Redesigning nucleic acids

Steven A. Benner, Thomas R. Battersby, Bernd Eschgfäller, Daniel Hutter, Janos T. Kodra, Stefan Lutz, Tuncer Arslan, Daniel K. Bäschlin, Monika Blättler, Martin Egli, Christophe Hammer, Heike A. Held, Jennifer Horlacher, Zhen Huang, Birgitte Hyrup, Thomas F. Jenny, Simona C. Jurczyk, Marcel König, Ulrike von Krosigk, Michael J. Lutz, Lawrence J. MacPherson, Simon E. Moroney, Eugen Müller, Krishnan P. Nambiar, Joseph A. Piccirilli, Christopher Y. Switzer, Johannes J. Vögel, Clemens Richert, Andrew L. Roughton, Jürgen Schmidt, K. Christian Schneider, and Joseph Stackhouse

Department of Chemistry, University of Florida, Gainesville FL 32611
Department of Chemistry, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

Abstract: A research program has applied the tools of synthetic organic chemistry to systematically modify the structure of DNA and RNA oligonucleotides to learn more about the chemical principles underlying their ability to store and transmit genetic information. Oligonucleotides (as opposed to nucleosides) have long been overlooked by synthetic organic chemists as targets for structural modification. Synthetic chemistry has now yielded oligonucleotides with 12 replicatable letters, modified backbones, and new insight into why Nature chose the oligonucleotide structures that she did.

INTRODUCTION

The "standard model" of nucleic acid structure dates back to 1953 and two classic papers by Watson and Crick.[1,2] It has been little altered since. The model holds that the energy of binding of two complementary DNA or RNA (oligonucleotide) strands arises from the stacking of the hydrophobic nucleobases, while the specificity of the association arises from base pairing following two simple rules ("A pairs with T, G pairs with C"). No other class of natural products has reactivity that obeys such simple rules. Nor is it obvious how one designs a class of chemical substances that does so much so simply. Despite this chemical conundrum, and the position of nucleic acids at the center of natural product chemistry, few organic chemists have chosen to apply their synthetic skills to explore reactivity at the level of the *oligo*nucleotide. Much work had been done, of course, in making structurally modified analogs of nucleo*sides*, both in industry and academia.[3] But most organic chemists, attracted by the structural intricacies of secondary metabolites, have neglected oligonucleotides as targets for structural modification.

Some 15 years ago we began a program to fill this gap, developing synthetic organic chemistry and organic structural theory as it applies to nucleic acids in their oligomeric form. This began with one of the first two total syntheses of a gene encoding a protein,[4] and has continued with the development of structurally altered oligonucleotides. As in all organic chemistry that alters the structure of natural products, our goal has been to learn more about how DNA and RNA work. We focus here on chemistry that has modified the bases, the sugars, and the backbones of oligonucleotides.

CHANGING THE NUCLEOBASES: A GENETIC ALPHABET WITH 12 LETTERS

The Watson-Crick base pair obeys two rules of complementarity: *size complementarity* (large purines pair with small pyrimidines) and *hydrogen bonding complementarity* (donors from one nucleobase pair with acceptors from the other). With three hydrogen bonds joining the base pair, eight hydrogen bonding patterns are possible. Six of these are readily written with conventional tautomeric forms (Figure 1).[5,6] Thus, nucleic acids *could* be built from an expanded genetic alphabet, containing 12 independently replicatable building blocks instead of four, forming six base pairs instead of two, each held together by a distinct pattern of hydrogen bonds. The chemist is able to implement this vision through synthesis.

In this discussion, we designate different nucleobases using a generalized nomenclature. The prefix "py" indicates a "small" single ring heterocycle analogous to a pyrimidine, while the prefix "pu" indicates a "large" 5-6 fused heterocycle analogous to a purine. The letters "D" and "A" indicate the pattern of hydrogen bond donors and acceptors that the ring presents to the opposite strand in a Watson-Crick duplex, starting from the major groove.

