A NEW MOLECULE OF LIFE?

Peptide nucleic acid, a synthetic hybrid of protein and DNA, could form the basis of a new class of drugs— and of artificial life unlike anything found in nature

By Peter E. Nielsen

KEY CONCEPTS

- A synthetic molecule called peptide nucleic acid (PNA) combines the information-storage properties of DNA with the chemical stability of a proteinlike backbone.

- Drugs based on PNA would achieve therapeutic effects by binding to specific base sequences of DNA or RNA, repressing or promoting the corresponding gene.

- Some researchers working to construct artificial life-forms out of mixtures of chemicals are also considering PNA as a useful ingredient for their designs.

- PNA-like molecules may have served as primordial genetic material at the origin of life.

—*The Editors*

For all the magnificent diversity of life on this planet, ranging from tiny bacteria to majestic blue whales, from sunshine-harvesting plants to mineral-digesting endoliths miles underground, only one kind of "life as we know it" exists. All these organisms are based on nucleic acids—DNA and RNA—and proteins, working together more or less as described by the so-called central dogma of molecular biology: DNA stores information that is transcribed into RNA, which then serves as a template for producing a protein. The proteins, in turn, serve as important structural elements in tissues and, as enzymes, are the cell's workhorses.

Yet scientists dream of synthesizing life that is utterly alien to this world—both to better understand the minimum components required for life (as part of the quest to uncover the essence of life and how life originated on earth) and, frankly, to see if they can do it. That is, they hope to put together a novel combination of molecules that can self-organize, metabolize (make use of an energy source), grow, reproduce and evolve.

A molecule that some researchers study in pursuit of this vision is peptide nucleic acid (PNA), which mimics the information-storing features of DNA and RNA but is built on a proteinlike backbone that is simpler and sturdier than their sugar-phosphate backbones. My group developed PNA more than 15 years ago in the course of a project with a rather more immediately useful goal than the creation of unprecedented life-forms. We sought to design drugs that would work by acting on the DNA composing specific genes, to either block or enhance the gene's expression (the production of the protein it encodes). Such drugs would be conceptually similar to "antisense" compounds, such as short DNA or RNA strands that bind to a specific RNA sequence to interfere with the production of disease-related proteins [see "Hitting the Genetic Off Switch," by Gary Stix; SCIENTIFIC AMERICAN, October 2004].
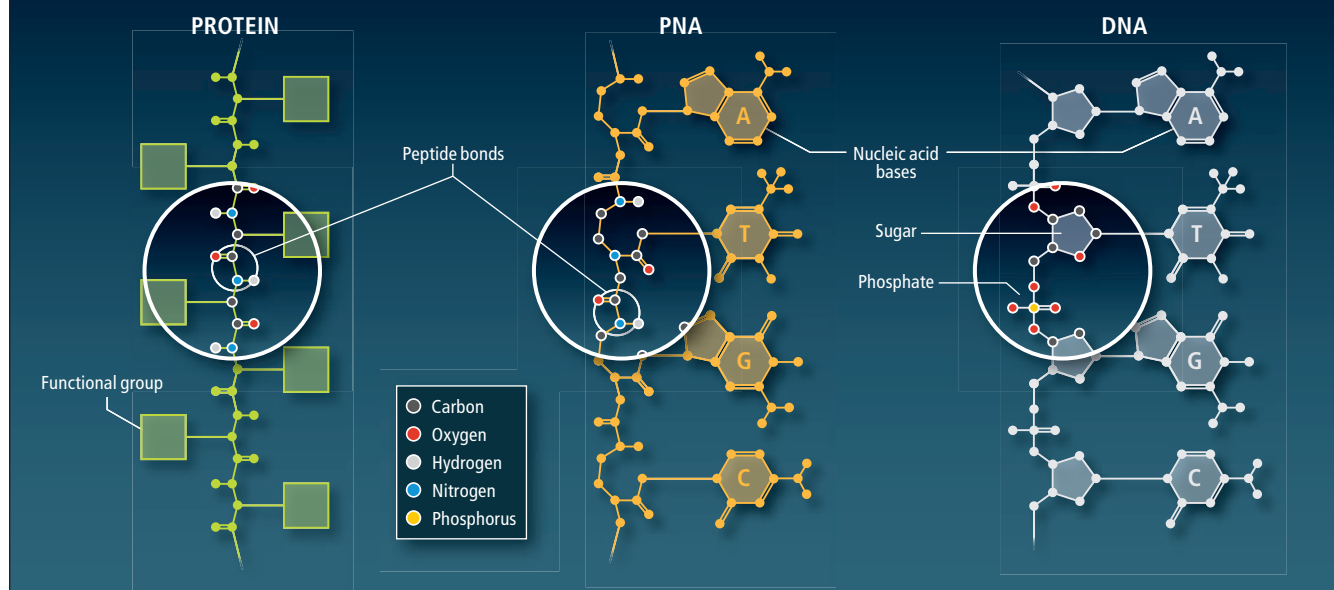
PNA's unique properties potentially give it several advantages over antisense DNAs and RNAs, including more versatility in binding to DNA as well as RNA, stronger binding to its target and greater chemical stability in the enzyme-laden cellular environment. Many studies have demonstrated PNA's suitability for modifying gene expression, mostly in molecular test-tube experiments and in cell cultures. Studies in animals have begun, as has research on ways to transform PNA into drugs that can readily enter a person's cells from the bloodstream.

