Perspective

The Evolution of Catalytic Function

PHILLIP A. SHARP AND DAVID EISENBERG

MORE THAN ANY OTHER RECENT DISCOVERY, THE CATALytic potential of RNA has injected new vigor into theories of the evolution of catalytic molecules. The old conceptual difficulty of how DNA could have been replicated by some simplification of the complex machinery we know today disappears if RNA simply replicated itself.

This new concept was the subject of the 52nd Symposium on Quantitative Biology at Cold Spring Harbor, which was entitled "The Evolution of Catalytic Function." The meeting, in May 1987, included discussion on early events in evolution of biological systems, along with several other areas of research such as mechanisms of catalytic processes, protein structure at the atomic level, and the molecular biology of processes that are, or may be, RNA-catalyzed. Thus, the discovery of Thomas Cech (University of Colorado, Boulder) of RNA catalysis has already profoundly influenced thinking about the early biological world. The possibility of an RNA world where RNA structures catalyzed many processes is being considered seriously. In fact, Francis Crick (Salk Institute) foreshadowed these ideas many years ago by suggesting for RNA a catalytic role as well as a central role in evolution.

The question of the role of RNA catalysis in translation has sparked a renewed interest in the study of ribosomes. M. Nomura (University of California, Irvine) emphasized that, although much has been learned over the intervening years, it is still unclear whether RNA or protein is the prime catalytic constituent of a ribosome. The possibility of RNA catalysis is supported by (i) the highly conserved nature of the primary and secondary structures of ribosomal RNA's, and (ii) the anticodon of charged transfer RNA (tRNA) can be specifically photon crosslinked to a site in the 16*S* ribsomal RNA (rRNA). (iii) Protection against chemical modification also maps the binding of the anticodon of tRNA's to the same site on rRNA. (iv) Mutations in the rRNA in the vicinity of the tRNA binding site render ribosomes resistant to antibiotics, and the direct binding of antibiotics can be mapped to these and other sites. (v) Nonsense mutations in many ribosomal proteins are nonlethal, suggesting that these proteins enhance the rate of translation but are not essential (also discussed by H. Noller, University of California, Santa Cruz, and S. Gerbi, Brown University). The critical question of the nature of the peptidyl transferase activity remains to be determined. In the RNA world view, the ribosome is an RNA catalytic machine that aligns the anticodons of tRNA's in a complementary fashion to the codons of messenger RNA and promotes a peptidyl transferase reaction between an aminoacyl tRNA and the adjacent peptidyl tRNA.

P. A. Sharp is at the Center for Cancer Research and the Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139. D. Eisenberg is at the Department of Chemistry and Biochemistry and the Molecular Biology Institute, University of California, Los Angeles, CA 90024-1569.

Among the best studied RNA catalysts (or "ribozymes") are the self-splicing introns. These can be assigned to two groups on the basis of a common secondary and probably tertiary structure. Group I type introns, typified by the intron of *Tetrahymena* rRNA gene, behave in trans as catalysts and increase the rate of hydrolysis of a phosphodiester bond by a factor of 10^9 (Cech). This catalytic RNA probably binds a guanosine cofactor and both splice sites in a stereospecific fashion so that the free energy of the transition state between two alternative phosphodiester bonds is reduced (T. Inoue, Salk Institute). This results in a transesterification reaction. Because the group I intron has a catalytic site for cleavage and formation of phosphodiester bonds, it is possible under different reaction conditions to entice this ribozyme to act as either an RNA polymerase, endonuclease, ligase, kinase, acid phosphatase, or phosphotransferase. Thus many processes related to reproduction of the genetic information in a prebiotic RNA world could have been catalyzed by self-splicing introns of the group I type.

The group II type self-splicing introns form lariat RNA's during splicing and thus do not require a guanosine cofactor. These introns have a set of consensus type sequences and structures that can be pictured as a wheel of single-strand RNA with six spokes of duplex secondary structure radiating from it. The site of branch formation during self-splicing of group II introns is at a single adenosine residue, which bulges from the sixth or 3′-most spoke. F. Michel (CNRS, France) showed that two sets of sequences in a loop of RNA that is part of the first spoke (or 5′-most) were complementary and interacted with two adjacent sets of sequences in the 5′ exon. Suprisingly, the results of C. Peebles (Pittsburgh) suggested that sequences in the first spoke of the intron were adequate to catalyze cleavage at the 5′ splice site. However, branch formation did not accompany cleavage, suggesting that, in the absence of the natural branch site, a water molecule can substitute for the 2′-OH of the ribose.

