Genes as **Parasites**

Parasites exist not only in the plant and animal kingdoms, they are also a part of us. Our genome contains myriad short stretches of DNA that propagate at the genome's expense. For this reason, these transposons, as they are called, are also referred to as parasitic DNA. **Oliver Weichenrieder** from the **Max Planck Institute for Developmental Biology** in Tübingen wants to shed light on the processes by which transposons are copied – not only because they can cause disease, but also because they may be an important engine of evolution.

TEXT **TIM SCHRÖDER**

The object of his research work is tiny and sparkles like fine diamond dust. Oliver Weichenrieder slides a plastic plate containing small wells back and forth under the microscope. He searches for a while. "Ah, I can see a few in there. Pretty, aren't they?" The small crystals are invisible to the naked eye, but under the microscope they glitter in violet, pink and blue hues. "It takes a bit of luck for the crystals to form. That's why we use these plates with many small depressions," Weichenrieder says. That increases the hit rate.

Crystal – that sounds like materials science, like glass and ceramics, but that's not at all what Oliver Weichenrieder is concerned with. He is a biochemist, and the crystals he grows at the Max Planck Institute in Tübingen are molecules from living cells, specifically proteins or ribonucleic acids (RNAs). Weichenrieder is studying the crystals to elucidate the structure of these molecules with a view to solving one of genetics' enduring secrets: the puzzle of parasitic DNA, or parts of the genome that self-propagate independently of the rest of the genome.

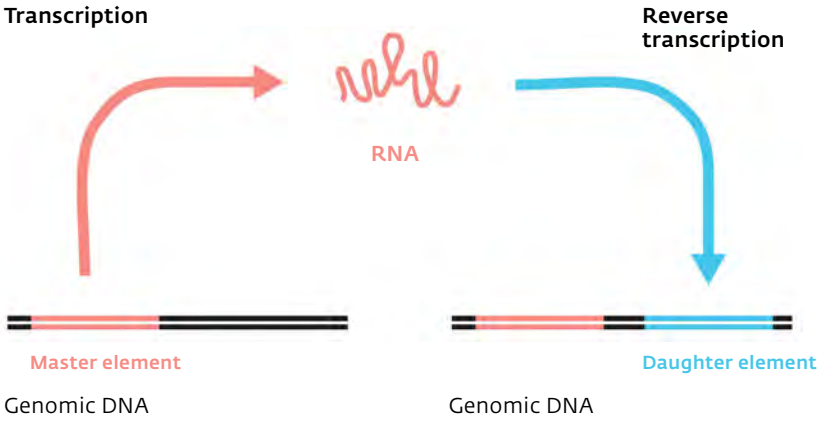
Parasitic DNA is a catchy description of genetic snippets known to scientists as transposons. Transposons are short sections of DNA that repeatedly replicate and insert themselves into new DNA sites in the genome. The term "transpose" is known from music and means writing or playing a piece of music in a different key. In the case of transposons, genetic information is transported from one site in the genome to another. This transfer process occurs during the development of germ cells and most notably in the early embryo when cells are dividing vigorously.

JUMP INTO THE GENOME

Each time this occurs, the DNA sequence is reshuffled – sometimes with serious consequences: "It can happen that a transposon lodges itself in a gene

Photo: Wolfram Scheible

The crystals of a transposon protein measure around one-tenth of a millimeter across. By analyzing the crystal structure, researchers in Tübingen have discovered how the protein is able to package transposon RNA.



Above For retrotransposons to be able to insert copies of themselves into the genome, their DNA, termed the master element, is first transcribed into RNA. The RNA is then transcribed back into DNA (reverse transcription), while the resulting DNA copy (the daughter element) is integrated into the genome.

Right Oliver Weichenrieder checks incubation cabinets for genetically altered bacteria. The microbes produce transposon proteins, which the researcher and his colleagues need in order to analyze the protein structure.

segment that contains information for a protein that is essential for metabolism," Weichenrieder explains. "The gene can then no longer be read correctly." The APC gene is a prime example. The protein it encodes can prevent the development of cancer. If it is disrupted by the incorporation of a transposon, colon cancer can result.