PEPTIDE NUCLEIC ACID (*gold*) readily enters DNA's major groove to form triple-stranded and other structures with DNA, allowing it to modify the activity of genes in new ways.

PNA: A PROTEIN-DNA HYBRID

Peptide nucleic acid (PNA) combines features of protein and DNA. The backbone of PNA consists of relatively simple units linked by sturdy peptide bonds (*box*), as in a protein, unlike the less stable and negatively charged phosphate-sugar backbone of DNA. Each unit of PNA carries a nucleic acid base, as in a strand of DNA.

PROTEIN   PNA   DNA

Peptide bonds

Nucleic acid bases

Sugar

Phosphate

Functional group

- ○ Carbon
- ● Oxygen
- ○ Hydrogen
- ● Nitrogen
- ● Phosphorus

[THE AUTHOR]

Peter E. Nielsen heads the Center for Biomolecular Recognition in the department of cellular and molecular medicine at the University of Copenhagen. He has been affiliated with the university since 1980, when he received his Ph.D. there. As well as being one of the inventors of peptide nucleic acid, he co-founded two biotechnology companies in Denmark (Pantheco and PNA Diagnostics, now part of Santaris and Applied Biosystems, respectively) to develop applications of the molecule, and he has more than 20 related patents and patent applications.

In addition to fomenting exciting medical research, these amazing molecules have inspired speculations relating to the origin of life on earth. Some scientists have suggested that PNAs or a very similar molecule may have formed the basis of an early kind of life at a time before proteins, DNA and RNA had evolved. Perhaps rather than creating novel life, artificial-life researchers will be re-creating our earliest ancestors.

Into the Groove

The story of PNA's discovery begins in the early 1990s. To generate drugs with broader capabilities than antisense RNA, my colleagues Michael Egholm, Rolf H. Berg, and Ole Buchardt and I wanted to develop small molecules able to recognize double-stranded, or duplex, DNA having specific sequences of bases—no easy task. The difficulty has to do with the structure of the familiar DNA double helix.

It is the bases—thymine (T), adenine (A), cytosine (C) and guanine (G)—that store information in DNA. (In RNA, thymine is replaced by the very similar molecule uracil, or U.) Pairs of these bases joined by hydrogen bonds form the "rungs" of the familiar DNA "ladder." C binds with G, and A binds with T, in what is called Watson-Crick base-pairing. A compound that binds with a stretch of double-helical DNA having a characteristic base sequence would therefore be one that acts on any gene containing that particular sequence of bases on one of its strands.

The task of recognition is relatively easy if a compound has to find a particular base sequence on *single*-stranded DNA or RNA. If two nucleic acid strands have complementary sequences, standard base-pairing can zip the two strands together. Thus, if one knows the sequence of a gene—from Human Genome Project data, for instance—producing a molecule to latch onto a section of the gene in a single strand is as simple as synthesizing the complementary sequence.

In duplex DNA, however, the task of recognizing a sequence is more challenging because the atoms responsible for Watson-Crick pairing are already involved in the hydrogen bonds linking the two strands together and thus are not available for linking with another molecule. Yet cells contain numerous so-called gene-regulatory proteins that recognize sequences in duplex DNA to carry out their function of controlling gene expression. So the feat can be accomplished. If my group could find molecules capable of the task, the molecules could potentially serve as gene-regulating drugs.