One of the first catalytic RNA's recognized was the M1 RNA of ribonuclease P of *Escherichia coli*. This RNA accurately processes the 5′ end of tRNA's and 4.5*S* RNA's. Testing various substrate RNA's, S. Altman (Yale) has shown that no more than a 20-nucleotide (nt) stem-type structure is recognized by the ribonuclease P RNA. Surprisingly, analysis by N. Pace (University of Indiana) of a series of overlapping deletions of the ribonuclease P RNA has failed to identify a single site that is absolutely essential for activity. This suggests that the recognition and catalysis by substrate ribonuclease P might not be simple.

The smallest and perhaps simplest catalytic RNA is the "hammerhead" structure initially discerned in plant viroid and virusoid RNA's. These small RNA's are thought to replicate by a rolling circle mechanism and to self-cleave into monomer units. Both strands of viroid RNA will catalyze self-cleavage and both contain a nonpalindromic sequence with a common hammerhead structure. R. H. Symon (University of Adelaide, Australia) showed that a 52-nt RNA containing this structure will self-cleave, yielding a cyclic 2′,3′-phosphate and 5′ hydroxyl termini. An active hammerhead structure can also be reconstituted from two RNA's, one of which is only 19 nt in length (O. Uhlenbeck, University of Colorado, Boulder). The longer RNA is cleaved at a specific site when these two RNA's are transiently bound by short tracts of sequence complementarity. In the hammerhead structure a 15-nt loop of RNA that is not base-paired is formed. The RNA sequence in this loop is conserved among self-cleaving RNA's, and nucleotide changes were shown to result in a catalytically inactive ribozyme. Self-cleaving "hammerhead" structures have also been identified in RNA transcribed from a satellite DNA of the newt (L. Epstein, Florida State, Tallahassee).

It was surprising to learn that a 3.0 Å structure for a ribozyme has

already been determined. This is a phenylalanine tRNA whose structure had been determined years ago by A. Klug and colleagues (Cambridge). They noticed that diffusion of divalent lead (Pb^{2+}) into the phenylalanine tRNA crystal lattice resulted in binding of the metal ion at a specific site and cleavage of a nearby phosphodiester bond. The cleavage was site-specific in the tRNA and generated a cyclic 2′,3′-phosphate and 5′ hydroxyl termini in common with the hammerhead structure above. It is thought that the Pb^{2+} ion is positioned by binding to bases in the tRNA and generates a nucleophile (RO^-) at one specific position in the RNA chain. This results in chain cleavage by transesterification in which the 2′ nucleophile interacts with the phosphate group at the 3′ position. Uhlenbeck showed that the half of the tRNA molecule containing the heavy metal binding site is a true catalyst; it will turn over during site-specific cleavage of a substance consisting of the other half of the tRNA.

Even sequence-specific DNA binding proteins can be viewed as having been derived from an RNA world. Many such proteins contain Zn^{2+} finger domains, where a short polypeptide is thought to be held in a folded structure by the chelation of Zn^{2+} by two cysteines and two histidines appropriately spaced. This structure was first recognized in the sequence of the TFIII A protein that promotes transcription of genes encoding 5*S* RNA. The nine Zn^{2+} finger domains in this promoter protein seem to make contact with the major groove of the DNA helix at periodic lengths of 5.5 bp. This suggested to Klug that the protein was extended along one face of the DNA helix with the finger domains extending into the major groove on alternative sides of the helix. A specific model for the zinc finger as an α helix and two strands of β sheet was presented by J. Berg (Johns Hopkins); these data raised the question of whether Zn^{2+} finger domain proteins might reflect a regulatory bridge between an RNA world and a DNA world. The TFIII A protein binds specifically to 5*S* RNA and was initially purified from oocytes as a specific protein-RNA complex. Prototype Zn^{2+} finger domains have also been identified in proteins that are thought to bind RNA. Thus, this particular protein structure may have the necessary flexibility to recognize RNA as well as DNA, a flexibility that is undoubtedly not available to a helix-coil-helix motif of DNA binding proteins of bacteria.

