9. Herrmann, W. A., *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 1297.
10. Harman, W. D. and Taube, H., *J. Am. Chem. Soc.*, 1990, **112**, 2682.
11. Harman, W. D. and Taube, H., *J. Am. Chem. Soc.*, 1988, **110**, 7555.
12. Harman, W. D., Hasegawa, T. and Taube, H., *Inorg. Chem.*, 1991, **30**, 453.
13. Ma, Y., Demou, P. and Faller, J. W., *Inorg. Chem.*, 1991, **30**, 62.
14. Faller, J. W. and Ma, Y., *Organometallics*, 1988, **7**, 559.
15. Hermann, W. A., Fischer, R. A. and Herdtweck, E., *J. Organomet. Chem.*, 1987, **329**, C1.
16. Bottomley, F. and Sutin, L, *Adv. Organomet. Chem.*, 1988, **28**, 339.
17. Bonnemann, H. and Brijoux, W., *New. J. Chem.*, 1987, **11**, 549.
18. Hermann, W. A., Fischer, R. A. and Herdtweck, E., *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 1509.
19. Hermann, W. A., Marz, D., Herdtweck, E., Schafer., Wagner, W. and Kneuper, K-J., *Angew. Chem., Int. Ed. Engl.*, 1987, **26**, 462.
20. Samuelson, A. G., unpublished results.
21. Luche, J. L., Allavena, C., Petrier, C. and Dupuy, C., *Tetrahedron Lett.*, 1988, **29**, 5373.
22. Wilson, S. R. and Guazzaroni, M. E., *J. Org. Chem.*, 1989, **54**, 3087.
23. Chan, T-H. and Li, C-J., *Organometallics*, 1990, **9**, 2649.
24. Baruah, J. B. and Samuelson, A. G., *J. Chem. Soc. Chem. Commun.*, 1987, 36.
25. Baruah, J. B. and Samuelson, A. G., *J. Org. Met. Chem.*, 1989, **361**, C57.
26. Schrock, R. R., *Acc. Chem. Res.*, 1986, **19**, 342.
27. Novak, B. M. and Grubbs, R. H., *J. Am. Chem. Soc.*, 1988, **110**, 7542.
28. Novak, B. M., *J. Am. Chem. Soc.*, 1991, **113**, 5065.
29. Demming, T. J. and Novak, B. M., *Macromolecules*, 1991, **24**, 326.
30. Demming, T. J. and Novak, B. M., *Macromolecules*, 1991, **24**, 6043.
31. Braca, G., Gulletti, A. M. and Sbrana, G., *J. Org. Met. Chem.*, 1991, **417**, 41.
32. Joh, T., Doyama, K., Fujiwara, K., Maeshima, K. and Takahashi, S., *Organometallics*, 1991, **10**, 508.
33. Collman, J. P., Hegedus, L. S., Norton, J. R. and Finke, R. G., *Principles and Applications of Organotransition Metal Chemistry*, University Science Books, Mill Valley California, USA, 1987, pp. 631–632.
34. Anderson, J. R., Campi, E. M. and Jackson, W. R., *Catal. Lett.*, 1991, **9**, 55.

RNA world and ribosomes

D. P. Burma

Molecular Biology Unit, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221 005, India

Earlier and recent evidences in favour of RNA as the original genetic material have been documented. Through this the diverse functions of RNAs have been focused on. That the function of ribosomal RNAs in protein synthesis will be the final word for the 'RNA world' has been emphasized through the recent observations of our own and other laboratories.

THE discovery of 'nuclein' by Miescher[1] in 1868 and the proposition of the double helical structure of DNA by Watson and Crick[2] in 1953 are two landmarks of life sciences in the nineteenth and twentieth centuries respectively. The series of discoveries that followed Watson and Crick's fundamental breakthrough finally established the DNA world where the master molecule plays the key role. Some breakthroughs in the recent past, specially the establishment of RNAs acting as catalysts[3] (so-called ribozymes) initiated rethinking in the area. Could it be that the DNA world developed at a later stage, subsequent to the ushering in of the RNA world? The latter term was coined first by Walter Gilbert in 1986 for obvious reasons and is not to be dismissed so early although it is disliked by many, for example Joyce[4]. The arguments in favour and against have led to healthy discussion and 'RNA world' terminology is being tossed about more and more instead of being discarded. It is not easy to look back into the history of the earth since its birth about billions of years ago and a lot of speculation has to be made to establish one theory or the other. Since there is no other alternative we have to take recourse to the present-day information to extrapolate to the past to have an idea about the primordial situation. That is the exercise I plan to do here.

The mechanism of synthesis of DNA is well-established although the same cannot be said of some of the intricacies[5]. However, it is safe to conclude at this stage that it is practically impossible for nature to synthesize DNA today without the help of RNA ingredients. DNA and RNA, as the names signify, differ primarily in the nature of the sugar in their structure and that leads to major differences. The 2'-OH present in RNA enables it to take part in a number of reactions and also leads to its basic instability. The absence of this OH imposes restrictions on the capabilities of DNA but provides remarkable stability to the genetic information it carries. The switch-over from RNA world to DNA world, if any, can be strongly justified on this ground only.