Gene expression takes place in two stages. First, in transcription, an enzyme constructs

messenger RNA (mRNA), which is a strand of RNA with a copy of the base sequence of one strand in the DNA helix. A molecular machine known as a ribosome, itself made of RNA and protein, carries out the second stage, translation of the mRNA into the protein coded by the gene. Antisense agents interfere with translation by binding to the mRNA. These compounds are typically small, chemically modified RNA or DNA molecules, designed with the appropriate sequence to identify their mRNA target by Watson-Crick base-pairing. By binding to its mRNA, the agent may trigger enzymes to degrade the RNA or may simply interfere physically with the mRNA's functioning.

Cells make use of proteins called transcription factors that recognize specific sequences in double-stranded DNA to control gene expression at the transcription stage. These proteins can repress a gene by obstructing the RNA polymerase enzyme that would otherwise transcribe the DNA's sequence into mRNA, or they can activate a gene by helping the RNA polymerase to attach to the DNA and start transcription.

Although these proteins offer a model of molecules capable of "reading" the DNA sequence from the outside of the helix, in the 1990s it was not yet possible for biochemists to start with a sequence and design a new protein to recognize it. A gene-regulatory protein recognizes its DNA sequence by having the correct overall shape and chemical composition on its surface to bind with the sequence in the so-called major groove of the DNA, which provides access to the base pairs that run along the center of the double helix. But the structure of the protein's active surface depends on how its chain of amino acids folds up, a process that researchers cannot model with any accuracy.

Some progress has been made since then by taking the lead from gene-regulatory proteins that include zinc-finger domains, which are lengths of about 30 amino acids that fold around a zinc ion, forming a characteristic "finger" structure that can fit in the major groove with a few amino acids lined up with the DNA's bases. Researchers have developed artificial proteins with zinc fingers, but in general it is still difficult to program a sequence of amino acids to match even a relatively short DNA sequence.

A discovery dating back to 1957, only four years after the discovery of the DNA double helix, provides another approach. That year Gary Felsenfeld, Alexander Rich and David Davies, all then at the National Institute of Mental

RECOGNIZING DNA

PNA binds to DNA and RNA both by Watson-Crick pairing of complementary bases (as occurs in ordinary double-stranded, or duplex, DNA) and by so-called Hoogsteen pairing, allowing a variety of structures to form and providing many ways for PNA to act as a drug in a cell.

BASE-PAIRING

Watson-Crick pairing (*blue bonds*) links complementary bases A and T (*left*) and G and C (*right*). Hoogsteen pairing (*green bonds*) allows another T (*orange, left*) to join an A-T pair or a C (with an added hydrogen) to join a G-C pair.
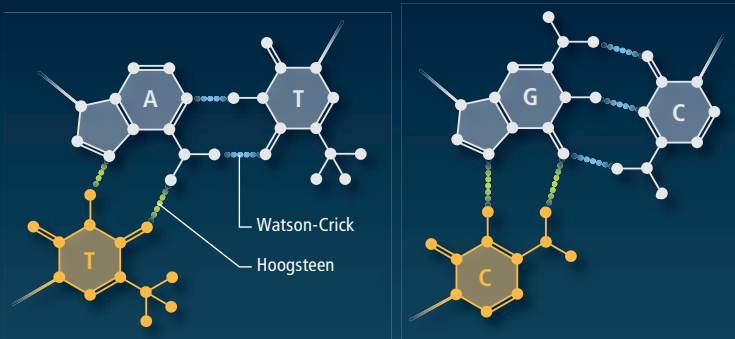

