *Biochemistry* **32**, 10489 (1993).
21. B. Bhat, Neelima, N. J. Leonard, H. Robinson, A. H. Wang, *J. Am. Chem. Soc.* **118**, 3065 (1996); C. Roberts, R. Bandaru, C. Switzer, *ibid.* **119**, 4640 (1997); F. Seela and A. Melenewski, *Eur. J. Org. Chem.* **1999**, 485 (1999).
22. R. Krishnamurthy et al., *Angew. Chem. Int. Ed. Engl.* **35**, 1537 (1996).
23. F. H. Westheimer, *Science* **235**, 1173 (1987).
24. D. H. Turner and P. C. Bevilacqua, in *The RNA World*, R. F. Gesteland and J. F. Atkins, Eds. (Cold Spring Harbor Laboratory Press, Plainview, NY, 1993), p. 447.
25. A. Eschenmoser, *Verh. Ges. Dtsch. Naturforsch. Aerzte* **116**, 135 (1990); A. Eschenmoser and E. Loewenthal, *Chem. Soc. Rev.* **21**, 1 (1992).
26. Homo-DNA's additional methylene group is positioned between carbons 1' and 2' of DNA's furanose ring, maintaining the nucleosidic character of the sugar building block. For a study of an isomeric (nonnucleosidic) homo-DNA, which, in contrast to homo-DNA, cross-pairs with RNA [see A. Van Aerschot, I. Verheggen, C. Hendrix, P. Herdewijn, *Angew. Chem. Int. Ed. Engl.* **34**, 1338 (1995)].
27. A. Eschenmoser and M. Dobler, *Helv. Chim. Acta* **75**, 218 (1992); M. Böhlinger et al., *ibid.*, p. 1416; J. Hunziker et al., *ibid.* **76**, 259 (1993); K. Groebke et al., *ibid.* **81**, 375 (1998).
28. Since then, the concept of backbone preorganization has been extensively used in the design of modified oligonucleotide backbones of high base-pairing strength in the context of “antisense” oligonucleotide chemistry [see, for example, M. Tarköy, M. Bolli, C. Leumann, *Helv. Chim. Acta* **76**, 481 (1993); R. Zou and M. D. Matteucci, *Tetrahedron Lett.* **37**, 941 (1996); S. K. Singh and J. Wengel, *Chem. Commun.* **1998**, 1247 (1998)].
29. P. Herdewijn, *Liebigs Ann.* **1996**, 1337 (1996) (review).
30. In contrast to fully hydroxylated hexoses, 2,3-di-deoxyhexoses are not considered to be potentially prebiotic natural products.
31. R. W. Fischer, thesis 9971, ETH, Zürich, 1992; A. G. Helg, thesis 10464, ETH, Zürich, 1994; K. Groebke, thesis 10149, ETH, Zürich, 1993; U. Diederichsen, thesis 10122, ETH, Zürich, 1993; R. Krishnamurthy, postdoctoral report, ETH, Zürich, 1994; see A. Eschenmoser, in *Chemical Synthesis: Gnosis to Prognosis*, C. Chatgililoglu and V. Sniekus, Eds. (Kluwer Academic, Dordrecht, Netherlands, 1996), p. 293.
32. G. Otting et al., *Helv. Chim. Acta* **76**, 2701 (1993).
33. R. Hammer, postdoctoral report, ETH, Zürich, 1992; C. Miculka, postdoctoral report, ETH, Zürich, and University of Frankfurt, Frankfurt, 1995.
34. Experimental confirmation would be required to determine whether this conclusion is also valid for other hexopyranosyl-(4'→6') oligonucleotide systems, as model considerations suggest.
35. S. Pitsch, S. Wendeborn, B. Jaun, A. Eschenmoser, *Helv. Chim. Acta* **76**, 2161 (1993); S. Pitsch et al., *ibid.* **78**, 1621 (1995).
36. M. Beier, F. Reck, T. Wagner, R. Krishnamurthy, A. Eschenmoser, *Science* **283**, 699 (1999).
37. For the (presumably) strongest artificial Watson-Crick base-pairing system, see A. A. Koshkin et al., *J. Am. Chem. Soc.* **120**, 13252 (1998).
38. O. Jungmann, H. Wippo, M. Stanek, H. Huynh, R. Krishnamurthy, A. Eschenmoser, in preparation.
39. F. Reck, H. Wippo, R. Kudick, M. Bolli, G. Ceulemans, R. Krishnamurthy, A. Eschenmoser, in preparation. The difference in base-pairing capability of the (L)- α -lyxopyranosyl- versus the (D)- β -ribofuranosyl-(3'→4') systems must be related to the 3',4'-dixial conformation of the phosphodiester bridge in the lyxo isomer. This