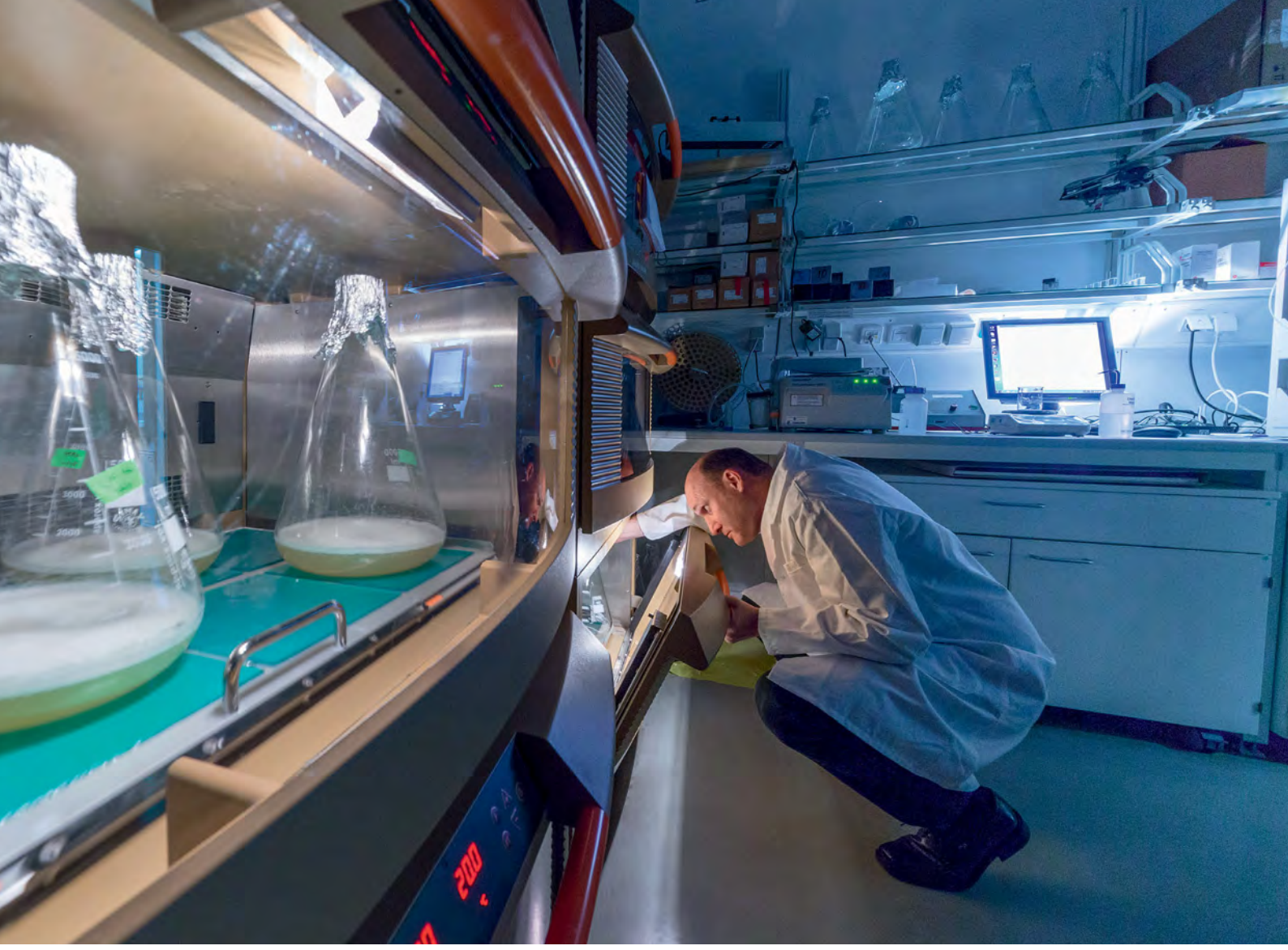
Proteins are essential for reading the information contained in DNA, transcribing it into RNA and translating the RNA into new proteins. Transposons, too, need various proteins to replicate and to insert themselves into the genome. "We still don't know exactly how transposons work," Weichenrieder says. "But if we elucidate the structure of the proteins involved in transposon replication, we might be able to better understand the whole process."