PNA-DNA STRUCTURES

A particularly interesting structure that has a number of potentially useful biological effects in cells is the so-called triplex invasion (*right*), in which two PNA strands (*orange*) displace one strand of a duplex DNA molecule and form a triple-stranded, or triplex, structure with its partner. The other structures shown below include the basic PNA-DNA triplex, also depicted on page 65.
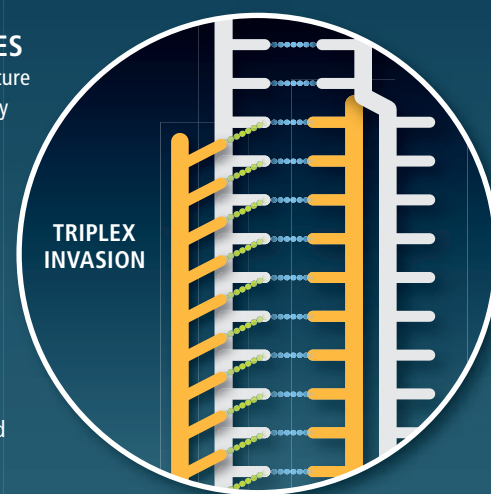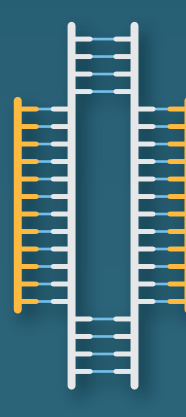

TRIPLEX INVASION

TRIPLEX

DUPLEX INVASION
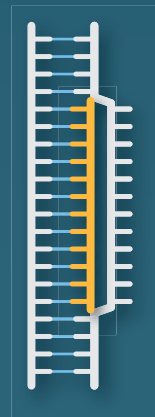
DOUBLE DUPLEX INVASION

GENE REGULATION BY PNA

PNA's ability to bind with double-stranded DNA as well as with RNA gives it several ways to modify the production of proteins encoded by specific genes.

CONTROLLING TRANSCRIPTION

At the transcription stage, an enzyme (RNA polymerase) transcribes the information encoded by a DNA molecule into a messenger RNA molecule. PNA can block this process by binding to a part of the gene that is to be transcribed (*inset*). Alternatively, triplex invasion by PNA can promote transcription by exposing an appropriate stretch of single-stranded DNA to enzymes that initiate transcription (*not shown*).

DNA   RNA polymerase

DNA   RNA polymerase

Messenger RNA

PNA
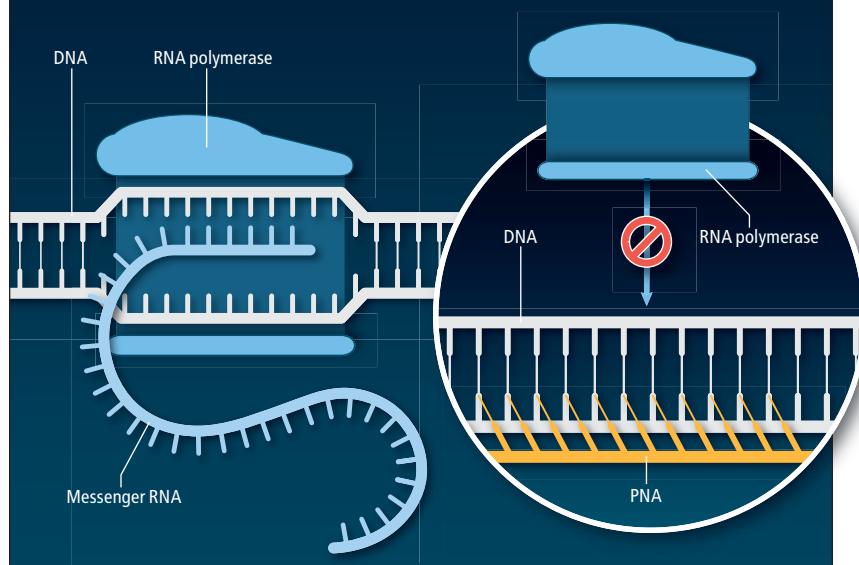
BLOCKING TRANSLATION

In the second stage of protein production, a ribosome translates the sequence on the messenger RNA into a sequence of amino acids, forming the protein. PNA can interfere with this stage by binding to the messenger RNA.

