Toward Safe Genetically Modified Organisms through the Chemical Diversification of Nucleic Acids

by **Piet Herdewijn**[a]) and **Philippe Marlière**[b])

[a]) Laboratory for Medicinal Chemistry, Rega Institute for Medical Research, Minderbroedersstraat 10, B-3000 Leuven

[b]) Isthmus Sarl, 31 rue Saint Amand, F-75015 Paris

It is argued that genetic proliferation should be rationally extended so as to enable the propagation *in vivo* of additional types of nucleic acids (XNA for 'xeno-nucleic acids'), whose chemical backbone motifs would differ from deoxyribose and ribose, and whose polymerization would not interfere with DNA and RNA biosynthesis. Because XNA building blocks do not occur in nature, they would have to be synthesized and supplied to cells which would be equipped with an appropriate enzymatic machinery for polymerizing them. The invasion of plants and animals with XNA replicons can be envisioned in the long run, but it is in microorganisms, and more specifically in bacteria, that the feasibility of such chemical systems and the establishment of genetic enclaves separated from DNA and RNA is more likely to take place. The introduction of expanded coding through additional or alternative pairing will be facilitated by the propagation of replicons based on alternative backbone motifs and leaving groups, as enabled by XNA polymerases purposefully evolved to this end.

Introduction. – A seemingly inexhaustible source of innovation and evolution is provided by the chemical capability to synthesize DNA oligomers coupled with enzymatic procedures for gene assembly and amplification. The implantation of the resulting synthetic DNA constructs in chromosomes and episomes of microorganisms and plants for application purposes in agriculture, energetics, and chemistry has triggered the proliferation of artificial genetic sequences of an ever expanding scope. These artificial sequences are viewed by some as a threat to wild ecosystems and even suspected to lead astray the evolution of terrestrial life [1].

The first nucleic acid sequence designed and constructed to be propagated *in vivo* took the form of a synthetic gene of 77 bases encoding a suppressor tRNA that was inserted, replicated, and expressed in yeast cells [2]. This was followed by the assembly of a recombinant plasmid enabling the production of a foreign protein in *E. coli* cells, dehydroquinase from *Neurospora crassa*, through the *in vitro* use of restriction endonucleases and DNA ligase [3]. The construction of artificial eukaryotic chromosomes of minimal size followed, first in yeast [4] and later in mice [5]. The advent of directed mutagenesis [6] and the elaboration of amplification techniques [7] through the judicious use of synthetic oligodeoxynucleotides provided potent and versatile protocols for genetic reprogramming of organisms [8] and directing their evolution [9][10]. Recently, an entire bacterial chromosome of 582 970 base pairs was synthesized through stages of hierarchical assembly in prokaryotic and then eukaryotic artificial chromosomes [11]. This synthetic chromosome is now expected to take over the genetic command of a bacterial cell *Mycoplasma genitalium*.

It is currently difficult to conceive biological experiments of scientific and industrial relevance without genetic constructions resorting to DNA oligonucleotides. The chemical synthesis and the enzymatic amplification, assembly, and recombination of artificial genetic sequences for industrial purposes have become routine to the extent that transformation of ecosystems through the dissemination of synthetic genetic constructs (generically and misleadingly referred to as transgenes) is now a major topic of debate between the lay public and experts [12]. Indeed, the risks of genetic pollution cannot be overlooked, considering the uniformity of genetic alphabets, the universality of the genetic code, and the ubiquity of genetic interchanges between domesticated and wild species. Therefore, technologies for preventing or restricting genetic cross-talk between natural species and the artificial biodiversity needed for scientific and industrial progress should be designed and deployed to anticipate this challenge. Here, we argue that nucleic acid proliferation should be rationally extended so as to enable the propagation *in vivo* of additional types of nucleic acids (XNA for 'xeno-nucleic acids'), whose chemical backbone motif would differ from deoxyribose and ribose, and whose polymerization would not interfere with DNA and RNA biosynthesis.

Because XNA building blocks do not occur in nature, they would have to be synthesized and supplied to cells which should be equipped with an appropriate enzymatic machinery for polymerizing them. The genetic information conveyed in XNA would thus be expected to vanish, if it could be ensured that its precursors could not be formed in existing metabolism and food chains. As a counterpart for the unavailability of such unnatural nutrients, the controlled propagation of XNA polymers would require the uptake of XNA precursors by reprogrammed host cells as well as specific polymerization enzymes.

Despite the empirical character of biological engineering and the vast evolutionary potential of natural biodiversity, we surmise that human health and natural ecosystems will be more safely preserved by embodying genetic instructions in artificial nucleic material, *i.e.*, in replicons chemically separate from the support on which natural selection has acted so far. Natural nucleic acids originate by enzymatic polymerization of activated metabolic precursors linking a backbone motif to a nucleobase on one hand and to a leaving group on the other hand, *e.g.*, 5′-phosphorylated D-ribose to the N(9) of adenine and to the O-atom of pyrophosphate in the RNA precursor ATP, and 5′-phosphorylated 2′-deoxy-D-ribose to the N(1) of thymine and to the O-atom of pyrophosphate in the DNA precursor dTTP. Numerous variations have been introduced in the structure of nucleobases, culminating in the elaboration of pairing schemes that respect the *Watson–Crick* base-pair geometry and can be combined with A:T and G:C pairs [13], and in the development of base-pair geometries separate from *Watson–Crick*'s [14]. Variant backbone motifs whose polymers adopt stable double-stranded structures akin to DNA and RNA have been the subject of less systematic explorations, and activation groups deviating from pyrophosphate even less.

Nevertheless, we believe that it is the *in vivo* creation of moieties alternative to the backbone motifs phosphoribose and phosphodeoxyribose, and also to the leaving group pyrophosphate that will lead to the fastest advances, and that the introduction of expanded coding through additional or alternative pairing will be facilitated by the propagation of replicons based on alternative backbone motifs and leaving groups.

The invasion of plants and animals with XNA replicons can be envisioned in the long run, but it is in microorganisms, and more specifically in bacteria, that the feasibility of such chemical systems and the establishment of genetic enclaves separated from DNA and RNA is more likely to take place. Bacteriophages featuring modified bases could be considered as natural forays toward additional types of nucleic acids.