Virtually every field of molecular science was recruited to make an expanded genetic alphabet a reality, including physical organic chemistry, synthesis, enzymology, molecular biology, and molecular evolution. First, a heterocycle was chosen to carry each of the non-standard hydrogen bonding patterns. Much work was done to optimize the reactivity of the heterocycle to address issues of tautomerism, acid/base reactivity, and stability. For example, problems with excessive basicity (pyDAD)[7,8] and facile epimerization (pyDDA)[10] were solved through synthetic modification. Once the

structures of the non-standard nucleobases were optimized, physical organic studies showed that the non-standard base pairs contribute to duplex stability in the expected fashion, both in DNA-DNA and RNA-DNA duplexes.[9]
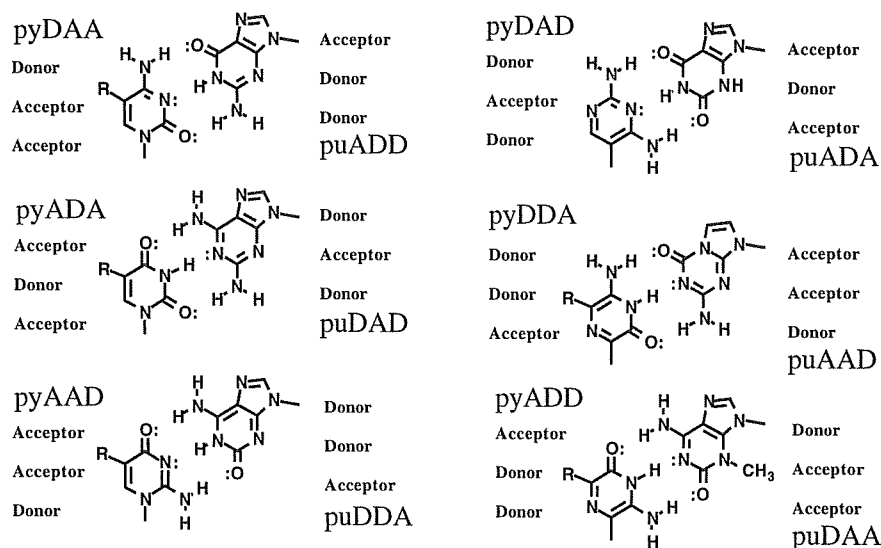

Figure 1. Twelve bases form 6 base pairs with a Watson-Crick geometry. Pyrimidines are designated by the prefix "py", purines by the prefix "pu". Following the prefix is the order, from the major groove to the minor groove, of acceptor (A) and donor (D) groups.

By manipulating the central element of nucleic acid molecular recognition without destroying the recognition behavior, synthetic organic chemistry has shown the remarkable versatility of the Watson-Crick formalism. Several properties of the non-standard nucleosides suggest, however, reasons why natural DNA does not use the non-standard nucleobases. For example, nucleosides with the pyDDA and pyADD heterocycles undergo slow, specific acid-catalyzed epimerization. Epimerization does not prevent these nucleosides from being used in *in vitro* experiments,[10] and the problem can be fixed by supporting the heterocycle on a carbocyclic ribose analog implementing the same hydrogen bonding pattern. These chemical properties are sub-optimal for a genetic encoding system, however.

Technological applications of an expanded genetic alphabet
Non-standard bases provide a molecular recognition system that has the properties of DNA but does not cross-react with natural oligonucleotides. Exploiting this specificity, Horn et al. at Chiron incorporated non-standard nucleobases into their branched DNA diagnostics systems.[11] In this system, an analyte DNA molecule serves as a "sandwich" to immobilize a branched DNA molecule carrying signalling units on to a solid support via a series of hybridizations. Incorporating non-standard bases into the hybridization oligonucleotides that assemble the signal molecules improves the signal to noise ratios of the system.[12,13]

Further, the expanded genetic alphabet permits us to enlarge the genetic lexicon, increasing the number of amino acids that can be translated into proteins. With four letters, only 64 triplet codons (4^3) are possible in the standard genetic lexicon., limiting the number of types of amino acids that can be built into proteins by ribosome-based translation. Moreover, the genetic code is degenerate; most individual amino acids are encoded by more than one triplet codon. Thus, all 64 codons possible with four letters are used to encode 20 proteinogenic amino acids, a small fraction of the thousands that are conceivable.

In collaboration with Jim Bain and Dick Chamberlin (University of California, Irvine),[14] the pyAAD was incorporated into a messenger RNA molecule, and the message was presented to a ribosome in the presence of a tRNA molecule that carried the complementary puDDA nucleobase in the anticodon loop, creating a new codon-anticodon pair. The non-standard tRNA was charged with a non-standard amino acid, and the extra "word" written with the non-standard nucleobase enabled the ribosome to incorporate iodotyrosine, a non-standard amino acid, efficiently into a polypeptide. These experiments

also developed new insights into how translation works. When a message containing the non-standard (iso-C)AG codon is incubated without the charged non-standard tRNA, translation does not stop. Rather, a frame shift occurs, the ribosome skips the non-standard base, and translation continues.[14]