Messenger RNA   Ribosome

Messenger RNA   Ribosome

Protein

PNA
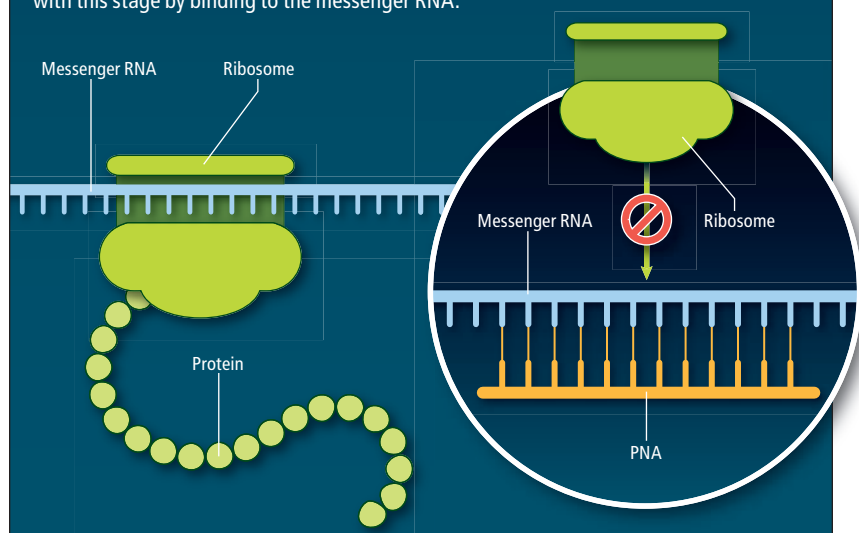
Health, created triple helix structures in which a nucleic acid strand attaches itself in the major groove of a duplex nucleic acid molecule. The extra strand exploits a different kind of bonding of the base pairs T-A and C-G called Hoogsteen pairing, after Karst Hoogsteen [*see box on preceding page*]. Each position along the triplex thus has a triplet of bases in which a T binds to a T-A pair (T-A=T, where the "=" indicates the Hoogsteen pairing) or a C binds to a C-G unit (C-G=C). This structure, however, can form only when the extra strand is a homopyrimidine—made entirely of C and T (or U, in RNA)—because each Hoogsteen pair requires a G or an A on the strand of the double helix.

In 1987 the late Claude Hélène, then at the National Museum of Natural History in Paris, and Peter B. Dervan of the California Institute of Technology independently demonstrated that the triple helix structure could indeed be exploited to design oligonucleotides (DNA strands about 15 nucleotides long) that read the sequence in double-stranded DNA and bind their Hoogsteen complementary target.

A Surprise Invasion

Inspired by this digital readout of the DNA double helix by groove-binding, triple helix–forming oligonucleotides, my group set out to synthesize a molecule that could do the same trick with fewer limitations. In particular, we hoped to find molecules that would not be limited to recognizing sequences made entirely of G and A. We also wanted our molecule to be neutral. The backbone of nucleic acids contains phosphate groups that carry a negative charge in solution. The repulsion caused by these negative charges on all three backbones weakens the binding of the third strand to the triplex.

We therefore decided to base the design on amide chemistry, involving the same kind of bond as links amino acids in proteins. Well-established techniques using amide, or peptide, bonds allow convenient synthesis of highly stable, neutral molecules. The peptide nucleic acid molecule that we came up with has a peptidelike backbone made of a much simpler repeating unit than the sugar and phosphate of DNA and RNA. Each unit may have a standard nucleic acid base (T, A, C or G) linked to it or bases that have been modified for special purposes. The spacing between bases along a PNA is very close to that of DNA and RNA, enabling short PNA strands, or PNA oligomers, to

form very stable duplex structures with DNA and RNA strands as well as with another PNA strand. The bases zip together with standard Watson-Crick bonding.

When we tried targeting duplex DNA with homopyrimidine PNA, to our surprise the PNA did not bind in the DNA's major groove as planned. Instead one PNA strand invaded the helix, displacing one of the DNA strands to form Watson-Crick bonds with its complement, and a second PNA strand formed Hoogsteen bonds to make a PNA-DNA=PNA triplex. The displaced length of DNA formed a single-stranded structure called a P-loop, alongside the triplex.

This triplex-invasion binding mode has several very interesting biological consequences, because the triplex has great stability and the P-loop affects central biological processes such as transcription, DNA replication and gene repair. For instance, the P-loop structure can initiate RNA transcription of the DNA. Furthermore, the single-stranded loop can be exploited in applications such as protocols to diagnose genetic disorders: the DNA in a sample must first be amplified (copied a large number of times), and the loop can serve as a specific attachment point for the copying process.