FLOOD OF TRANSPOSONS IN THE GENOME

Oliver Weichenrieder and his colleagues focus primarily on identifying the structure of two important parasitic DNA snippets known as the LINE-1 retrotransposon and the Alu retrotransposon, as both are extremely common in the human genome. A LINE-1 segment is about 6,000 DNA base pairs long – about as long as an average gene. The genome contains around 500,000 LINE-1 copies and fragments. In fact, the LINE-1 element alone makes up 17 percent of the genome. "This immense number has resulted from the creation of ever new copies in the course of evolution over millennia," Weichenrieder says. It's possible that more than half of our genome has been created from transposable elements. The proportion of transposons may be even higher in other organisms. In corn plants, for example, as much as 85 percent of the entire genome can be traced back to transposons.

Of course, the transpositions could only be passed on if they didn't kill the

Graphic: MPI for Developmental Biology; photo: Wolfram Scheible



individual, for example as a result of cancer. "Not every transposition is necessarily fatal for the individual or their offspring. It depends on which part of the genome the LINE-1 copy is incorporated into." In addition, many have been deactivated over time by mutations. At present, only around 100 of the 500,000 LINE-1 sections in the human genome are active and capable of parasitic behavior. The rest no longer work.

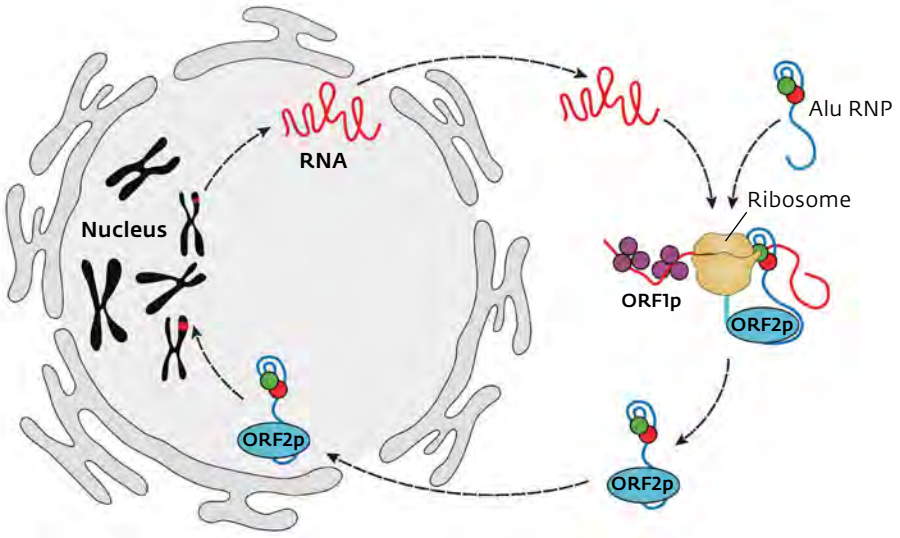
Weichenrieder is particularly fascinated by Alu retrotransposons, which are also distributed throughout the genome – not only because, with more than a million copies, they occur in great numbers and make up about 10 percent of the genome, but because, from a parasitic point of view, they take things a step further. They are parasites of a parasite in that they hijack the LINE-1 machinery and use it for their own replication. LINE-1 is able to replicate by itself. Alu, in contrast, requires LINE-1 proteins and uses them for its own purposes.

Transposons are therefore concerned only with their own propagation and don't appear to contribute anything to the organism's survival. The organism is just a means to an end. On closer inspection, however, it becomes clear that it can nevertheless benefit from transposons. Whenever transposons insert copies of themselves into the genome, they reshuffle the DNA, thus driving evolution. "That keeps the genome flexible," says Weichenrieder. "During times when the environment changes dramatically as a result of climate change or natural disasters, the organism is able to adapt more quickly." In fact, it has since been shown in plants that transposons are particularly active when the plants are under stress – for instance during a hot spell.

Many transposons are harmful, but sometimes they give an organism new characteristics that are vital to survival. "It has long been believed that evolution is driven by the exchange of individual letters of the genetic code, known as point mutations," Weichenrieder says. Now, however, many of his colleagues believe that transposons play an important role in vigorously reshuffling the genome, thus creating new variations.

PROTEINS FOR JUMPING GENES

In recent years, Oliver Weichenrieder and his team have used the small protein crystals to probe deep into the molecular structure of LINE-1 and Alu. Their laboratories in Tübingen contain equipment that allows them to extract the proteins required for the transposition of LINE-1 and Alu. These high-performance liquid chromatography systems separate a mixture of different proteins into its component parts so



that tiny amounts of a pure protein trickle out at the end. These are then mixed with a variety of substances on the microtiter plates and are concentrated in the hope that the small crystals will form.

