Perspective

Directed Evolution of Nucleic Acids by Independent Replication and Selection

JOHN ABELSON

SPECIFIC INTERACTIONS BETWEEN NUCLEIC ACID AND PROTEIN mediate the control of gene expression. The understanding of these interactions has been facilitated by the construction of sequence variants of both nucleic acid and protein. For example, it is now commonplace to construct numerous variants of a promoter sequence in order to understand which nucleotides are important for gene expression and how regulatory factors bind to them. The decision of which residues to change in these experiments is necessarily dictated by the starting sequence. This naturally occurring sequence is the result of evolution and is therefore the idiosyncratic product of history and the complex environment of the cell. Thus, in general, a small percentage of the possible nucleic acid sequences to which a particular protein could bind are explored in these experiments. This is because single mutants are generally constructed, but multiple mutants are rarely explored.

An exciting new experimental procedure is described by Tuerk and Gold (see *Science*, p. 505) which provides a general way to study protein nucleic acid interactions. In their method, called SELEX (for the systematic evolution of ligands by exponential enrichment, but generally known as the "Tuerk-o-matic" in Boulder), a random collection of RNA molecules is enriched for those that can bind to a particular protein. Those RNA molecules that bind are converted to complementary DNA and are replicated by the polymerase chain reaction (PCR). Continued cycles of enrichment and amplification lead to a collection of sequences that bind with the highest affinity.

This technique was used to study the interaction between bacteriophage T4 DNA polymerase and the 5′ leader sequences of its mRNA. The T4 DNA polymerase translationally controls its own synthesis by binding to a sequence of 36 nucleotides in the mRNA that overlaps the ribosome binding site. Included in this sequence is a stem loop structure consisting of a five–base pair stem and a loop of eight nucleotides. A DNA template for this RNA was synthesized in which the nucleotides in the eight-base loop had been completely randomized. This collection of 4^8 (66,536) molecules was transcribed by T7 RNA polymerase to provide the starting material for the SELEX enrichment. After four rounds of selection and amplification, two sequences with equal affinity to T4 DNA polymerase predominated in the population. One was the wild-type sequence and the second differed in four bases of the loop. The sequence of the variant is such that two more base pairs could be added to the stem leaving a loop of four bases. This unexpected result forces us to consider new models for the structure of the wild-type loop, at least as it is interacting with the protein. The variant sequence has not appeared in any of the phage T4 relatives that have been sequenced and perhaps it could not evolve from the wild-type sequence since the intermediates would bind with lower affinity. Thus SELEX has provided a sequence that might have evolved but did not. It is unlikely that anyone would have chosen to make this mutant in a standard mutagenesis experiment since it is a quadruple mutant.

There is an intellectual precedent to this experiment. In a series of experiments done in Spiegelman's laboratory nearly 20 years ago, it was discovered that small, rapidly evolving RNA molecules arise in Qβ RNA replicase reaction mixtures (*1*). In this case, the selection and replication steps were not independent, and the variants arose because of the infidelity of Qβ replicase. The environment could be altered, however, providing additional selective pressure. For example, variants were selected that replicated in 15 μM ethidium bromide (*2, 3*).

Once automated oligonucleotide synthesis machines were available, many people began to use them as mutators (*4, 5*). In mutagenesis the strategy is often biased toward the selection of single mutants, but Struhl randomized contiguous blocks of sequence in promoters and selected functional sequences either by genetic selection or by physical selection (*6*) of sequences that could bind to a transcription factor (*7*).

Joyce used a similar approach to the isolation of variants of the *Tetrahymena* self-splicing intron (*8*). In this case, a trans-splicing reaction was used to join the two RNA molecules, a cDNA copy was produced and amplified by T7 RNA polymerase. The selection in this case is for variants that were functional in the trans-splicing reaction.