Other binding modes also occur, depending on the target DNA sequence and on how we modify the PNA's bases. Of these, double duplex invasion is particularly interesting. In this mode, we prepare two pseudocomplementary PNA oligomers—that is, their bases are modified enough to prevent formation of a PNA-PNA duplex but not enough to disrupt their individual binding to an ordinary complementary DNA strand. The PNAs thus invade double-stranded DNA and form two PNA-DNA duplexes. In contrast to triplex formation, which requires a long stretch of purines (A and G) in the target DNA, the double-duplex-invasion binding mode has less restrictive sequence requirements: with the present technology, the target sequence must contain at least 50 percent A-T base pairs. Even that constraint would be relaxed with discovery of suitable modified forms of the G and C bases.

PNA binds in these ways to complementary RNA or DNA molecules with even greater specificity and affinity than that exhibited by natural DNA. PNA oligomers with fluorescent groups attached are thus attractive as probes to detect specific genes in diagnostic tests. For instance, so-called fluorescence in situ hybridization analyses highlight the positions on chromosomes where specific sequences are present.

PROOFS OF CONCEPT

Many experiments have demonstrated how PNA oligomers can affect biological processes in ways that could be therapeutically useful.

Peter M. Glazer's group at Yale University showed in animal cells that PNA forming a triplex-invasion structure with DNA activated the repair of a gene carrying a mutation responsible for the human blood disease thalassemia. This process may eventually be exploited in therapy of diseases caused by single base mutations or small deletions in a gene.

David R. Corey of the University of Texas Southwestern Medical Center at Dallas and his co-workers showed in human breast cancer cells that PNA can block transcription by binding to a stretch of DNA that is separated into two strands just before the RNA polymerase enzyme starts transcribing the DNA into RNA. This process will provide insights into the role of specific genes in turning cells malignant.

Researchers have produced PNAs that block expression of essential genes in bacteria, inhibiting the bacteria's growth or even killing them. This discovery may lead to novel classes of antibiotics to combat the problem of bacterial resistance to existing drugs.

Prospects for Drugs

Many studies, in cell cultures as well as solutions in vitro, have demonstrated proof of concept for using PNA oligomers to suppress or activate the transcription, replication or repair of specific genes by binding to DNA in various ways. Researchers have also reported numerous experiments showing that PNA oligomers can function somewhat in the manner of antisense RNA interference, inhibiting gene expression at the translation stage, both in cell cultures and in a few studies with mice. PNA achieves these effects by physically blocking key processes involving RNA. In contrast, DNA or RNA oligomers used for RNA interference are assisted by enzymes in cells that break down the RNA-DNA or RNA-RNA duplexes that are formed. The RNA-PNA structure is unlikely to receive this kind of assistance because the enzymes cannot recognize such a foreign structure, although so far researchers have studied the question only for one of the relevant enzymes. Yet the alien nature of PNA oligomers also makes them exquisitely stable in biological environs—enzymes that break down other peptides do not recognize them, so PNAs have more time to encounter matching RNA and disable it.

In some cases, blocking an RNA process can restore a healthy protein. Matthew Wood of the University of Oxford and his co-workers demonstrated in 2007 that PNA can exploit this effect. When they injected PNA into mice with muscular dystrophy, the injected muscles showed increased levels of the protein dystrophin, whose absence causes muscular dystrophy. The PNA prevented a bad segment of the dystrophin gene from being translated from RNA to protein, thus eliminating a debilitating mutation present in that segment while leaving intact enough of the dystrophin to function.

PNA oligomers and conventional nucleic acids share a common problem of poor bioavailability because they are large and predominantly hydrophilic (water-loving) molecules, making it difficult for them to enter cells, whose walls are made of hydrophobic lipid membrane. Despite the great stability of PNAs, they do not remain in an animal for long, being quickly excreted in urine thanks to their hydrophilicity. For instance, half of the PNA in a mouse is gone in less than half an hour. Thus, the advent of PNA-based drugs awaits the development of suitable chemical modifications or pharmaceutical formulations (that is, mixtures with other substances) to improve PNA bioavailability. Indeed, the

ARTIFICIAL LIFE

Researchers striving to construct new life-forms out of combinations of nonliving chemicals are considering PNA as the genetic (information-carrying) component of their designs because it is simpler and more stable than DNA or RNA. In the proposal shown below, PNA is embedded in the surface of a container that self-assembles out of surfactant molecules. Light-sensitive molecules in the cell power the generation of more surfactant molecules from precursor esters. If the PNA replicates (*top*), the expanding protocell may divide into two similar copies (*bottom*). By self-organizing, metabolizing (exploiting an energy source) and self-replicating, the protocell exhibits some of the hallmarks of life.