Many of the constituents of an RNA world, such as the aromatic bases, are readily generated during electrical discharges through prebiotic-type gas mixtures. However, the amount of ribose produced during these experiments is low, and this has led to the suggestion that the original biological system might not have been based on ribose-containing moieties (L. Orgel, Salk Institute). It is unlikely that RNA could be the sole macromolecule in a membrane-enclosed, reproducing system. Thus, a protein synthesis mechanism would have had to evolve early. A. Weiner and N. Maizels (Yale) suggested that tRNA's might have evolved from the termini of replicating RNA segments. Parts of tRNA structures are found at the termini of plant virus RNA genomes (T. C. Hall, Texas A&M), and the addition of CCA to the 3′ terminus of tRNA's could reflect an earlier telomeric-type function. An extension of this hypothesis suggests that the group I RNA catalysis could have charged primitive tRNA's by a transesterification reaction with the carboxyl group of amino acids.

Principles of catalysis for protein enzymes may indeed hold for catalytic RNA and for catalytic antibodies. Catalytic antibodies have been produced in the laboratories of R. Lerner (Scripps Clinic) and P. Schultz (University of California, Berkeley); they used monoclonal antibodies to haptens that mimic the chemical properties of the transition state complex of the reaction to be catalyzed. This work is a logical outgrowth of L. Pauling's insight, 40 years ago, that enzymes have evolved to bind and stabilize the transition state of the catalyzed reaction. Pauling's insight was discussed further by W. Jencks (Brandeis), who pointed out that lowering the energy of the transition state necessarily has the implication that the energy of the ground state cannot be lowered too much, otherwise the net difference in free energy for the catalyzed reaction is no different from that for the uncatalyzed reaction. J. Knowles (Harvard) defined the ultimate goal in catalytic evolution as "enzymatic perfection." Such perfection exists when the covalent reactions on enzymes can occur at diffusion velocity rates, so that diffusion of substrate to the enzyme is the rate-limiting step. D. Koshland (University of California, Berkeley) discussed the cases of enzymes that were selected for their catalytic "imperfection," for example, reactions of kinases and phosphatases in which an enzyme with high turnover would inevitably waste more ATP in a futile cycle. In these cases, evolution would select for the lowest catalytic rate that could accommodate the regulatory role.

Another theme was the organization of larger proteins into folding domains that, in some cases in higher cells, seem to be encoded in the genome by individual exons. One example of a larger enzyme with several folding domains is bacterial glutamine synthetase. It is made up of 12 identical subunits each having two folding domains. The active site is formed at the interface of two domains on different subunits. Also having an active site at subunit boundaries is RuBisCo (ribulose bisphosphate carboxylase-oxygenase). In the structure of the dimeric form of the *Rhodospirillum rubrum* RuBisCo (C.-I. Branden, Uppsala), the active site is at the surface of an α–β barrel domain adjacent to a second domain. Branden noted that this α–β barrel domain with eight alpha helices surrounding a cylindrical core of eight strands of beta sheet has been found in ten enzymes. In tobacco RuBisCo (D. Eisenberg, UCLA) there are 16 subunits including four pairs of dimers of large subunits. Both glutamine synthetase and RuBisCo are primary biological catalysts in that they catalyze the entry of nitrogen and carbon into the metabolic pathway. The observation that their active sites are shared between subunits shows they must have evolved from simpler ancestors.

Among the multisubunit proteins described was hemoglobin [M. L. Perutz (Cambridge), who presented a movie illustrating the structural changes during oxygenation], phospholipase A_2 (P. Sigler, Chicago), periplasmic binding proteins (F. A. Quiocho, Baylor), and the Klenow fragment of *E. coli* DNA polymerase I (T. Steitz, Yale). The Klenow fragment has two active sites, one for polymerase activity and one in a separate domain for 3′,5′-exonuclease activity. Four periplasmic binding proteins (for arabinose, sulfate, or galactose, and for Leu, Ile, or Val) sequester their ligand within a cleft between domains. What is suprising here is that the doubly charged sulfate ion is completely shielded from solvent with no nearby counterions or salt linkages. The sulfate is bound by hydrogen bonds, which must provide electrostatic stabilization through ion-dipole forces. All four binding proteins have structures similar to each other, but not much amino acid sequence similarity.

How amino acid sequence dictates the three-dimensional structure of a protein remains an active field of inquiry. Improved capacity to synthesize peptides has opened the way for preparation and characterization of small proteins designed to fold in a given way. Quite small peptides can exist as independent folding units and can associate noncovalently to form protein-like aggregates. R. Baldwin (Stanford) described experiments on 17-residue α helices indicating that helix stabilization comes from both interaction of side-chain charges with the intrinsic helix dipole and from side-chain charges with each other. Amphiphilic helices were discussed by W. DeGrado (E. I. du Pont) who prepared synthetic helices that form noncovalent complexes with each other to create four α-helical

(continued on page 807)

Table 2. Frequency of *Ac*-containing regenerated (R_0) plants among total R_0 plants from *Ac*-containing embryo cell lines 4-41 and 1-42.