Known base modifications, incorporated prior to polymerization, have been all found to be linked to the DNA backbone [15], and consist of close variations of the A:T and G:C pairing structures, from which they deviated, presumably to escape restriction endonucleases in host bacteria [16]. So far, only the four ribonucleoside triphosphates of A, U, C, and G, as well as the eight deoxyribonucleoside triphosphates of A, T, G, C, U, 5-(hydroxymethyl)uracil, 5-(hydroxymethyl)cytosine, and 2-aminoadenine are thought to serve as substrates for nucleic acid polymerization in nature. Backbone diversification of nucleic acids thus seems entirely alien to extant living species, which suggests that the evolutionary barrier has remained too high for spontaneous innovation and natural selection to overcome.

If valid, this postulate should incite synthetic chemists and geneticists to take up the challenge, diversify nucleic acid scaffolds *in vivo*, and hence access a safer level of informational transactions in engineered life forms. These artificial constructs would be known structurally as 'xeno-nucleic acids' and functionally as 'orthogonal episomes' (*Fig. 1*).

Specifications for an Orthogonal Episome. – Briefly stated, the specifications of an orthogonal episome are as follows:

I) The chemical nature of its backbone must differ from those of DNA and RNA in at least one of its two complementary strands.

II) It must carry sequences of at least two different nucleobases, either belonging to the canonical A:T and G:C pairs, or to other pairs that can be combined with them.

III) It should be polymerized by a dedicated polymerase from activated precursors that do not interfere with DNA and RNA synthesis of the host cell.

IV) It should convey information for at least one selectable function indispensible for growth of the host cell in specific culture media.

Several additional options can be envisaged. The orthogonal episome could be devised to act as a template that is transcribed into a functional RNA (mRNA, tRNA, or rRNA) from regular rNTPs normally present in cells by a dedicated RNA polymerase. Alternatively, it could act as a messenger that is translated by a specialized ribosome from regular aminoacyl-tRNAs into a functional protein. A third possibility is that it could directly catalyze a metabolic reaction vital to the host cell. Its activated precursors could be derived from metabolic intermediates of the host cell. Eventually, the orthogonal episome could be elaborated to encode the genes required for its replication, recombination, transcription, translation, or supply of metabolic precursors.

Fourteen different experimental scenarios combining four types of replication duplexes and seven types of selection modalities are listed in the *Table*. Of the four types of replication duplexes, XNA:DNA is the least attractive, because the risk of

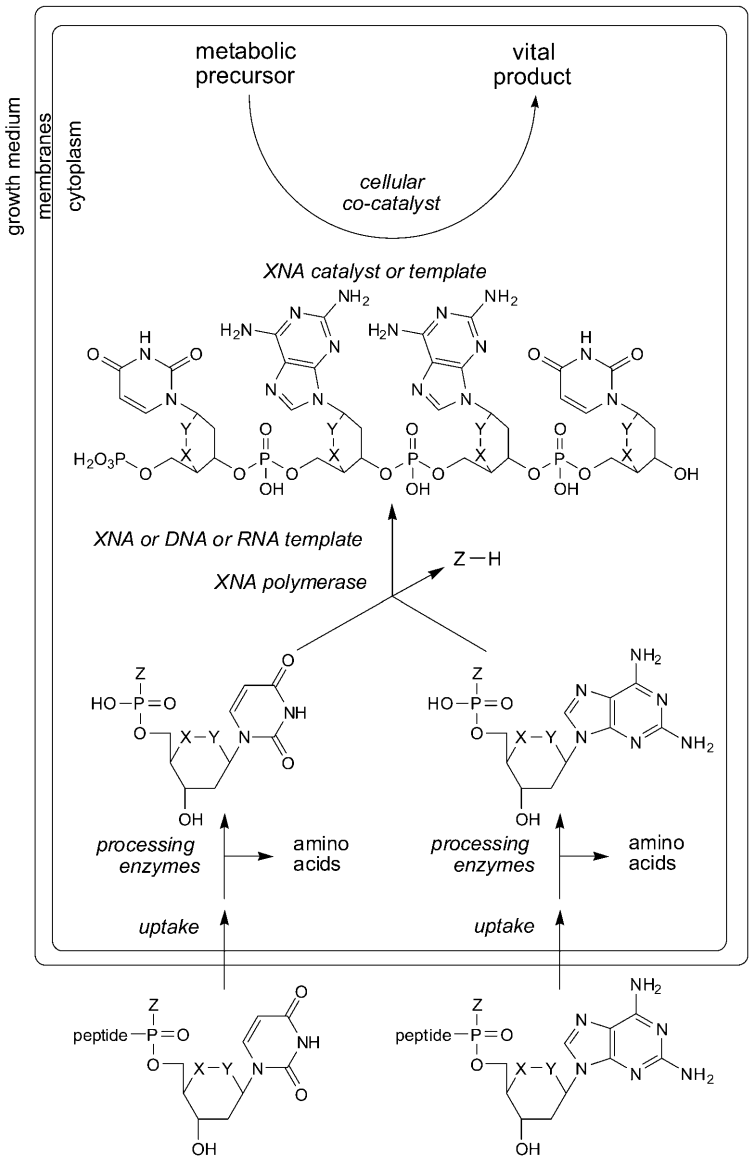

Fig. 1. *General scheme for selection of XNA biosynthesis* in vivo. The formulae and symbols used for nucleobases (uracil and diaminopurine), leaving group (Z), backbone motif (six-membered heterocycles with X and Y variables), and import vectors (peptide) are arbitrary and based on already existing constructions. At least two nucleobases are required for propagating *in vivo* the informational biosynthesis of XNA sequences. Propagation through replication of XNA : XNA duplexes would further imply that, at least, a couple of complementary nucleobases are incorporated, as depicted. The vital product could correspond to any metabolite consisting of, or that can be converted into an indispensible building block of the host cell, and that the host cell is unable to synthesize. Alternatively, the vital product could be an essential RNA, the precursor substrates then corresponding to ribonucleoside triphosphates and the cellular co-catalyst to an RNA polymerase capable of taking XNA as template. Alternatively, the vital product could be an essential protein, the precursor substrates then corresponding to aminoacyl-tRNAs and the cellular co-catalyst to a ribosome capable of translating a messenger XNA.

illegitimate recombination of the DNA strand with the host genome, including plasmids, cannot be ruled out.

Direct translation of messenger XNA templates into proteins whose function can be selected *in vivo* by specialized versions of ribosomes is tantalizing in the long term. However, because of the complexity of ribosome assembly and function, this scheme does not seem amenable to straightforward selection protocols *in vivo* or *in vitro*. An additional difficulty would result from the opposite templating directions in translation (5′ to 3′) and in transcription and replication (3′ to 5′), as exemplified by the intricate regulation of RNA virus cycles (Qβ) [17].