Light

PNA

Light-sensitive molecule

Ester

Waste

Surfactant
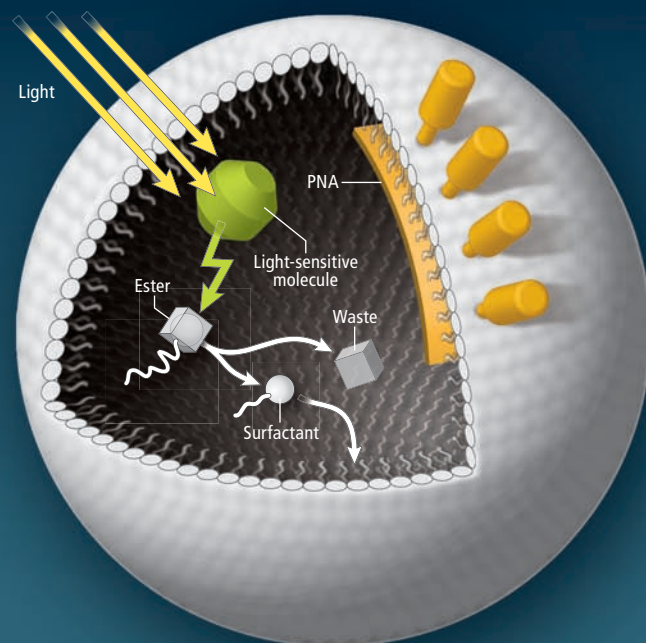
The PNA replicates when short, complementary PNA fragments attach to it and it migrates to the protocell's lipid interior, where the fragments join to form a second PNA strand.
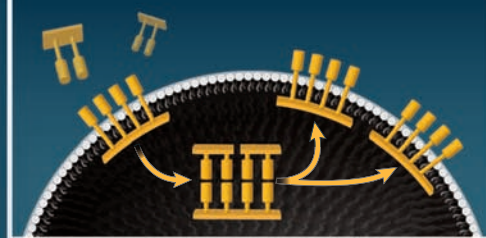
When the protocell grows large enough, it becomes unstable and divides in two.
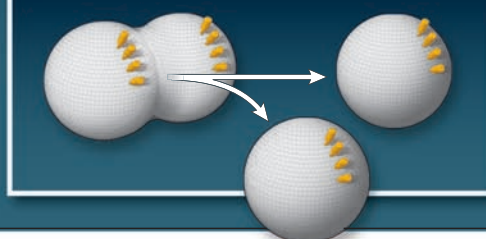
main focus of research into genetic medicines in general is work on overcoming the problem of delivery to cells in the body. Researchers believe that hurdle is the last obstacle holding back medical breakthroughs in this field.

Artificial Life

By bridging the realms of nucleic acids and proteins, PNA might be able to serve both as a store of information, like DNA, and as the catalytic machinery of an artificial cell, like the many protein-based enzymes in natural cells. It is that potential dual ability, along with PNA's other properties, that has attracted the interest of scientists seeking to create artificial life.

In many respects, however, RNA is ahead of PNA in this game. Natural and synthesized examples of catalytic RNA abound. Catalytic PNA molecules, in contrast, remain to be discovered. Yet just like proteins and RNA, PNA oligomers do fold up into the kinds of shapes (so-called secondary and tertiary structures) that are the key to performing catalysis, so I believe it is just a matter of time before a cata-

lytic variation on the PNA theme is developed.

The most advanced approaches to creating life from the bottom up, by assembling collections of molecules, seek to identify self-replicating RNA molecules that catalyze their own synthesis. In principle, the RNA molecules in these schemes could be substituted with PNA or a very similar synthetic molecule. Autocatalytic replication systems using short oligonucleotides have been discovered, as have self-replicating short peptides. Thus, it should be possible to develop analogous self-replicating PNA systems. A self-replicating system based on PNA would have the advantage of chemically robust peptide bonds, along with the versatility and specificity of base-sequence recognition.

Yet a genetic replication system is only one component of life, albeit a central one. The essence of life is a network of chemical reactions functioning in a state that is relatively stable yet not in equilibrium and that is open to both inputs and outputs [see "A Simpler Origin for Life," by Robert Shapiro; SCIENTIFIC AMERICAN, June 2007]. A major challenge will there-

TOMMY MOORMAN

fore be to incorporate the self-replicating molecule in a larger system that carries out other catalytic activity and has a metabolic cycle and to integrate the system with a physical compartment such as a lipid vesicle, forming what some researchers call a "protocell."