FROM STRUCTURE TO FUNCTION

And then it's time to take a trip: the scientists bombard the crystals with intense X-ray light in particle accelerators in Zurich or Hamburg. The atoms of the proteins diffract the X-rays differently, producing a characteristic diffraction pattern. From this, computer programs determine the electron distribution in the crystal and, to a certain extent, the chemical structure of the protein molecule. "Nevertheless, there's still plenty of manual work for us to do. We need a great deal of chemical knowledge to reconstruct the entire protein from the fragments in the computer." This is how the researchers in Tübingen uncovered the structure of the LINE-1 proteins and of Alu RNA.

LINE-1 requires two main proteins to replicate successfully: ORF1p and ORF2p – that much was already known. The blueprint for those proteins is contained in the LINE-1 element. If LINE-1 is active, an RNA copy is first made. This migrates to the cell's protein factory, the ribosome, where it provides a template for synthesizing the proteins.

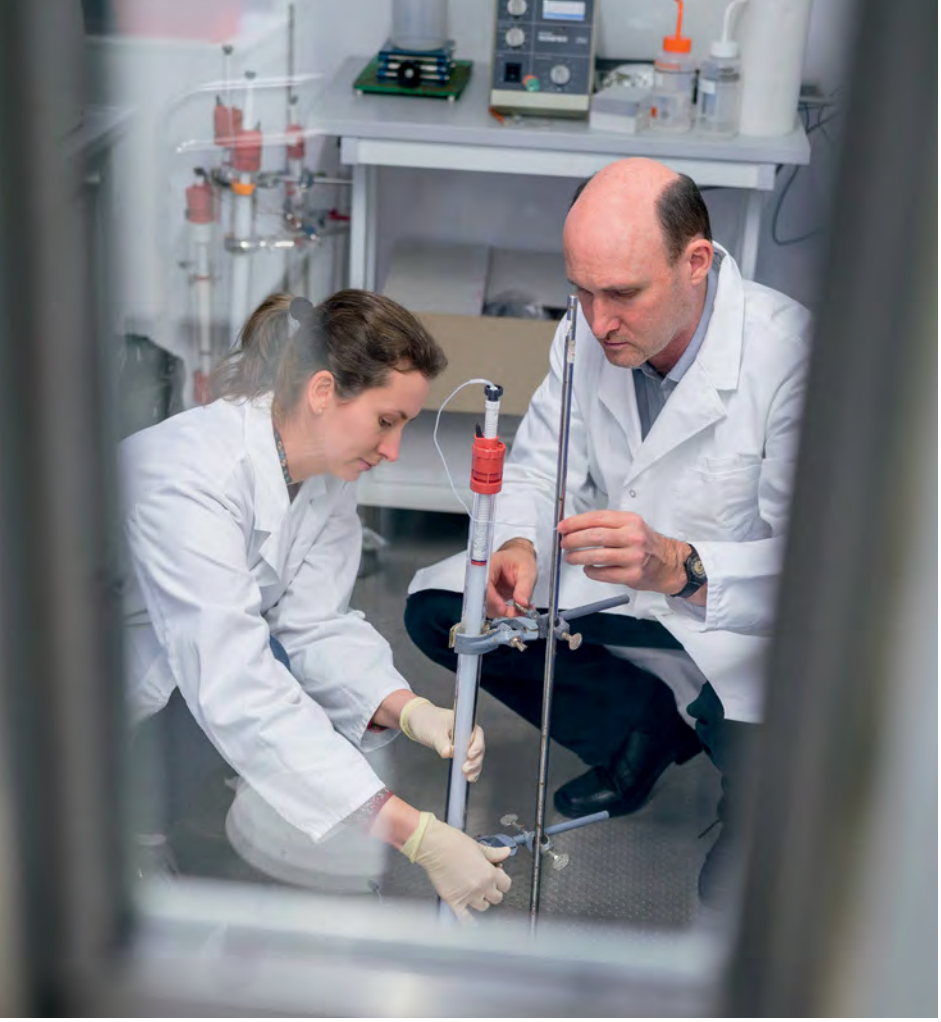
For a long time, it wasn't known how the two proteins function, but the identification of their spatial structure by the scientists in Tübingen provided clarity: ORF1p consists of three subunits, and the RNA is able to nestle between them. In addition, the ORF1 proteins always occur in bundles of three ORF1 protein molecules. "By all indications, the LINE-1 RNA is wound around the trimers of this protein to protect it from the cell's defense weapons," Weichenrieder explains. The cell is by no means completely at the mercy of a transposon; it can switch off the copying process in various ways, for instance with the help of proteins that destroy the transposon's RNA or attack ORF1p. The ORF1 protein protects the LINE-1 RNA from these countermeasures so that the RNA is able to go on to produce ORF2p.