Resorting to the smallest possible number of experimental steps of directed evolution, it would appear that the best case scenarios in the *Table* would consist of:

a) replicating XNA : XNA duplexes with one of the XNA strands acting as a metabolic enzyme, *i.e.*, a 'xenozyme';
b) replicating XNA : RNA heteroduplexes with the RNA strand acting as a metabolic ribozyme, or a tRNA or rRNA essential for the host cell survival;
c) transcribing a chromosomal or plasmidic DNA:DNA replicon into an XNA single-strand which would act as a metabolic xenozyme.

In scenario *a*, only one replicase enzyme would need to be evolved for reproducing the XNA : XNA duplex, *i.e.*, an XNA-dependent XNA polymerase. The same replicase could be used for PCR amplification and selection of catalytic XNA aptamers destined to function *in vivo* as xenozymes, once a reliable metabolic reaction and appropriate bacterial mutants lacking it have been validated.

In scenario *b*, two polymerase enzymes would have to be evolved for replicating the XNA : RNA heteroduplex, a transcriptase and a reverse transcriptase, *i.e.*, an XNA-dependent RNA polymerase and an RNA-dependent XNA polymerase.

Both scenarios *a* and *b* correspond to proliferation modalities in which at least one XNA strand serves as template. Such is not the case in scenario *c* in which DNA serves as template for XNA, but XNA does not direct the polymerization of any nucleic acid. XNA would thus participate in the phenotype of the host cell permitting its own synthesis, yet it would not create an informational enclave propagating its genotype. This scenario could nevertheless provide a bridgehead for the launching of *bona fide* XNA replicons in further steps of elaboration. It would also enable the *in vivo* selection of uptake systems for XNA precursors and of diversified leaving groups.

The big advantage of scenario *c* is that it requires evolution of only a DNA-dependent XNA polymerase. It would benefit from the intrinsic ability of transcriptases to prime polymerization from a DNA signal, *i.e.*, a promoter, to end polymerization when another DNA signal is reached, *i.e.*, a terminator, and to disrupt the association of the polymerized XNA transcript from the DNA template.

A crucial issue in the implementation of all scenarios will be the separation of XNA strands from replication or transcription duplexes. Known possibilities elaborated by natural RNA viruses to overcome this hurdle include heteroduplex-specific RNases like RNase H, helicases, and the formation of intrastrand secondary structures. Presumably, these functional expedients could be transposed in the XNA world, by evolving XNA-heteroduplex-specific RNases, XNA helicases, and XNA stems and loops, at the cost of painstaking additional experimentation. Alternatively, backbone

Table. *Divide-and-Conquer Plan for Propagating XNA in Bacterial Cells.* Each block lists the series of experimental tasks to be carried out in order to reach the stage of stable selection *in vivo* (n.a. = not applicable)

In vitro selection modality	Replication duplex			
	XNA : XNA	XNA : RNA	XNA : DNA	DNA : DNA
	XNA Generation process			
	Replication	Transliteration of RNA template	Transliteration of DNA template	Transcription of DNA genes from promoter
XNA Strand as metabolic enzyme	• select *in vitro* an XNA aptamer • evolve XNA-dependent XNA polymerase • initiate and separate complementary strands • import *in vivo* XNA precursors	• select *in vitro* an XNA aptamer • evolve RNA-dependent XNA polymerase • evolve XNA-dependent RNA polymerase • evolve heteroduplex specific RNase • initiate complementary strands • import *in vivo* XNA precursors	• select *in vitro* an XNA aptamer • evolve DNA-dependent XNA polymerase • evolve XNA-dependent DNA polymerase • evolve heteroduplex-specific DNase • initiate complementary strands • prevent DNA recombination with host genome • import *in vivo* XNA precursors	• select *in vitro* an XNA aptamer • evolve DNA-dependent XNA polymerase • import *in vivo* XNA precursors
RNA Strand as metabolic ribozyme or tRNA or rRNA	n.a.	• select *in vitro* an RNA aptamer • evolve RNA-dependent XNA polymerase • evolve XNA-dependent RNA polymerase • initiate and separate complementary strands • import *in vivo* XNA precursors	n.a.	n.a.
DNA Strand as metabolic enzyme	n.a.	n.a.	• select *in vitro* a DNA aptamer • evolve DNA-dependent XNA polymerase	n.a.

Table (cont.)

In vitro selection modality	Replication duplex			
	XNA : XNA	XNA : RNA	XNA : DNA	DNA : DNA
	XNA Generation process			
	Replication	Transliteration of RNA template	Transliteration of DNA template	Transcription of DNA genes from promoter
			• evolve XNA-dependent DNA polymerase • initiate and separate complementary strands • prevent DNA recombination with host genome • import *in vivo* XNA precursors	
XNA Strand as RNA transcription template	• evolve XNA-dependent XNA polymerase • evolve XNA-dependent RNA polymerase • initiate and separate complementary strands • import *in vivo* XNA precursors	n.a.	• evolve DNA-dependent XNA polymerase • evolve XNA-dependent DNA polymerase • evolve XNA-dependent RNA polymerase • evolve heteroduplex-specific DNase • initiate complementary strands • prevent DNA recombination with host genome • import *in vivo* XNA precursors	n.a.
DNA Strand as RNA transcription template	n.a	n.a	• evolve DNA-dependent XNA polymerase • evolve XNA-dependent DNA polymerase • initiate and separate complementary strands	n.a

Table (cont.)

In vitro selection modality	Replication duplex			
	XNA : XNA	XNA : RNA	XNA : DNA	DNA : DNA
	XNA Generation process			
	Replication	Transliteration of RNA template	Transliteration of DNA template	Transcription of DNA genes from promoter
			• prevent DNA recombination with host genome • import *in vivo* XNA precursors	
XNA Strand as translation template	• evolve XNA-dependent XNA polymerase • evolve specialized ribosome • initiate and separate complementary strands • prevent ribosome polymerase collision • import *in vivo* XNA precursors	• evolve RNA-dependent XNA polymerase • evolve XNA-dependent RNA polymerase • evolve specialized ribosome • initiate and separate complementary strands • prevent ribosome–polymerase collision • import *in vivo* XNA precursors	• evolve DNA-dependent XNA polymerase • evolve XNA-dependent DNA polymerase • evolve specialized ribosome • evolve heteroduplex-specific DNase • initiate complementary strands • prevent DNA recombination with host genome • import XNA precursors	• evolve DNA-dependent XNA polymerase • evolve specialized ribosome • import *in vivo* XNA precursors
RNA strand as translation template	n.a.	• evolve RNA-dependent XNA polymerase • evolve XNA-dependent RNA polymerase • initiate and separate complementary strands • prevent ribosome–polymerase collision • import *in vivo* XNA precursors	n.a.	n.a.

motifs and nucleobases could be chosen so as to destabilize XNA homoduplexes or heteroduplexes and avoid replication dead-ends.