Embryo cell line	Number with *Ac* activity/total R_0 plants tested	Percent
4-41	9/26	34.6
1-42	1/7	14.3

species have demonstrated the occurrence of both chromosomal abnormalities and heritable phenotypic changes. This study has revealed a third type of change, the activation of transposable elements. Researchers who have observed both chromosomal alterations and phenotypic mutants in their materials have generally been unable to directly correlate these occurrences. Activation of transposable elements by chromosome breakage may provide a link between the chromosome aberrations and mutations. A transposable element, once released, could insert into a gene locus rendering it inactive or otherwise altered. Elucidation of the mechanisms governing somaclonal variation may enable researchers to better control the amount, and perhaps the type, of variability produced by the in vitro culture of plants.

REFERENCES AND NOTES

1. S. Edallo, C. Zucchinali, M. Perenzin, F. Salamini, *Maydica* **26**, 39 (1981).
2. G. Benzion, thesis, University of Minnesota, St. Paul (1984).
3. M. Lee and R. L. Phillips, *Genome* **29**, 122 (1987); *ibid.*, in press.
4. C. L. Armstrong and R. L. Phillips, *Crop Sci.*, in press.
5. R. I. S. Brettell, R. Thomas, D. S. Ingram, *Theor. Appl. Genet.* **58**, 55 (1980); P. F. Umbeck and B. G. Gengenbach, *Crop Sci.* **23**, 584 (1983).
6. C. A. Rhodes, R. L. Phillips, C. E. Green, *Can. J. Genet. Cytol.* **28**, 374 (1986).
7. C. E. Green, R. L. Phillips, A. S. Wang, *Maize Genet. Coop. News Lett.* **51**, 53 (1977).
8. P. J. Larkin and W. R. Scowcroft, *Theor. Appl. Genet.* **60**, 197 (1981).
9. B. McClintock, *Proc. Natl. Acad. Sci. U.S.A.* **36**, 344 (1950).
10. ———, *ibid.* **25**, 405 (1939); *ibid.* **28**, 458 (1942).
11. M. G. Neuffer, *Genetics* **53**, 541 (1966); E. B. Doerschug, *Theor. Appl. Genet.* **43**, 182 (1973).
12. A. Bianchi, F. Salamini, R. Parlavecchio, *Genet. Agrar.* **22**, 335 (1969).
13. N. Fedoroff, S. Wessler, M. Shure, *Cell* **35**, 235 (1983).
14. B. McClintock, *Stadler Genet. Symp.* **10**, 25 (1978).
15. ———, *Science* **226**, 792 (1984).
16. B. Burr and F. Burr, *Stadler Genet. Symp.* **13**, 115 (1981).
17. Three hundred and one regenerated plants from 94 embryo cell lines were represented in the present study. Seed from 161 R_0 plants was obtained from Lee and Phillips (*3*), who initiated organogenic (Type I) callus cultures with immature embryos produced by sib pollinations within F_2 rows of the cross Oh43 × A188. Ninety R_0 plants or their progeny (from 41 embryo cell lines) were obtained from Armstrong and Phillips (*4*); the materials were initiated from lines derived from A188 × B73 crosses, and both organogenic (Type I) and embryogenic (Type II) cultures were represented. Seed from 50 R_0 plants representing 29 embryos was obtained from H. Hartloff [thesis, University of Minnesota, St. Paul (1984)], who had initiated Type I cultures from the inbred A188.
18. H.-P. Döring and P. Starlinger, *Cell* **39**, 253 (1984).
19. B. McClintock, *Cold Spring Harbor Symp. Quant. Biol.* **16**, 13 (1951).
20. N. W. van Schaik and R. A. Brink, *Genetics* **44**, 725 (1959); E. R. Orton, *ibid.* **53**, 17 (1966); E. M. Nowick and P. A. Peterson, *Mol. Gen. Genet.* **183**, 440 (1981).
21. S. V. Evola, F. A. Burr, B. Burr, abstract, Eleventh Annual Aharon Katzir-Katchalsky Conference, Jerusalem, 8 to 13 Janaury 1984; S. V. Evola, A. Tuttle, F. Burr, B. Burr, abstract, First International Congress of Plant Molecular Biology, Savannah, GA, 27 October to 2 November 1985, p. 10.
22. R. W. Groose and E. T. Bingham, *Crop Sci.* **24**, 655 (1984); *Plant Cell Rep.* **5**, 104 (1986).
23. A. Pryor *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6705 (1980).
24. S. S. Johnson, R. L. Phillips, H. W. Rines, *Genome* **29**, 439 (1987).
25. A. Lima-de-Faria, in *Handbook of Molecular Cytology*, A. Lima-de-Faria, Ed., vol. 15 of *Frontiers of Biology* (North-Holland, Amsterdam, 1969), pp. 277–325.
26. M. D. Sacristan, *Chromosoma* **33**, 273 (1971).
27. T. J. McCoy, R. L. Phillips, H. W. Rines, *Can. J. Genet. Cytol.* **24**, 37 (1982).
28. P. S. Chomet, S. Wessler, S. L. Dellaporta, *EMBO J.* 6, 295 (1987); D. Schwartz and E. Dennis, *Mol. Gen. Genet.* **205**, 476 (1986).
29. V. L. Chandler and V. Walbot, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1767 (1986).
30. V. M. Peschke, thesis, University of Minnesota, St. Paul (1986).