ORF1p also resembles proteins that allow cell membranes to fuse. Accordingly, ORF1p might also help LINE-1 RNA conquer the nuclear membrane

Above A parasite attacks a parasite: A complex consisting of Alu RNA and proteins (Alu-RNP) recognizes the ribosomes that are currently translating the RNA of the parasitic LINE-1 element to produce ORF2p. The Alu complex hijacks the protein, which actually belongs to LINE-1, and hitches a ride with it into the cell's nucleus (gray), where ORF2p reverse-transcribes the Alu RNA and integrates the resulting DNA into the genome as a new Alu element.

Right-hand page Oliver Weichenrieder and Elena Khazina prepare a chromatography column (top). They inject a purified cell extract into the apparatus and then separate the protein mixture into its constituents (bottom).

Graphic: MPI for Developmental Biology



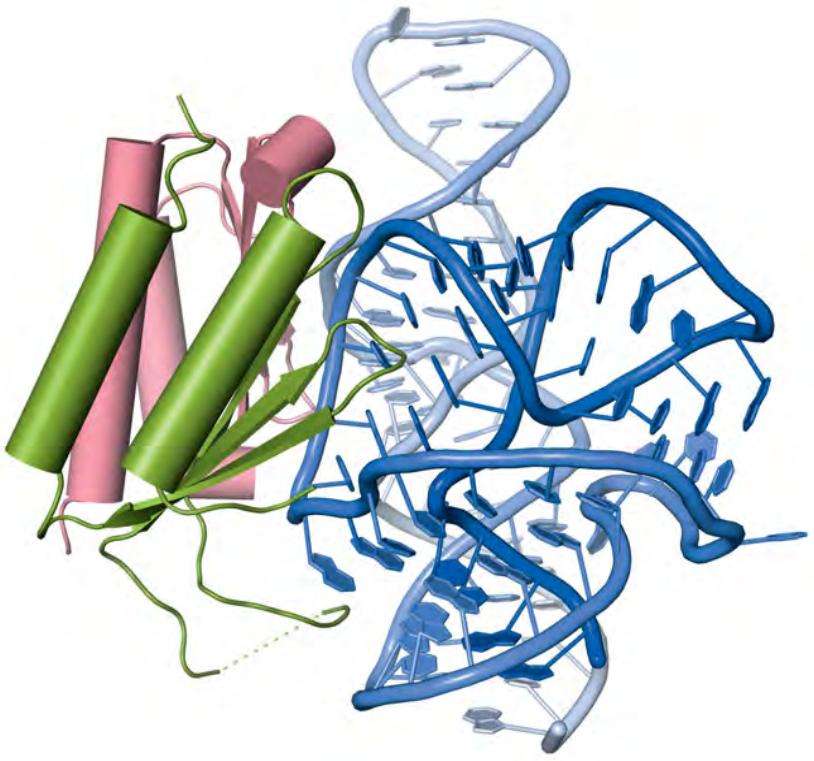
Photos: Wolfram Scheible (2)

to reach the DNA in the nucleus, to which it must return in order to replicate itself.

CUTTING AND TRANSCRIBING

ORF2p is needed for LINE-1 to replicate, and the researchers now have a more detailed understanding of its spatial structure, as well. It consists of two subunits: an endonuclease and a reverse transcriptase. The endonuclease first cuts the DNA. In doing so, it recognizes specific DNA structures that occur during the duplication of DNA or shortly before cell division, which clears the way for incorporation of the retrotransposon sequence. The reverse transcriptase then converts the RNA back into DNA and incorporates it into the genome.