Diversification of the Backbone Motif. – The backbone motif of a third type of nucleic acid (XNA) should chemically resemble the natural polymers but differ structurally enough from DNA and RNA so that its functions and biosynthesis could be uncoupled from those of natural polymers, which would enable the establishment of a genetic enclave.

A prerequisite might be the potential of XNA to form helical structures that are similar if not identical to that of natural nucleic acids and to hybridize with natural nucleic acids by *Watson–Crick* base pairing. It is now well-established that several XNA backbone motifs bearing canonical bases are able to form regular homoduplexes. So far, the backbone motifs which have been validated for this purpose are anhydrohexitol (HNA) [18], threose (TNA) [19], glycerol (GNA) [20], and cyclohexene (CeNA) [21] (*Fig. 2*).

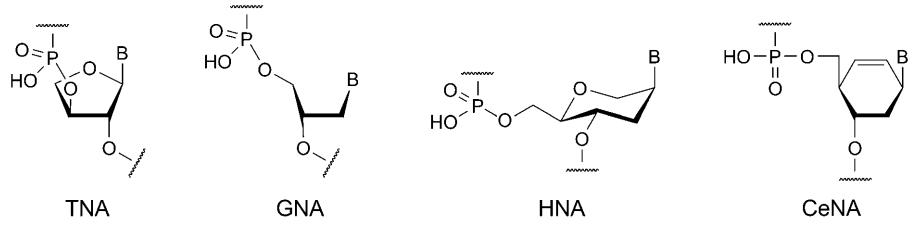

Fig. 2. *Structures of threose nucleic acid* (TNA), *glycerol nucleic acid* (GNA), *hexitol nucleic acid* (HNA), *and cyclohexenyl nucleic acids* (CeNA)

HNA was designed by inserting a CH_2 group between the ring O-atom and the anomeric C-atom of the furanose ring of a natural nucleoside. As a result, the base moiety is no longer positioned at the anomeric C-atom. A chemically and enzymatically stable nucleic acid is obtained, which is still able to convey information to and from natural nucleic acids. HNA hybridizes with DNA and RNA, and with itself [18]. An interesting observation is that the ΔT_m/modification of an HNA:RNA duplex decreases with increasing chain length, ranging from a high stabilizing effect for short duplexes to practically no stabilization/destabilization effect for long duplexes. HNA fulfills the requirement for being similar but not identical to natural nucleic acids. Its production *via* an appropriate metabolic pathway remains to be investigated. Likewise, it must be demonstrated that the structure of HNA is different enough from those of RNA and DNA to construct a fully orthogonal system.

Initial DNA-dependent HNA polymerase activity, using *Watson–Crick* base-pair fidelity and including the potential for chain elongation, has been observed for family B DNA polymerases [22] and the M184V mutant of HIV-1 reverse transcriptase [23]. Likewise, T7 RNA polymerase and terminal transferase use hexitol nucleoside triphosphates (hNTP) as substrate for HNA synthesis, although not as efficiently as the natural nucleoside triphosphates [24]. T4 DNA Ligase and T4 RNA ligase, however, do not accept HNA as 5′- nor 3′-substrate, nor as template for ligation of single-stranded DNA [24]. Marginal HNA-dependent DNA polymerase activity and HNA-dependent

HNA polymerase activity was observed for DNA polymerase I from *E. coli* and for *Thermus aquaticus* polymerase [25]. The fact that hNTPs and HNA can be used by certain polymerases as both substrate and template to a limited extent shows the potential to evolve natural polymerases for affording a *bona fide* HNA replication system.

For performing experiments on *in vivo* replication, a selection scheme was designed to assay transmission of a genetic message from HNA to a DNA plasmid by the replication enzymes in *E. coli* [25]. *E. coli* cells bearing a chromosomal deletion of the gene for thymidylate synthase were transformed with plasmids containing HNA stretches encoding the active-site region surrounding the catalytic Cys146 of the enzyme. Utilization of the HNA stretch as template was mandatory for propagation of an active gene for thymidylate synthase. Since active thymidylate synthase is absolutely required for growth of *E. coli* thyA mutants in nutrient media devoid of thymine or thymidine, this scheme can be used for the selection of strains that are able to copy HNA into DNA *in vivo*.

Prototrophic transformants of a host strain lacking an active gene for thymidylate synthase, *i.e.*, bacteria capable of growth without added thymidine, could be obtained with a plasmid construct containing two HNA codons (six nucleotides). Sequencing experiments demonstrated that the message conveyed in the HNA residues was correctly copied into DNA [25].

We could also demonstrate that the replacement of two codons of a 33mer mRNA (AUG start codon and UUC second codon) by HNA (mHNA) did not hinder the main steps of translation, as indicated by the same level of ribosomal binding of mRNA with hexitol residues under P-site conditions, and the same yield of tRNA binding to the P- and A-sites. Both peptide formation and translocation took place on mRNA containing the two-codon-long mHNA [26]. The enzymatic stability of HNA may ensure that protein synthesis will go on for a longer time than using mRNA.

Other successful examples of polymerization of sugar-modified nucleoside triphosphates using polymerases as catalyst on a DNA template (or polymerization of deoxynucleoside triphosphate on modified nucleic acids template) are those of GNA [27] and TNA [28][29]. The structures of GNA and TNA, featuring phosphorylated glycerol and phosphorylated tetrofuranose backbones, respectively, deviate more substantially from those of natural DNA than HNA.

The fourth case is that of cyclohexenyl nucleic acid (CeNA) [21]. DNA Polymerase B and HIV-1 reverse transcriptase show good DNA-dependent CeNA polymerase and CeNA-dependent DNA polymerase activity which renders it as a better starting point than HNA [30].