31. We thank M. Lee, C. L. Armstrong, and H. H. Jessen for providing the regenerated plant materials used in this study, and B. McClintock and E. H. Coe, Jr., for the transposable element tester stocks. This work was supported by the Pillsbury Company and by U.S. Department of Agriculture grant USDA/85-CRCR-1-1683. Paper No. 15265, Scientific Journal Series, Minnesota Agricultural Experiment Station.

28 January 1987; accepted 31 July 1987

(continued from page 730)

bundles similar to those that occur naturally. A new computational approach to protein folding was described by F. Richards (Yale), who determined all possible amino acid sequences compatible with a given protein backbone fold and found relatively few sequences that can fit a given backbone conformation.

Another difficult question is how amino acid replacements in a given protein change the structure over evolutionary time. This question was addressed by C. Chothia and A. Lesk (Cambridge, United Kingdom) by comparing structures of families of proteins such as the globins and the immunoglobulins. They find that α helices change their directions by a few degrees and that β sheets experience local changes in conformation, with atoms moving up to a few angstroms. However, the structural changes in different regions of each protein tend to be coupled to maintain the geometry at the binding sites. These examples of how protein structures can adapt to amino acid replacements give some insight into the capacity of proteins for evolutionary change.

A striking example of convergent evolution of lysozymes was presented by C.-B. Stewart (University of California, Berkeley). Sequence data were used to suggest that the bacteriolytic lysozyme c was recruited twice during evolution as a digestive enzyme in the stomachs of foregut-fermenting mammals, occurring once in the cow and once in the tree-dwelling, leaf-eating langur monkey. Analysis of evolutionary trees suggests that after foregut fermentation arose in monkeys, the langur stomach lysozyme gained sequence similarity to the cow stomach lysozyme, permitting it to adapt to the acidic environment of the stomach fluid.

Another session dealt with the construction of evolutionary trees based on sequence information. G. Olsen (Indiana) groups thermophilic, methanogenic, and halophilic archaebacteria together—to the exclusion of eukaryotes and eubacteria. On the basis of unequal rates of sequence change, J. Lake (UCLA) suggests that the eukaryotic RNA genes evolved from those of sulfur-metabolizing prokaryotes.

The correlation between protein domains and exons in genes of contemporary organisms was emphasized by W. Gilbert (Harvard) in hypothesizing that genes evolved in an RNA world as RNA's containing single exons that encoded short functional polypeptide domains. Genes then evolved by recombination into DNA. This scenario implies that introns are ancestral remnants of primordial genes and that the general absence of introns in the genes of prokaryotes reflects their loss. This is contested by others (T. Cavalier-Smith, Kings College, London), who propose that introns were inserted into prokaryotic-type genes. At the heart of this problem is the question of the evolutionary lineage relations between eukaryotes and prokaryotes (F. W. Doolittle, Dalhousie).

The many deep mysteries concerning evolution of catalytic function include how enzymes arose, how amino acid replacements affect structure and function, and how present-day catalysts are descended from more primitive ones. However, the recent discovery of RNA catalysis has provided at least one of the missing links and it gives hope that we are on the path to solving some of the questions obscured by the 3.5 billion years of birth, change, and death since the emergence of cells.