The researchers' findings could lead to the development of drugs for forms of cancer in which the genome is damaged by transposons. Such drugs could block LINE-1 proteins, preventing transposition. However, Oliver Weichenrieder first wants to gain a thorough understanding of the underlying process, and Alu, being a parasite of a parasite, is particularly suitable for this purpose. "The Alu RNA evidently iden-



The crystal structure of an Alu ribonucleoprotein particle (RNP). The Alu RNA (blue) consists of ribonucleotides (adenine, guanine, cytosine and uracil), which fold into a well-defined structure stabilized by two proteins: SRP9 (red) and SRP14 (green). The actual molecular parasite is therefore the folded RNA.

tifies ribosomes that are currently producing ORF2p without directly recognizing the protein," he explains. "Once the ribosome has completed the ORF2p protein, Alu strikes and fishes it out." The Alu RNA then uses the hijacked endonuclease and reverse transcriptase to insert its own DNA into the genome. "So it depends on your point of view," Weichenrieder says, "because strictly speaking, the actual parasite isn't the transposon DNA but the LINE-1 or Alu RNA read from it."

STUTTERING ASSEMBLY LINE

It is estimated that only about one in twenty newborn babies carries a new Alu copy and passes it on to his or her descendants. The reason transposition isn't that common is that transposons are usually inactivated by cellular defense mechanisms. The transposition of LINE-1 is also fairly rare. "Alu RNA thus needs to be able to sense very precisely where LINE-1 is currently active and synthesizing ORF2p on a ribosome," Weichenrieder says – like a parasite that unerringly tracks down its host.

His results suggest that Alu may recognize a sort of stuttering that occurs during the synthesis of ORF2p on the ribosome. Some of the amino acids that are incorporated into the ORF2p protein during production are lysines, which don't readily pass through the ribosome's production channel. If they occur frequently, the assembly line can falter and stutter. It is very likely that Alu recognizes such stuttering and then arrives in time to pick up the finished ORF2p.

To date, no researcher has observed this process live, but all the evidence the scientists have gathered by analyzing protein structures points to this conclusion. In addition, the Alu retrotransposon is closely related to another particle known as the signal recognition particle (SRP), which regulates the production of specific proteins on the ribosome. Alu only emerged from SRP with the evolution of primates, which was a very recent event in evolutionary terms. The researchers in Tübingen have found that Alu RNA has retained substantial parts of the SRP structure and must also bind two

Graphic: MPI for Developmental Biology

Three scientists, one goal: By analyzing transposon proteins, Gabriele Wagner, Oliver Weichenrieder and Elena Khazina (from left) hope to gain an understanding of how parasitic DNA can spread in our genome.

small SRP proteins for it to replicate successfully as a retrotransposon.

For Weichenrieder, transposons are, in and of themselves, neither good nor bad. Nevertheless, he is intrigued by the idea of a molecular parasite: "Alu is so simple in structure: a short RNA segment and two small SRP proteins that fold the Alu RNA. It appropriates everything else it needs. By comparison, even viruses are enormously complex with their wealth of information on protein construction."

This raises the question of how small and simple a parasite can actually be and still function. Weichenrieder is therefore testing Alu variants to determine whether they can replicate. To this end, he gradually reduces their size and the number of proteins involved. In this way, he hopes to achieve his goal and find the "ultimate parasite" – a piece of RNA, pared down to the barest of necessities, that can still replicate successfully with the help of an organism's genome and at the organism's expense. ◀

www.tinyurl.com/y79ad2yp
(available only in German)

TO THE POINT

- **Parasitic DNA molecules are DNA segments that are able to replicate at the expense of the organism. In the case of LINE-1 and Alu, however, the corresponding RNA molecules are the actual molecular parasites.**
- **Large parts of the genome have developed from these transposons, which account for more than 80 percent of the genome of some plants. Most, however, are no longer active.**
- **Some RNA molecules, such as Alu RNA, exploit the copying machinery of other transposons, making them parasites of parasites.**

GLOSSARY

Reverse transcriptase: An enzyme used to transcribe RNA into the corresponding DNA sequence. Retrotransposons require this enzyme to transcribe a DNA copy from its RNA, and the copy is then reincorporated into the genome. The enzyme was first discovered in retroviruses, which have a genome consisting of RNA. In order for it to be incorporated in the host's genome, the reverse transcriptase must produce a DNA version.

Transposons: DNA segments that can change their location in the genome. They are sometimes referred to as "jumping genes." However, complete transposons often contain multiple genes. Transposons whose DNA-derived RNA is used not only for producing proteins, but also as a template for building a new DNA molecule, are termed retrotransposons. Retrotransposons probably also gave rise to retroviruses, which also insert themselves into the genome of the host, but can also leave it again as complete viruses.

Photo: Wolfram Scheible