HNA featuring a cyclic ether and CeNA endowed with an unsaturated carbocyclic scaffold both appear as hardly accessible from common cell metabolites, if biosynthetic pathways have to be designed for producing them endogenously. By contrast, GNA and TNA appear more realistic in terms of biosynthetic accessibility from metabolites present in cells. Enabling the spontaneous emergence of XNA metabolic precursors will represent a feat of great scientific significance for the propagation of XNA-bearing hosts. In this case, drastic containment procedures will be mandatory for safely growing XNA prototrophs, *i.e.*, organisms able to synthesize episomes with deviant chemistry. Therefore, an intrinsic safety device will be provided by the requirement for xeno-

nutrients to propagate XNA in next-generation GMOs. Metabolic inaccessibility should thus be considered as an advantage, when it comes to the chemical estrangement of XNA from metabolism of extant cells and food chains of natural ecosystems. This entails that XNA precursors will have to be devised and synthesized so as to permit their uptake when they cannot be synthesized within the host cell.

A potent delivery system for nucleoside mono(di- and tri-)phosphates could make use of oligopeptide permeases, which are part of a larger group of transport systems, *i.e.*, ATP-binding cassette transporters. The periplasmic oligopeptide-binding protein is the initial receptor for the uptake of peptides by the oligopeptide permease in *Gram*-negative bacteria. The ligand-binding site shows a broad substrate specificity and accepts chemical groups of large diversity (it accepts di- to pentapeptides regardless of the identity of the side chains) [31]. The first conjugates for delivery of nucleotides that could function as substrate for the transporter have been synthesized [32]. They consist of a pyroglutamyl protected tripeptide, a lateral pyridoxal moiety, and a nucleotide loaded on a serine residue (*Fig. 3*). Following transport into the bacterial cytoplasm, the pyroglutamyl group could be deblocked by a specific aminopeptidase liberating a free terminal amino group that could be involved in the catalytic process to deliver the laterally attached nucleotide. Intracellular delivery of the nucleotide could indeed be accomplished by a pyridoxal-catalyzed elimination of the nucleotide, bound to the free amino group of the serine residue, *via* formation of a *Schiff* base.

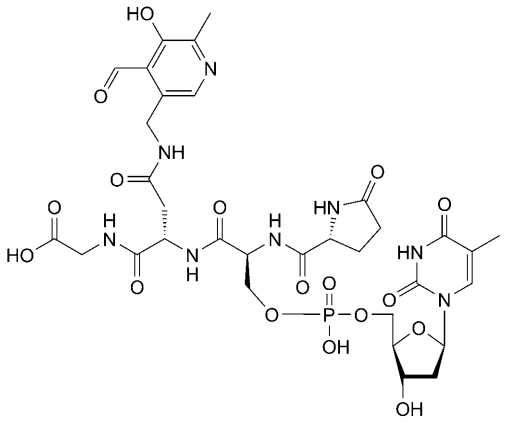

Fig. 3. *Structure of a potential delivery system for nucleotides into bacterial cells*

Diversification of the Leaving Group. – DNA and RNA are synthesized from nucleoside triphosphates through an iterative catalytic process where one nucleotide is attached (*via* its 5′-phosphate) to the 3′-hydroxy group of the growing nucleic acid chain, releasing pyrophosphate in the process. This pyrophosphate is further hydrolyzed into two phosphate molecules, thus ensuring the irreversibility of nucleic polymerization. The nucleoside triphosphates themselves are synthesized from the monophosphates through the action of nucleotide kinases and nucleoside diphosphate kinases. The entanglement of the various functions of nucleoside triphosphates can be seen as a trap when trying to propagate an artificial biopolymer such as novel types of

nucleic acid *in vivo*. It would be difficult indeed to install additional nucleoside triphosphates without interfering with DNA and RNA metabolism, cell energy supply *via* respiration, or substrate-level phosphorylation.

From a chemical standpoint, the simplest way out of this intricacy is to substitute the pyrophosphate moiety with alternative leaving groups in XNA precursors. From an informational standpoint, the use of alternative leaving groups would *ipso facto* result in genetic enclaves distinct from canonical nucleic acids without having to physically separate precursors of XNA from those of DNA and RNA. Activated nucleotides with alternative leaving groups should not be recognized by polymerases and other enzymes involved in cellular functions where nucleoside triphosphates play a role. If an alternative leaving group corresponded or could be converted to a common metabolite, they could be fitted into recycling pathways, and thus prevent the accumulation of by-products of XNA polymerization.

The ideal properties for XNA precursors with an alternative leaving group could be listed as follows:

– they should be very soluble and not too instable in H_2O;
– they should be accommodated in the active site of polymerase and react as substrate;
– they should undergo productive elongation;
– the leaving group should be released by polymerases in a mechanism-based manner;
– the leaving group should be actively degraded or recycled to common metabolites so as to render polymerization irreversible;
– the nucleotide precursor equipped with the leaving group should be taken up actively in cells.

Modification of the triphosphate moiety of natural nucleosides with the aim of discovering alternative substrates for polymerases is a research area that has been little explored. Generally, such modifications have been investigated in the search for inhibitors of enzymes such as adenylate cyclase and reverse transcriptase, or inhibitors of nucleotide binding receptors. The enzymatic synthesis of modified nucleic acids (phosphoramidates) using sugar-modified nucleoside triphosphates was pioneered by *Letsinger et al.* [33] using dNTTP as substrate and DNA polymerase I of *E. coli* as catalyst. The greatest amount of information about sugar- and triphosphate-modified nucleosides, and their incorporation into DNA-using enzymes (reverse transcriptase) can be found in the literature on potential inhibitors of HIV replication [34]. The modifications that have been investigated so far are mainly minor modifications within the triphosphate moiety itself. The most studied are methylene phosphonate [35–37], phosphoramidate [35], and thiophosphate [38–40] analogues of ATP.

The most deviant instances of a leaving group that retains the capacity to be recognized by polymerases are those of Asp-dAMP and His-dAMP (*Fig. 4*) in the case of reverse transcriptase [41][42]. L-Aspartate and L-histidine were found to be efficient pyrophosphate mimics despite the structural differences. The fact that L-Glu-dAMP is not recognized as substrate, while L-Asp-dAMP is recognized, underscores the precise geometric and spatial arrangement of functional groups that enables the formation of a tertiary complex in the active site of the enzyme.