Tuerk and Gold have used a protein as the SELEX target, but the target could also have been a small molecule. We know that binding sites for small molecules can form in RNA molecules. For example, the *Tetrahymena* ribozyme has an active site for the guanosine cofactor (*9, 10*), and Yarus has shown that L-arginine can bind to this site and competitively inhibit the reaction (*11*). In fact, an experiment to evolve RNA molecules that bind to specific small molecule targets has already been done. In independent work, Ellington and Szostak (*12*) have used a very similar approach to SELEX to isolate RNA molecules that can selectively bind to small ligands. They randomized a sequence of 100 nucleotides in a DNA template. In this case, the RNA transcripts were selected by binding to dye columns, for example, Cibacron Blue, a dye that has structural similarity to nicotinamide adenine dinucleotide. Six different dye columns were used and five cycles of selection and amplification were carried out. Four of the six pools of RNA resulting from this experiment showed dye selectivity. Individual sequences in the pool had binding constants to the dye matrix of $\sim 10^{-4}$ M. Potentially, the randomization of 100 nucleotides can lead to a population of 4^{100} or 10^{60} molecules, but only a small fraction of that can be obtained; in this case the pools had a complexity of $\sim 10^{13}$. Of these, a population of 10^2 to 10^5 molecules (or aptamers as Ellington and Szostak term the product of the selection) could bind specifically to a ligand. The next step will be to select RNA molecules that can bind to transition state analogues in order to generate an RNA enzyme. By this approach, the many enzymatic activities required in the prebiotic RNA world might be evolved anew.

Tuerk and Gold suggest additional exciting and novel applications of the SELEX procedure. The plasticity of RNA molecules and the early successes imply that it will be possible to evolve RNA molecules that can bind to any target. Since the active site of an enzyme generally represents a cleft or pocket, it is a possible binding site for RNA. With the SELEX procedure, they suggest, arbitrarily tight binding RNA inhibitors of pharmacological targets might be

The author is in the Division of Biology, California Institute of Technology, Pasadena, CA 91125.

evolved. Provided that RNA or RNA analogues can be made to function as drugs, SELEX may offer a new and rapid approach to drug design. Even more daring is their idea that selection for binding of random proteins to a target can be used in SELEX. To do this the binding must be by nascent peptides in polysomes. The ribosome-associated mRNA coding for the selected peptide would then be isolated and amplified for use in repeated cycles of protein synthesis coupled selection. The advantages of this technique over genetic selection or large-scale screening is that far larger populations can be screened and organism viability is not an issue. Can it be that the "awesome power of genetics" has met its match?

REFERENCES AND NOTES

1. D. R. Mills *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 217 (1967).
2. R. Saffhill *et al.*, *J. Mol. Biol.* **51**, 531 (1970).
3. F. R. Kramer *et al.*, *ibid.* **89**, 719 (1974).
4. J. B. McNeill and M. Smith, *Mol. Cell. Biol.* **5**, 3545 (1985).
5. C. A. Hutchinson III *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 710 (1986).
6. A. R. Oliphant and K. Struhl, *Nucleic Acids Res.* **16**, 7673 (1988).
7. A. R. Oliphant, C. J. Brandl, K. Struhl, *Mol. Cell. Biol.* **9**, 2944 (1989).
8. G. F. Joyce, in *RNA: Catalysis, Splicing, and Evolution*, M. Belfort and D. Shub, Eds. (Elsevier, Amsterdam, 1977), pp. 83–87.
9. B. Bass and T. Cech, *Biochemistry* **25**, 4473 (1987).
10. F. Michel *et al.*, *Nature* **342**, 391 (1989).
11. M. Yarus, *Science* **240**, 1751 (1988).
12. A. D. Ellington and J. W. Szostak, *Nature*, in press.

Telomeres and Their Synthesis

ELIZABETH H. BLACKBURN

TELOMERES, THE SPECIALIZED DNA-PROTEIN STRUCTURES found at the ends of every eukaryotic chromosome, are required to stabilize chromosomes. Cytogenetic studies indicate that true terminal deletions of chromosomes, with loss of the telomere, are rare and that they lead to chromosome instabilities or progressive loss of terminal sequences from chromosome ends (*1–3*). Until quite recently, studies of the molecular mechanisms of telomere function and synthesis in higher eukaryotes were confounded by the low abundance of telomeres in their genomes. Such studies therefore were concentrated on certain lower eukaryotes, especially the ciliated protozoa, which possess short linear chromosomal DNA molecules. However, it is now known that the structure, function, and metabolism of telomeres are remarkably conserved among protozoans, fungi, slime molds, animals, and plants (*1–7*). As I will outline in this Perspective, this conservation appears to reflect the specialized manner in which telomeric DNA is synthesized.