Molecular biology of the expanded genetic alphabet

An expanded genetic alphabet needs DNA and RNA polymerases able to copy it. In academic and industrial collaborations with Stephen Hughes, Samuel Wilson, Catherine Joyce, William Beard, Clyde Hutchison, Bernard Angerer, David Gelfand and others, DNA and RNA polymerases were identified that accept non-standard nucleobases, although generally not with the same efficiency and fidelity as standard bases.[15,16,17] Mutation of these polymerases has improved their ability to accept non-standard nucleobases,[18] and we are now making the first attempts to amplify oligonucleotides containing non-standard nucleic acids using a procedure similar to the polymerase chain reaction.

CHANGING THE BACKBONE

The standard model of nucleic acid structure proposes little role for the nucleic acid backbone, other than to hold the nucleobases together. Therefore, many alternative backbones have suggested under the presumption that they will support the molecular recognition properties of nucleic acids as well. In some of these, the elements of the backbone are simpler than ribose, and some have suggested that they might have preceded RNA in the origin of life on earth.[19] We have tested one flexible linker based on glycerol (Figure 2b), proposed by Joyce et al.[19] as a possible prebiological genetic element. The stability of duplexes between DNA strands containing the flexible linker was much lower than expected from the standard model.[20] These results suggest that the backbone serves more than a passive role in strand-strand interaction. The more rigid ribose backbone evidently provides a degree of preorganization necessary for formation of a more stable duplex.
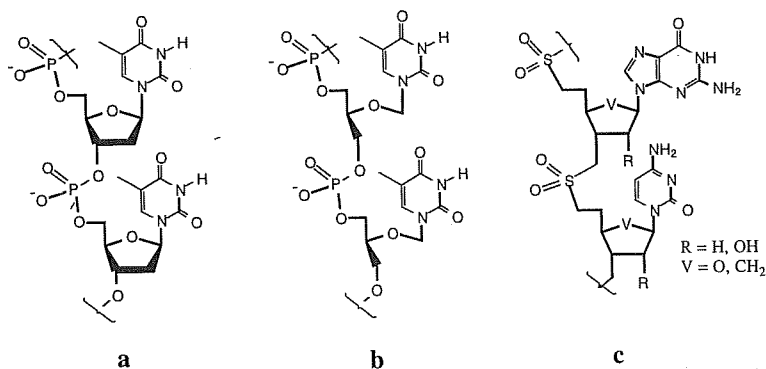

Figure 2. Standard phosphate backbone (a), a flexible sugar analog (b), and sulfone-linked analogs of DNA (c).

Why phosphates in the backbone?

To further explore the role of the backbone on molecular recognition in nucleic acids, DNA and RNA analogs (SNAs and rSNAs) were synthesized where uncharged dimethylenesulfone units replaced the phosphates (Figure 2c). Dimethylene sulfone units are largely isosteric and isoelectronic analogs of phosphates, so structural perturbations should originate primarily from the lack of associated anionic charge in the sulfones. In the crystal, the self-complementary $rGSO_2C$ dimer forms a duplex very similar to that formed by the natural ribodinucleotide GPO_2^-C.[21] Short SNAs behave as expected as nucleic acid analogs in other respects as well.

Oligosulfones (SNAs) can fold and be "catalysts"

More complex oligosulfones display more complex behavior. For example, the rSNA sequence:
$HO-rA_{SO_2}U_{SO_2}G_{SO_2}G_{SO_2}U_{SO_2}C_{SO_2}A_{SO_2}U-OH$
shows no binding to complementary oligonucleotides and no biological activity. Instead, this molecule self-associates, with a melting transition at 83°C.[22] Unhindered by intrastrand coulombic repulsion, it appears as if oligosulfones fold. Other unexpected properties are observed. For example, a partially protected SNA molecule transformed itself by removing a benzoyl protecting group 30 fold faster than in a simple analog, the 2',3"-deprotected adenosine monomer.[23] This debenzoylation is not true catalysis, of course, as the molecule effecting the transformation is altered in structure by the transformation. It is, however, "catalysis" in the same sense as the self-splicing ribozyme of Cech and coworkers is a catalyst. Without the polyanionic backbone, this RNA analog has a reactivity more reminiscent of proteins than RNA.

These results help to explain the role of the polyanionic backbone in DNA and RNA, beyond simply conferring aqueous solubility. The polyanion backbone has often been viewed as destabilizing to duplex formation, due to coulombic repulsion between polyanionic complementary strands. In fact, the backbone assists in duplex formation.