Likewise, Asp-dGMP, Asp-dCMP, and Asp-dTMP are substrates for HIV-RT, while retaining the canonical base-pair selectivity [43]. The kinetics of incorporation are,

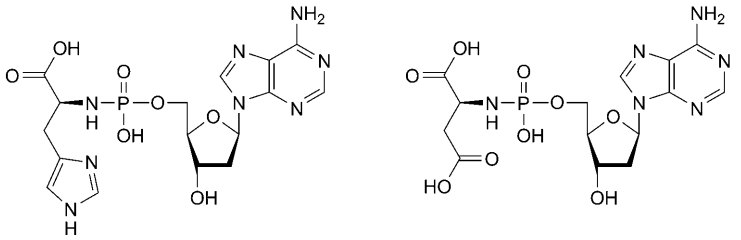

Fig. 4. *Structure of* L-*His-dAMP and* L-*Asp-dAMP*

however, much slower than that of the natural substrates [41]. Encouraging results are that the primer can be extended with up to six deoxyadenines (although stalling starts after incorporation of three nucleotides), that the incorporation reaction follows *Watson–Crick* rules and is stereoselective (L- *vs.* D-amino acids), and that the incorporation of the four building blocks, L-Asp-dAMP, L-Asp-dCMP, L-Asp-dGMP, and L-Asp-dTMP, follows *Michaelis–Menten* kinetics. That a moiety as simple and common as aspartate linked by a P–N bond can serve as alternative leaving group in the polymerization of DNA augurs well for the further elaboration of parallel biosynthetic and energetic pathways for XNA synthesis through metabolic engineering of host cells, and directed evolution of DNA and RNA polymerases. A promising aspect of chemical diversification of the leaving group is that it could be applied first to a transcription system for generating XNA. To encode functional sequences, only two different monomers would have to be introduced, presumably complementary.

Diversification of Nucleic Bases. – The main synthetic effort to explore macromolecular structures of nucleic acids not known in nature has mainly focused on diversifying pairing schemes beyond those originally proposed by *Watson* and *Crick*. The genetic alphabet has been expanded from a two base-pair system to a six base-pair system [44][45]. This approach is based on finding new H-bonding topologies. For example, a 6-amino-5-nitropyridin-2(1*H*)-one:2-aminoimidazo[1,2-*a*]triazin-4-one base pair has been proposed for synthetic-biology purposes [46] (*Fig. 5*).

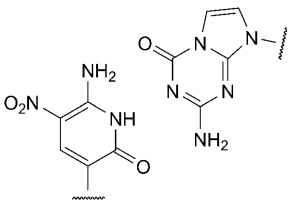

Fig. 5. *A pyDDA/puAAD system that could function as informational system* in vivo

Others have enlarged the genetic helix with expanded-size base pairs [47][48]. Another successful attempt involved the incorporation of hydrophobic nucleobase analogues [49–51]. Starting from a large screening effort using five different polymerases, additional hydrophobic base-pair systems were identified consisting of

bases built from phenyl, isoquinoline, and the pyrrolo[2,3-*b*]pyridine scaffold [52]. The base-pair system depicted in *Fig. 6* is proposed as promising for *in vivo* use.

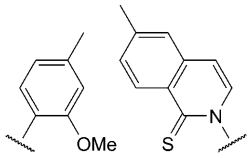

Fig. 6. *A hydrophobic base-pair system for potential use* in vivo

An additional base pair could be accommodated in XNA by providing a host cell with the corresponding activated precursors only if the replication of XNA does not involve a complementary strand made of DNA or RNA. Extensive diversification of the nucleobases might thus only fit in the XNA : XNA scenario *a* of the *Table*.

However, base modifications compatible with the natural base pairs such as diaminopurine, 5-substituted uracil, and cytosine, as well as 8-azapurines, 7-substituted deaza-adenine, and guanine would suit all scenarios. Some such modifications might be rationally introduced for modulating duplex stability, easing purification of XNA, or enhancing catalytic activities of XNA aptamers.

Backbone and leaving-group diversification thus appear more pertinent than base diversification for establishing XNA *in vivo*. Once established *in vivo*, XNA replicons will provide an opportunity to recruit alternative base pairs for protein coding or biocatalytic purposes.

Diversification of Nucleic Acid Polymerases. – Enzymatic activities that should be considered for propagating XNA *in vivo* following the various scenarios of the *Table* can be listed as follows:

– a DNA-dependent XNA polymerase, or XNA transliterase, would be needed for conveniently synthesizing XNA strands;
– an XNA-dependent DNA polymerase, or reverse XNA transliterase;
– an XNA-dependent XNA polymerase or XNA replicase;
– an XNA-dependent RNA polymerase or XNA transcriptase to copy XNA into mRNA, tRNA, or rRNA that can be used by the ribosome for protein synthesis;
– a RNA-dependent XNA polymerase to copy RNA information into XNA strands.

The XNA polymerase activities, whether templated by XNA itself, RNA, or DNA, could have the additional capability of condensing precursors with a leaving group other than pyrophosphate and to reject nucleoside triphosphates.

DNA Polymerases perform a feat of molecular recognition, incorporating the correct four different nucleotide substrates as specified by the template base with minimal error rates. As a result of this high fidelity and tight geometric control within the polymerase active site, catalysis becomes very sensitive to distortions in the primer–template–dNTP tertiary complex. This precludes or diminishes the replication of modified DNA templates, and restricts the enzymatic incorporation or replication of synthetic nucleic acids.

The majority of synthetic nucleic acids are poor substrates for enzymatic polymerization, and studies with unnatural nucleic acids are limited to short oligomers accessible by solid-phase synthesis. Construction of artificial genetic systems both *in vitro* and *in vivo* implies the development of an efficient enzymatic synthesis of the nucleic acid analogues. The efficiency and selectivity of modified nucleoside triphosphate incorporation are usually low. To facilitate the development of XNA-based replicons and episomes, it will be necessary to develop mutant polymerases able to synthesize the new polymers and to use them as templates for the propagation of information. These mutants could be obtained by rational design, activity-based library screening, and enrichment [53–55]. Activity-based selection such as compartmentalized self-replication [56][57] and phage display [58–60] could also be exploited for this purpose. Indeed, incomplete understanding of the detailed mechanism of polymerization makes rational design approaches difficult. Therefore, selection approaches and directed evolution methods have been developed [61].