Telomeric DNA, comprising the extreme molecular ends of chromosomes, consists of simple tandemly repeated sequences, characterized by clusters of G residues in one strand (Table 1). An overall strand composition asymmetry results in G-rich and complementary C-rich strands. The 3′ end of each strand of the duplex linear chromosomal DNA molecule is the G-rich telomeric strand, and it forms a 3′ terminal overhang, 12 to 16 nucleotides in length, protruding from the duplex (*3*). Each eukaryotic species has a characteristic telomeric repeat sequence. Limited sequence variations are found in some species (Table 1). However, widely divergent species can have the same telomeric repeat unit: for example, 5′-AGGGTT-3′ is the telomeric repeated sequence of acellular slime molds and humans. In human germline (sperm) nuclei, about 10 to 15 kb of this tandemly repeated sequence is found at every telomere (*7*), so that ~0.03% by weight of the total genome is telomeric DNA.

The enzyme telomerase is responsible for synthesis of the G-rich strand of telomeric DNA. Telomerase was first identified in the ciliate *Tetrahymena* (*8, 9*) and was shown to polymerize nucleotides into tandem repeats of the *Tetrahymena* telomeric DNA sequence, TTGGGG. Polymerization occurs by adding onto the 3′ end of a G-rich strand telomeric oligonucleotide primer, independent of an exogenously added nucleic acid template. The enzyme requires a DNA primer: it can use the G-rich strand telomeric sequences from all eukaryotes tested, but not random sequence DNA oligonucleotides (*8*). Similar findings were subsequently made for the telomerase activities of the ciliates *Oxytricha* and *Euplotes* (*10, 11*) and of human cells (*5*). Each telomerase synthesizes its species-specific G-rich strand sequence and has primer requirements similar to those of the *Tetrahymena* telomerase. Identification of telomerase activity in human cells suggests the generality of this enzyme in telomere synthesis in eukaryotes outside the ciliated protozoa.

The existence of telomerase can explain many properties of telomeres in vivo. Telomeres from one species can stabilize linear DNA molecules or chromosomes in another species, even though the two organisms have different telomeric DNA sequences (*3*). DNA sequencing showed that yeast telomeric DNA sequences are added to the end of a ciliate telomere in vivo (*12*). Human telomeres also function in yeast (*4*), demonstrating that this functional conservation is not limited to the lower eukaryotes. The basis for conservation of telomere function between distantly related eukaryotes may be the recognition properties of telomerases. In vitro, all telomerases require a minimum length (10 to 12 nucleotides) of G-rich strand telomeric DNA, similar to the length of the 3′ overhang of telomeric DNA, for high-affinity recognition as a primer in vitro (*5, 8, 10, 11, 13*). The ability of such G-rich synthetic DNA oligonucleotides to assume intra- and intermolecular two- and four-stranded folded structures stabilized by non–Watson-Crick base pairs (*14, 15*) correlates with the ability to be recognized as a primer by telomerase (*8, 13, 14*). However, the exact structure responsible for recognition by telomerase is not yet known.

A linear duplex DNA such as a chromosomal DNA is thought to require special means, other than normal semiconservative DNA replication, for completing the replication of its 5′ termini. The action of conventional DNA polymerases, which synthesize DNA in the 5′ to 3′ direction and usually require a nucleic acid primer, is expected to leave 5′-terminal gaps after each replication round (*1–3*). A central function of telomerase therefore appears to be to counterbalance this terminal DNA loss. Considerable variation in telomere length is common in eukaryotes, including humans (*3, 6*). If mean telomere length is determined by the balance between the addition of sequences of telomerase and their terminal loss through 5′ end attrition as described above, the variability of telomere lengths in

Department of Molecular and Cell Biology, University of California at Berkeley, Room 229 Stanley Hall, Berkeley, CA 94720.