- observation led to the study of the (L)- α -threofuranosyl-(2'→3') oligonucleotide system (Fig. 1, bottom right), which, according to recent observations, is in fact a weak pairing system [see (47)].
40. I. Schlönvogt et al., *Helv. Chim. Acta* **79**, 2316 (1996).
 41. R. Micura, R. Kudick, S. Pitsch, A. Eschenmoser, *Angew. Chem. Int. Ed. Engl.* **38**, 680 (1999); R. Micura, M. Bolli, N. Windhab, A. Eschenmoser, *ibid.* **36**, 870 (1997).
 42. M. Bolli, R. Micura, S. Pitsch, A. Eschenmoser, *Helv. Chim. Acta* **80**, 1901 (1997).
 43. M. Renz, R. Lohrmann, L. E. Orgel, *Biochim. Biophys. Acta* **240**, 463 (1971); see also (12).
 44. R. Micura, M. Bolli, A. Eschenmoser, unpublished work.
 45. M. Bolli, R. Micura, A. Eschenmoser, *Chem. Biol.* **4**, 309 (1997).
 46. The data (except those for the hexose series) were determined for duplexes containing thymine instead of uracil as a nucleobase (due to the protection group strategy used in the synthesis of the pentopyranosyl members).
 47. K.-U. Schöning, P. Scholz, A. Eschenmoser, unpublished work.
 48. T. Vivemore, S. Pitsch, A. Eschenmoser, unpublished work.
 49. The T_m -value for the $r(A_a) \cdot r(T_b)$ -duplex given in table 1 and figure 3 of (36) mistakenly refers to a concentration of 3 μ M instead of the standard 10 μ M.
 50. L. E. Orgel, *Nature* **358**, 203 (1992).
 51. ——— and R. Lohrmann, *Acc. Chem. Res.* **7**, 368 (1974); L. E. Orgel, *ibid.* **28**, 109 (1995); T. Wu and L. E. Orgel, *J. Am. Chem. Soc.* **114**, 317 (1992); *ibid.*, p. 5496; *ibid.*, p. 7963.
 52. M. Kurz, K. Göbel, C. Hartel, M. W. Göbel, *Angew. Chem. Int. Ed. Engl.* **36**, 842 (1997).
 53. G. Ertem and J. P. Ferris, *Nature* **379**, 238 (1996).
 54. T. Li and K. C. Nicolaou, *ibid.* **369**, 218 (1994); K. D. James and A. D. Ellington, *Chem. Biol.* **4**, 595 (1997).
 55. G. v. Kiedrowski, *Angew. Chem. Int. Ed. Engl.* **25**, 932 (1986); D. Sievers and G. v. Kiedrowski, *Chem. Eur. J.* **4**, 629 (1998); A. Luther, R. Brandsch, G. v. Kiedrowski, *Nature* **396**, 245 (1998).
 56. Of great interest in the etiological context, however, are recent demonstrations that, in contrast to previous belief, auto- and cross-catalytic molecular replication with high turnover numbers can occur in the oligopeptide series [D. H. Lee, J. R. Granja, J. A. Martinez, K. Severin, M. R. Ghadiri, *Nature* **382**, 525 (1996); D. H. Lee, K. Severin, Y. Yokobayashi, M. R. Ghadiri, *ibid.* **390**, 591 (1997); see also S. Yao, I. Ghosh, R. Zutshi, J. Chmielewski, *ibid.* **396**, 447 (1998)]. For work on autocatalytic replication in other chemical systems, see T. Tjirikua, P. Ballester, J. Rebek Jr., *J. Am. Chem. Soc.* **112**, 1249 (1990); E. A. Wintner, M. M. Conn, J. Rebek Jr., *ibid.* **116**, 8877 (1994); K. Soai, T. Shibata, H. Morioka, K. Choji, *Nature* **378**, 767 (1995).
 57. J. A. Doudna and J. W. Szostak, *Nature* **339**, 519 (1989); A. A. Beaudry and G. F. Joyce, *Science* **257**, 635 (1992); E. H. Eklund and D. P. Bartel, *Nature* **382**, 373 (1996).
 58. J. Oró, *Nature* **191**, 1193 (1961); J. P. Ferris and L. E. Orgel, *J. Am. Chem. Soc.* **88**, 1074 (1966); J. P. Ferris and W. J. Hagan Jr., *Tetrahedron* **40**, 1093 (1984) (review).
 59. G. F. Joyce, *Nature* **338**, 217 (1989).
 60. On the other hand, substantial progress has recently been made with regard to the formation of (3'→5')/(2'→5')-ribofuranosyl oligonucleotide mixtures by mineral-assisted oligocondensation of activated 5' mononucleotides [J. P. Ferris and G. Ertem, *J. Am. Chem. Soc.* **115**, 12270 (1993); J. P. Ferris, A. R. Hill Jr., R. Liu, L. E. Orgel, *Nature* **381**, 59 (1996); (53)].
 61. Of possible relevance in this context are the preference for ribose formation in kinetically controlled aldolization reactions [D. Müller et al., *Helv. Chim. Acta* **73**, 1410 (1990); G. Zubay, *Origins Life Evol. Biosphere* **28**, 13 (1998)] and the fixation of the furanose form of ribose by preferential functionalization of the sterically least hindered primary hydroxyl group [B. Bennua-Skalmowski, K. Krolkiewicz, H. Vorbrüggen, *Tetrahedron Lett.* **36**, 7845 (1995); J. Wengel and C. Scheuer-Larsen, *Bioorg. Med. Chem. Lett.* **7**, 1999 (1997)] or by annelation of the ribofuranose ring by cis-fused five-membered ring α -diol derivatives, for example, cyclophosphates (C. Spinner, S. Pitsch, R. Krishnamurthy, A. Eschenmoser, unpublished work).
 62. Although the numerous nucleic acid analogs synthesized recently in antisense oligonucleotide chemistry [for reviews, see J. Hunziker and C. Leumann, in *Modern Synthetic Methods*, B. Ernst and C. Leumann, Eds. (Verlag Helvetica Chimica Acta, Basel, Switzerland, 1995), vol. 7, p. 331; B. Hyrup and P. E. Nielsen, *Bioorg. Med. Chem. Lett.* **4**, 5 (1996); (29)] are an additional illustration of the remarkably broad spread of the nucleobase-pairing potential among organic oligomer structures, the backbones of most of these systems do not correspond to the criteria adopted here for potentially natural nucleic acid alternatives. In a broader etiological context, however, pairing systems such as Nielsen's "polyamide nucleic acid" [P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **254**, 1497 (1991); M. Koppitz, P. E. Nielsen, L. E. Orgel, *J. Am. Chem. Soc.* **120**, 4563 (1998)] or amino acid-based systems of the type described by, for example, U. Diederichsen [Angew. Chem. Int. Ed. Engl. **35**, 445 (1996)] or M. Fujii et al. [Chem. Commun. **1998**, 717 (1998)] may have to become part of the range of structure types that must be screened with respect to their accessibility under natural conditions.
 63. The author is deeply indebted to R. Lerner (TSRI, La Jolla) and to R. Hütter (ETH, Zürich) for their support. I thank my present or former collaborators O. Jungmann, H. Wippo, M. Stanek, H. Huynh, R. Kudick, F. Reck, G. Ceulemans, and M. Bolli at TSRI; K.-U. Schöning, P. Scholz, T. Vivemore, S. Pitsch, and R. Micura at ETH for providing unpublished results mentioned in this article; and R. M. Wolf (Novartis Pharma AG, Basel, Switzerland) for producing Fig. 4. The other figures were drawn by R. Krishnamurthy, to whom I express my appreciation for his assistance in operating the La Jolla branch of my research group. I thank S. Weinberg (University of Texas, Austin) for the source of the Einstein quote; I also thank colleagues who helped with the translation. Finally, I am grateful to C. Wintner (Haverford College, Haverford, PA), who critically read the manuscript, made valuable suggestions, and improved the English.