Loeb and co-workers pioneered the genetic complementation approach for polymerase selection [54]. Using a similar approach, *Chelliserrykattil* and *Ellington* have selected variants of T7 RNA polymerase with an increased ability to synthesize transcripts from 2′-substituted ribonucleotides [62]. Along this experimental line, one could attempt to evolve a DNA-dependent XNA polymerase responding to promoters and terminators specified by DNA, and also enzyme variants specific for leaving groups different from pyrophosphate that can be recycled in metabolism. Phage display has been used for the selection of a polymerase function by proximal display of both primer–template duplex and polymerase on the same phage particle [58] which led to a variant of the *Stoffel* fragment that incorporates ribonucleoside triphosphates (rNTPs) with efficiencies approaching those of the wild-type enzyme for dNTP substrates.

A different assay for the evolution of polymerases is compartmentalized self-replication [57]. Polymerases and substrates are encapsulated in discrete non-communicating compartments akin to artificial cells. Individual polymerase variants are isolated in separate compartments. Each polymerase replicates only its own encoding gene to the exclusion of those in other compartments. Consequently, only genes encoding active polymerases are replicated, while inactive variants disappear from the gene pool. The more active will make proportionally more copies of their own encoding gene. This method has been shown to be a powerful method for the directed evolution of polymerases. The first step towards the generation of polymerases capable of replication of unnatural nucleic acids have already been taken using this approach [63]. Unfortunately, the molecular type synthesized in this way, a S-DNA, is not sufficiently different from DNA to be treated as an orthogonal entity.

It is mandatory for an artificial genetic system to become fully and not partially disconnected from DNA replication and RNA transcription of the host cell in order to sever links of information transmission by mistemplating and misincorporation of activated monomers.

An additional feature that would need to be evolved in a XNA polymerase would be the initiation step using a tRNA primer as for reverse transcriptases, or a protein primer as for DNA polymerase from bacteriophage phi29. Another very helpful enzyme for diversifying nucleic acids *in vivo* would be an XNA ligase, *i.e.*, a ligase that

accepts XNA as substrate to produce long stretches of XNA for making whole genes and episomes.

Conclusions. – Propagating a third type of nucleic acid in a microbial cell will require the molecular alliance of several entities (metabolic precursors, templates, and enzymes) each of which will require systematic efforts to elaborate individually.

The successful construction of at least an uptake and processing apparatus for exogenous activated precursors of minimally two XNA precursors, an XNA polymerase, and either a xenozyme (XNA-aptamer) catalyzing an essential metabolic reaction, or an XNA-dependent RNA polymerase and an XNA complement for a tRNA, an rRNA, a short mRNA, or a metabolic ribozyme (RNA-aptamer) will have to be accomplished (*Table*).

Research in fields as different as organic chemistry, evolutionary biotechnology, and microbiology will need to be coordinated and trustfully funded for several years before sustained proliferation of XNA becomes operational.

Once it is operational, an entirely new adaptive landscape will become accessible to experimental exploration through prolonged cultivation or successive selection steps for usage of deviant exogenous XNA precursors and function of XNA-encoded sequences.

The persistent requirement of vast populations of reprogrammed microorganisms for exogenous XNA precursors will establish conclusively the functional role of XNA polymers as well as the inaccessibility of metabolic reactions for endogenously producing these XNA precursors.

Such experiments should be instrumental for estimating the risks of horizontal transfer and counteracting genetic pollution in next-generation GMOs.

In the long run, we anticipate that XNA propagation *in vivo* will lend itself to evolutionary processes of exaptation unprecedented in natural history and the experimental record, by allowing the recruitment under selection of ever different nucleic monomers deviating in nucleobase pairing, leaving groups, and backbone motifs within the limits of organic synthesis.

Forging the limited arsenal of molecular machinery needed to launch such an open-ended and safe evolutionary regime seems feasible by relying on the theoretical understanding and experimental procedures of current nucleic acids research.

We are grateful to *Susan Cure* for correcting and improving the text, and to *Mia Vanthienen* for editorial help. Creative discussions with *Phil Holliger* and *Markus Schmidt* are gratefully acknowledged.

REFERENCES

[1] J. Rifkin, 'The Biotech Century', Penguin Putnam, New York, 1998.
[2] H. G. Khorana, *Pure Appl. Chem.* **1971**, *25*, 91.
[3] D. Vapnek, J. A. Hautala, J. W. Jacobson, N. H. Giles, S. R. Kushner, *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 3508.
[4] A. W. Murray, J. W. Szostak, *Nature* **1983**, *305*, 189.
[5] P. Giraldo, L. Montoliu, *Transgenic Res.* **2001**, *10*, 83.
[6] C. A. Hutchison III, S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke, M. Smith, *J. Biol. Chem.* **1978**, *253*, 6551.