It is difficult to push a similar argument for the exchange of the base uracil in RNA with thymine in DNA. The advantage of introducing a methyl group in uracil does not appear to make much of a difference. The widespread occurrence of thymine in RNA as well

makes one think that this base substitution is not highly critical. However, this unique base in DNA is retained throughout the living world. If we look at the modes of synthesis of these two constituents, deoxyribose and thymine of DNA, its origin from RNA ingredients becomes more than obvious. This will be clear from the reactions shown in Figure 1.

Deoxyribonucleotide ingredients of DNA are not directly produced like ribonucleotides of RNA but have to be synthesized from ribonucleotides by reduction at the diphosphate level. Similarly thymidylate is produced from uridylate by the methylation reaction. Direct synthesis of thymidylate as in case of uridylate is not known. Finally it is not possible to synthesize DNA without RNA primer. Synthesis of Okazaki fragments, intermediates in DNA synthesis, is primed by RNA. These reactions of the present-day world can be justifiably fitted into the past history, rather picture, to conclude that RNA evolved first as the genetic message and then the necessity arose to produce a similar but more stable molecule from the ingredients that were already available by diverting those to the synthesis of a new material without synthesizing entirely new ingredients. Although it is far-fetched thinking, it is still not untenable that if genetic information originally evolved as RNA, an unstable material, there was

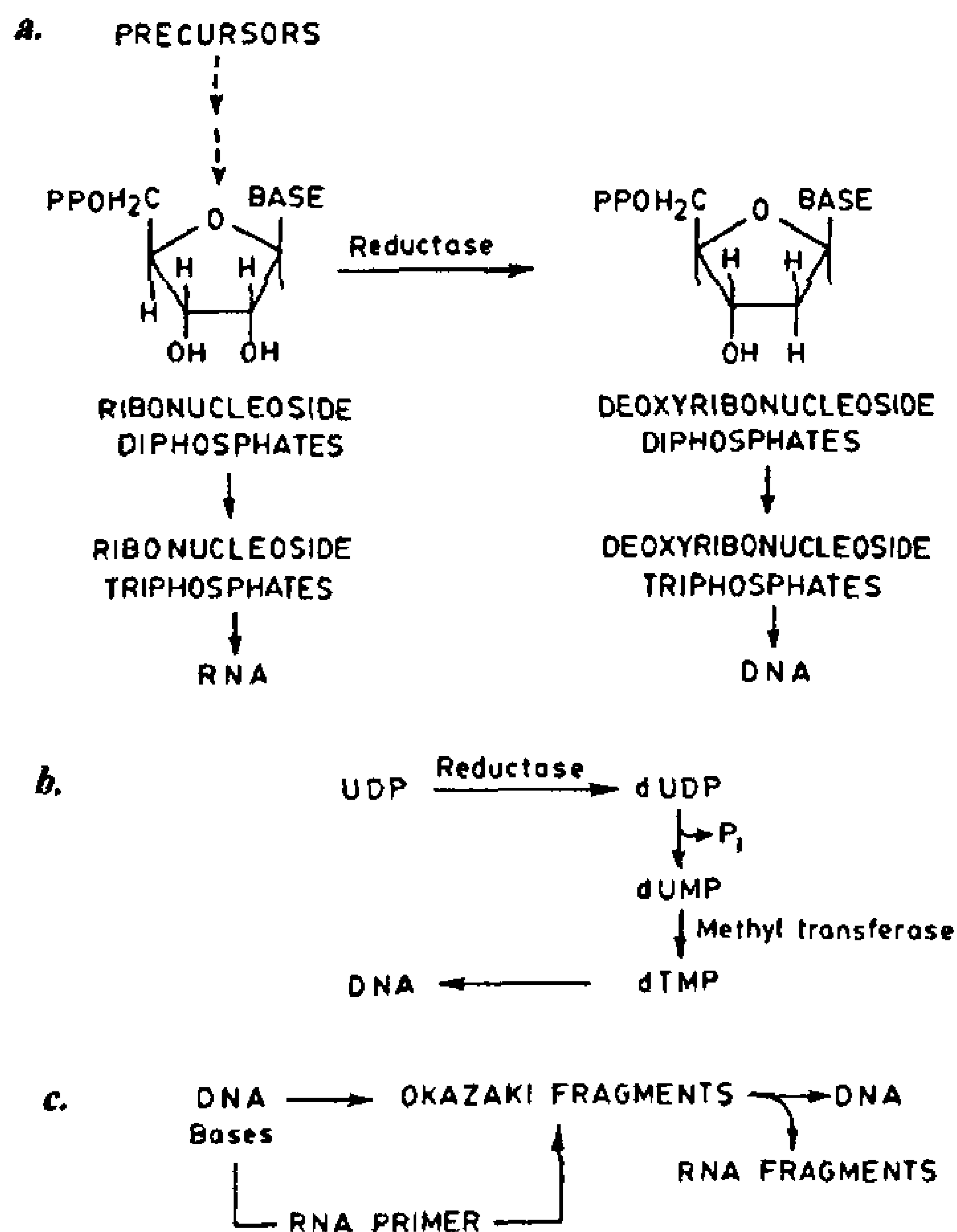

Figure 1. Dependence of DNA synthesis on RNA and RNA components.

impelling reason to evolve a more stable molecule similar to RNA and the 2'-OH group of RNA was the target. The other change, the change of a base (uracil to thymine) is not readily understood but the necessity to keep a separate tag for the newly evolved molecule can be reconciled with readily.