Steen Rasmussen of Los Alamos National Laboratory and Liaohai Chen of Argonne National Laboratory have suggested a primitive protocell design based on PNA [*see box on opposite page*]. The protocell container self-assembles out of surfactant molecules—lipid chains with hydrophilic, or water loving, heads. The PNA's backbone is modified to be lipophilic, or oil-loving, so that the PNA embeds itself in the protocell's surface. Short pieces of PNA pair up with the protocell's PNA to form a second strand with the complementary sequence. A light-sensitive molecule powers the production of more surfactant molecules, which increases the protocell's size. When it grows large enough, the protocell becomes unstable and naturally fissions. This proposal is, however, highly speculative, and it still suffers from a basic problem that chemists have yet to solve—the stability of double-stranded PNA greatly inhibits its separation into two daughter strands. A long, tortuous road remains before researchers develop robust artificial cells.

Origin of Life?

A major goal of these efforts to create life de novo in the laboratory is to better understand how life may have started on earth. Considering the detailed microbiology of contemporary life-forms, it seems very clear that RNA is probably more primordial and central to life than DNA and proteins. This one molecule can carry both the genotype (the genetic sequence information) of an organism and the phenotype (catalytic functions). For this reason as well as other evidence, many scientists now accept the idea that our DNA/RNA/protein world was preceded by an RNA world [see "The Origin of Life on the Earth," by Leslie E. Orgel; SCIENTIFIC AMERICAN, October 1994].

Yet it is very unclear how primitive prebiotic conditions could have produced RNA molecules, in particular the sugar ribose in the RNA backbone. Further, even if RNA molecules were produced, RNA's very poor chemical stability hardly would have allowed the molecules to survive unprotected long enough to play a central role in the initial chemical evolution of life.

PNA WORLD?

Some researchers have suggested that the earliest life on earth may have been based on PNA or a similar information-carrying molecule, because it is simpler and more stable than RNA, widely believed to have evolved before DNA. This "PNA world" would have evolved into an "RNA world" of RNA-based life and thence to the world of life as we know it, based on DNA, RNA and protein.

Thus, a molecule like PNA appears very attractive as a candidate for a pre-RNA world: it is extremely stable and chemically simple, and it carries sequence information.

In 2000 Stanley L. Miller, famous for his seminal experiments more than 50 years ago showing that amino acids can form under conditions believed to simulate those on the primitive earth, identified precursors of PNA in similar experiments. Researchers have also shown that sequence information in a PNA oligomer can be transferred by "chemical copying" to another PNA oligomer or to an RNA molecule—processes needed for a PNA world and then a following transitional PNA/RNA world. Admittedly, it is a long leap from these scanty observations to building a strong case for a pre-RNA world based on PNA or some very similar molecule, and for the hypothesis to have any legs at all, scientists must uncover PNA molecules possessing catalytic activity.

Much remains to be learned about PNA 15 years after its discovery: Are catalytic PNA molecules possible? What is a good system for delivering therapeutic PNA into cells? Can a totally alien, PNA-based life-form be created in the lab? I am confident these questions and many others will be well answered over the next 15 years. ■

➥ MORE TO EXPLORE

Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide. Peter E. Nielsen, Michael Egholm, Rolf H. Berg and Ole Buchardt in *Science,* Vol. 254, pages 1497–1500; December 6, 1991.

Template Switching between PNA and RNA Oligonucleotides. Christof Böhler, Peter E. Nielsen and Leslie E. Orgel in *Nature,* Vol. 376, pages 578–581; August 17, 1995.

Peptide Nucleic Acid: A Molecule with Two Identities. Peter E. Nielsen in *Accounts of Chemical Research,* Vol. 32, No. 7, pages 624–630; July 1999.

Synthesizing Life. Jack W. Szostak, David P. Bartel and P. Luigi Luisi in *Nature,* Vol. 409, pages 387–390; January 18, 2001.

Prebiotic Chemistry and the Origin of the RNA World. Leslie E. Orgel in *Critical Reviews in Biochemistry and Molecular Biology,* Vol. 39, No. 2, pages 99–123; March/April 2004.