[7] R. K. Saiki, S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, N. Arnheim, *Science* **1985**, *230*, 1350.
[8] E. Stolovicki, T. Dror, N. Brenner, E. Braun, *Genetics* **2006**, *173*, 75.
[9] G. Pósfai, G. Plunkett III, T. Fehér, D. Frisch, G. M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S. S. Sharma, M. de Arruda, V. Burland, S. W. Harcum, F. R. Blattner, *Science* **2006**, *312*, 1044.
[10] K. A. Powell, S. W. Ramer, S. B. Del Cardayré, W. P. Stemmer, M. B. Tobin, P. F. Longchamp, G. W. Huisman, *Angew. Chem., Int. Ed.* **2001**, *40*, 3948.
[11] D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, T. B. Stockwell, A. Brownley, D. W. Thomas, M. A. Algire, C. Merryman, L. Young, V. N. Noskov, J. I. Glass, J. C. Venter, C. A. Hutchison III, H. O. Smith, *Science* **2008**, *319*, 1215.
[12] C. M. Delude, K. W. Mirvis, J. Davidson, *Nat. Biotechnol.* **2001**, *19*, 911.
[13] J. Horlacher, M. Hottiger, V. N. Podust, U. Hubscher, S. A. Benner, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 6329.
[14] A. T. Krueger, E. T. Kool, *Chem. Biol.* **2009**, *16*, 242.
[15] R. A. Warren, *Annu. Rev. Microbiol.* **1980**, *34*, 137.
[16] S. Bron, E. Luxen, G. Venema, *J. Virol.* **1983**, *46*, 703.
[17] N. Ichihashi, T. Matsuura, H. Kita, K. Hosoda, T. Sunami, K. Tsukada, T. Yomo, *ChemBioChem* **2008**, *9*, 3023.
[18] C. Hendrix, I. Verheggen, H. Rosemeyer, F. Seela, A. Van Aerschot, P. Herdewijn, *Chem.–Eur. J.* **1997**, *3*, 110.
[19] K. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347.
[20] L. Zhang, A. Peritz, E. Meggers, *J. Am. Chem. Soc.* **2005**, *127*, 4174.
[21] J. Wang, B. Verburen, I. Luyten, E. Lescrinier, M. Froeyen, C. Hendrix, H. Rosemeyer, F. Seela, A. Van Aerschot, P. Herdewijn, *J. Am. Chem. Soc.* **2000**, *122*, 8595.
[22] K. Vastmans, S. Pochet, A. Peys, L. Kerremans, A. Van Aerschot, C. Hendrix, P. Marlière, P. Herdewijn, *Biochemistry* **2000**, *39*, 12757.
[23] K. Vastmans, M. Froeyen, L. Kerremans, S. Pochet, P. Herdewijn, *Nucleic Acids Res.* **2001**, *29*, 3154.
[24] K. Vastmans, J. Rozenski, A. Van Aerschot, P. Herdewijn, *Biochim. Biophys. Acta* **2002**, *1597*, 115.
[25] S. Pochet, P. A. Kaminski, A. Van Aerschot, P. Herdewijn, P. Marlière, *C. R. Biol.* **2003**, *326*, 1175.
[26] I. N. Lavrik, O. N. Avdeeva, O. A. Dontsova, M. Froeyen, P. A. Herdewijn, *Biochemistry* **2001**, *40*, 11777.
[27] C. H. Tsai, J. Chen, J. W. Szostak, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14598.
[28] J. C. Chaput, J. W. Szostak, *J. Am. Chem. Soc.* **2003**, *125*, 9274.
[29] V. Kempeneers, K. Vastmans, J. Rozenski, P. Herdewijn, *Nucleic Acids Res.* **2003**, *31*, 6221.
[30] V. Kempeneers, M. Renders, M. Froeyen, P. Herdewijn, *Nucleic Acids Res.* **2005**, *33*, 3828.
[31] J. R. Tame, G. N. Murshudov, E. J. Dodson, T. K. Neil, G. G. Dodson, C. F. Higgins, A. J. Wilkinson, *Science* **1994**, *264*, 1578.
[32] A. Marchand, D. Marchand, R. Busson, P. Marlière, P. Herdewijn, *Chem. Biodiversity* **2007**, *4*, 1450.
[33] R. L. Letsinger, J. S. Wilkes, L. B. Dumas, *J. Am. Chem. Soc.* **1972**, *94*, 292.
[34] A. A. Arzumanov, D. G. Semizarov, L. S. Victorova, N. B. Dyatkina, A. A. Krayevsky, *J. Biol. Chem.* **1996**, *271*, 24389.
[35] R. G. Yount, D. Babcock, W. Ballantyne, D. Ojala, *Biochemistry* **1971**, *10*, 2484.
[36] T. Myers, K. Nakamura, J. Flesher, *J. Am. Chem. Soc.* **1963**, *85*, 3292.
[37] D. B. Trowbridge, D. M. Yamamoto, G. L. Kenyon, *J. Am. Chem. Soc.* **1972**, *94*, 3816.
[38] R. S. Goody, F. Eckstein, *J. Am. Chem. Soc.* **1971**, *93*, 6252.
[39] F. Eckstein, H. Gindl, *Biochim. Biophys. Acta* **1967**, *149*, 35.
[40] F. Eckstein, R. S. Goody, *Biochemistry* **1976**, *15*, 1685.
[41] O. Adelfinskaya, M. Terrazas, M. Froeyen, P. Marlière, K. Nauwelaerts, P. Herdewijn, *Nucleic Acids Res.* **2007**, *35*, 5060.
[42] O. Adelfinskaya, P. Herdewijn, *Angew. Chem., Int. Ed.* **2007**, *46*, 4356.
[43] M. Terrazas, P. Marlière, P. Herdewijn, *Chem. Biodiversity* **2008**, *5*, 31.
[44] S. A. Benner, *Acc. Chem. Res.* **2004**, *37*, 784.

[45] C. R. Geyer, T. R. Battersby, S. A. Benner, *Structure* **2003**, *11*, 1485.
[46] Z. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, S. A. Benner, *Nucleic Acids Res.* **2007**, *35*, 4238.
[47] S. R. Lynch, H. Liu, J. Gao, E. T. Kool, *J. Am. Chem. Soc.* **2006**, *128*, 14704.
[48] H. Liu, J. Gao, S. R. Lynch, Y. D. Saito, L. Maynard, E. T. Kool, *Science* **2003**, *302*, 868.
[49] S. Matsuda, J. D. Fillo, A. A. Henry, P. Rai, S. J. Wilkens, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. G. Schultz, G. Spraggon, F. E. Romesberg, *J. Am. Chem. Soc.* **2007**, *129*, 10466.
[50] A. A. Henry, A. G. Olsen, S. Matsuda, C. Yu, B. H. Geierstanger, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 6923.
[51] A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 2336.
[52] G. T. Hwang, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 14872.
[53] J. B. Sweasy, L. A. Loeb, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 4626.
[54] M. Suzuki, D. Baskin, L. Hood, L. A. Loeb, *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9670.
[55] A. Shinkai, P. H. Patel, L. A. Loeb, *J. Biol. Chem.* **2001**, *276*, 18836.
[56] D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* **1998**, *16*, 652.
[57] F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4552.
[58] J. Jestin, P. Kristensen, G. Winter, *Angew. Chem., Int. Ed.* **1999**, *38*, 1124.
[59] M. Fa, A. Radeghieri, A. A. Henry, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 1748.
[60] G. Xia, L. Chen, T. Sera, M. Fa, P. G. Schultz, F. E. Romesberg, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 6597.
[61] R. C. Holmberg, A. A. Henry, F. E. Romesberg, *Biomol. Eng.* **2005**, *22*, 39.
[62] J. Chelliserrykattil, A. D. Ellington, *Nat. Biotechnol.* **2004**, *22*, 1155.
[63] F. J. Ghadessy, N. Ramsay, F. Boudsocq, D. Loakes, A. Brown, S. Iwai, A. Vaisman, R. Woodgate, P. Holliger, *Nat. Biotechnol.* **2004**, *22*, 755.

Received March 20, 2009