(a) Phosphates force strands to interact with other polyanionic strands through contacts as far from the backbone as possible. Without the negative charges, Hoogsteen interactions and interactions between sugars can dominate strand-strand association. Thus, the polyanionic backbone guarantees that the Watson–Crick rules for base pairing are obeyed.

(b) *Phosphates force the single strand to adopt an extended structure, pre-organizing it for association with its complement.* Coulombic repulsions between phosphates on the *same* strand (intra-strand repulsions) are more important than inter-strand repulsion; phosphates on the same strand are closer to each other than to any phosphates on the opposite strand. Intra-strand coulombic repulsion causes the single strand to stretch out in solution. This effect is well known in the statistical mechanical theory of polymers,[24] and prepares the single strand for binding to its complement.

(c) *Repeating anionic groups make the physico-chemical properties of DNA and RNA largely independent of sequence.* We found that each SNA has its own unique set of physico-chemical properties (solubility, folding, chemical reactivity), much like proteins. The interaction between phosphates (bearing the point charges) dominates interactions involving hydrogen bonds or hydrophobic effects in DNA and RNA. The resulting physico-chemical properties are, to a first approximation, largely independent of their nucleobase sequences. Thus, DNA mutations can change the encoded information without creating molecules that no longer function chemically. Compare this, for example, with sickle cell hemoglobin, where a single amino acid substitution radically alters the solubility of the molecule.

CONCLUSIONS

In 1986, essentially no work had been done to apply synthetic organic chemistry to oligonucleotides as polymers. Synthesis has now yielded DNA and RNA molecules with an expanded genetic alphabet. Molecular biology has yielded enzymes that copy these molecules. Systematically altered oligonucleotides provide insights into the structural importance of the nucleobases and backbone in DNA and RNA. A "second generation" model for nucleic acid structure is now emerging. Last, oligonucleotide analogs are now being used to solve practical problems in diagnostic systems, and will shortly be the basis for new methods for creating receptors, catalysts, and pharmaceuticals.

ACKNOWLEDGEMENTS

REFERENCES

1. J. D. Watson and F. H. Crick. *Nature* **171**, 737-738 (1953).
2. J. D. Watson and F. H. Crick. *Nature* **171**, 964-965 (1953).
3. G. B. Elion. *Science* **244**, 41-47 (1989).
4. K. P. Nambiar *et al. Science* **223**, 1299-1301 (1984).
5. C. Y. Switzer *et al. J. Am. Chem. Soc.* **111**, 8322 (1989)
6. J. A. Piccirilli, *et al. Nature* **343**, 33-37 (1990).
7. J. A. Piccirilli *et al. Helv. Chim. Acta* **74**, 397-406 (1991).
8. U. von Krosigk, Ph.D. Dissertation No. 10164, ETH Zürich (1993).
9. J. J. Voegel and S. A. Benner. *J. Am. Chem. Soc.* **116**, 6929 (1994).
10. U. von Krosigk and S. A. Benner. *J. Am. Chem. Soc.* **117**, 5361-5362 (1995).
11. M. Urdea. *Clin. Chem.* **39**, 725-726 (1993).
12. T. Horn *et al. Tetrahedron Lett.* **36**, 203-206 (1995).
13. T. Horn *et al. Nucleosid. Nucleotid.* **14**, 1023-1026 (1995).
14. J. D. Bain, *et al. Nature* **356**, 537-539 (1992).
15. C. Y. Switzer *et al. Biochemistry* **32**, 10489-10496 (1993).
16. J. Horlacher *et al. Proc. Natl. Acad. Sci. USA* **92**, 6329-6333 (1995).
17. M. J. Lutz *et al. Nucl. Acids Res.* **24**, 1308-1313 (1996).
18. M. J. Lutz, unpublished results.
19. G. F. Joyce *et al. Proc. Nat. Acad. Sci. USA* **84**, 4398-4402 (1987).
20. K. C. Schneider and S. A. Benner. *J. Am. Chem. Soc.* **112**, 453 (1990).
21. A. L. Roughton *et al. J. Am. Chem. Soc.* **117**, 7249-7250 (1995)
22. C. Richert *et al. J. Am. Chem. Soc.* **118**, 4518-4531 (1996).
23. Z. Huang, Ph.D. Dissertation No. 10429, ETH Zürich (1993).
24. P. Flory. *Principles of Polymer Chemistry*, Cornell Univ. Press, Ithaca (1953).