REVIEW

Phylogenetic Classification and the Universal Tree

W. Ford Doolittle

From comparative analyses of the nucleotide sequences of genes encoding ribosomal RNAs and several proteins, molecular phylogeneticists have constructed a "universal tree of life," taking it as the basis for a "natural" hierarchical classification of all living things. Although confidence in some of the tree's early branches has recently been shaken, new approaches could still resolve many methodological uncertainties. More challenging is evidence that most archaeal and bacterial genomes (and the inferred ancestral eukaryotic nuclear genome) contain genes from multiple sources. If "chimerism" or "lateral gene transfer" cannot be dismissed as trivial in extent or limited to special categories of genes, then no hierarchical universal classification can be taken as natural. Molecular phylogeneticists will have failed to find the "true tree," not because their methods are inadequate or because they have chosen the wrong genes, but because the history of life cannot properly be represented as a tree. However, taxonomies based on molecular sequences will remain indispensable, and understanding of the evolutionary process will ultimately be enriched, not impoverished.

The impulse to classify organisms is ancient, as is the desire to have classification reflect the "natural order." Before Darwin, biologists thought that God or some other eternal principle

created that order (1, 2). After Darwin (3), they knew the ordering principle to be shared descent from an ever more limited number of common ancestors (Fig. 1), back to the last common

ancestor of all living things. A phylogenetic classification is thus the only natural one, and it should be "inclusively hierarchical" (1): Each species should be part of one and only one genus, each genus should be part of one and only one family, and so forth.

Much of modern phylogenetics is molecular phylogenetics. Microbial phylogeneticists in particular depend on molecular sequence characters, because prokaryotes (Bacteria and Archaea) offer relatively little in the way of complex morphology and behavior. Beyond this practical consideration is the understanding that molecular sequences define, in the words of Zuckerkandl and Pauling, "the essence of the organism"—not only do genes reveal the phy-

Canadian Institute for Advanced Research, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada. E-mail: ford@is.dal.ca