The first RNA species for which function became known is transfer RNA[6]. It was discovered as part of the cell extract required for protein synthesis and designated as soluble RNA. The name was changed when its function of transfer of amino acids from soluble pool to the sites of protein synthesis which were eventually termed as ribosomes (ribonucleoprotein particles of microsomes) became known. The primary structure of one of the tRNAs (specific for alanine) was worked out by Holley and coworkers[7] and the secondary structure (clover leaf model) was built on the basis of that. It took quite some time before its tertiary structure was worked out by X-ray diffraction technique[8]. That ribosomes are composed of both RNA and protein was known from the time of discovery of the particles. The complete sequence of both the RNAs (16S and 23S RNAs) present in *E. coli* became known through the work of Noller and his coworkers[9,10]. Secondary structures based on base pairing was worked out almost at the same time. However, the tertiary structures are still largely unknown although a few important clues have already been obtained[11,12]. The primary and secondary structures of 5S RNA which is part of the larger subunit of *E. coli* ribosomes and slightly larger in size than transfer RNAs have also been worked out but its complete three-dimensional structure is yet to emerge. It should be mentioned here that another small molecular weight RNA (5.8S RNA) besides 5S RNA occurs in eukaryotic ribosomes. The latter contain slightly larger size RNAs (18S and 28S) than prokaryotic ones. The 5.8S RNA is thought to be equivalent to 5S RNA of *E. coli* in function although its actual function is not yet known.

Several approaches including the search for an intermediary between DNA residing in the nucleus and the site of protein synthesis in the cytoplasm, the famous Jacob–Monod concept based on genetic data, discovery of an enzyme capable of synthesizing RNA on a DNA template and finally breaking of the genetic code culminated in the recognition of messenger RNA as conceived by Jacob and Monod[13]. The original notion that mRNA is structureless had to be abandoned soon. Further, several new concepts in terms of capping, tailing, splicing, processing, etc., specially of this species of RNA, led to better understanding of a number of complexities in the biological system. However, it was very soon realized that these (tRNA, rRNA and mRNA) are not the only species of RNA of functional importance but a larger number are yet to be recognized.

There were two central dogmas in biochemistry and molecular biology. The first one, DNA makes RNA and RNA makes protein, had to be revised when reverse transcriptase was discovered and it became known that RNA can pass on its message to DNA. The second dogma, enzymes are invariably proteins, took a long time for revision. There was considerable justification for that. When the first enzyme (urease) was claimed to be a protein some did not believe it but eventually all the enzymes discovered were invariably found to be protein molecules. The variety of amino acids (20) and their divergent nature made them suitable for numerous catalytic functions. Acidic, basic, hydrophilic, hydrophobic groups, etc. are the attributes of natural catalysts. The distinction between chemical catalysis and biological catalysis lies in the macromolecular structure of the latter. Other than the functional groups the remaining structure plays a very significant role in the recognition of the reactants, intermediate complex formation as well as product formation. Among all the naturally occurring molecules proteins are most suitable for the purpose. Therefore it was no wonder that the biochemists could never think of any other type of molecule acting as an enzyme.

Both DNA and RNA have some basic as well as acidic (phosphate) groups although the variety is less (4 basic and one acidic group) but that does not deprive them of the catalytic power. RNA has an extra fillip over DNA in its 2'-OH group which can be involved in nucleophilic reactions. It was Altman[14] who had the first evidence that RNA can act as an enzyme. He was working with RNase P which is present in both prokaryotes and eukaryotes and is responsible for the production of mature tRNAs from precursor tRNAs. By this way the mature 5'-terminus of RNA molecule is produced. Under physiological condition the enzyme exists as an RNA-protein complex. Altman showed for the first time that RNA is capable of catalysing the reaction and that the protein component has no such ability. However, this work raised doubts about the central dogma and it was widely believed that his RNA was contaminated with some protein. Eventually, however, it was unequivocally shown with the recombinant RNA product that RNA is the catalytic unit and the protein component enhances its activity considerably. The RNA component of RNase P of *E. coli* is 377 residues long and is known as M1RNA (Figure 2). The RNase P reaction is naturally dependent on structure of RNA and a divalent cation (Mg^{++} or Mn^{++}) and does not require any energy input. It is believed that a hydroxyl ligand of Mg^{2+} acts as a general base responsible for catalytic activity. Although protein component of RNase P is dispensable *in vitro* its presence is essential under *in vivo* conditions. It is believed that the main function of the protein is to act as an electrostatic shunt for catalytic RNA.

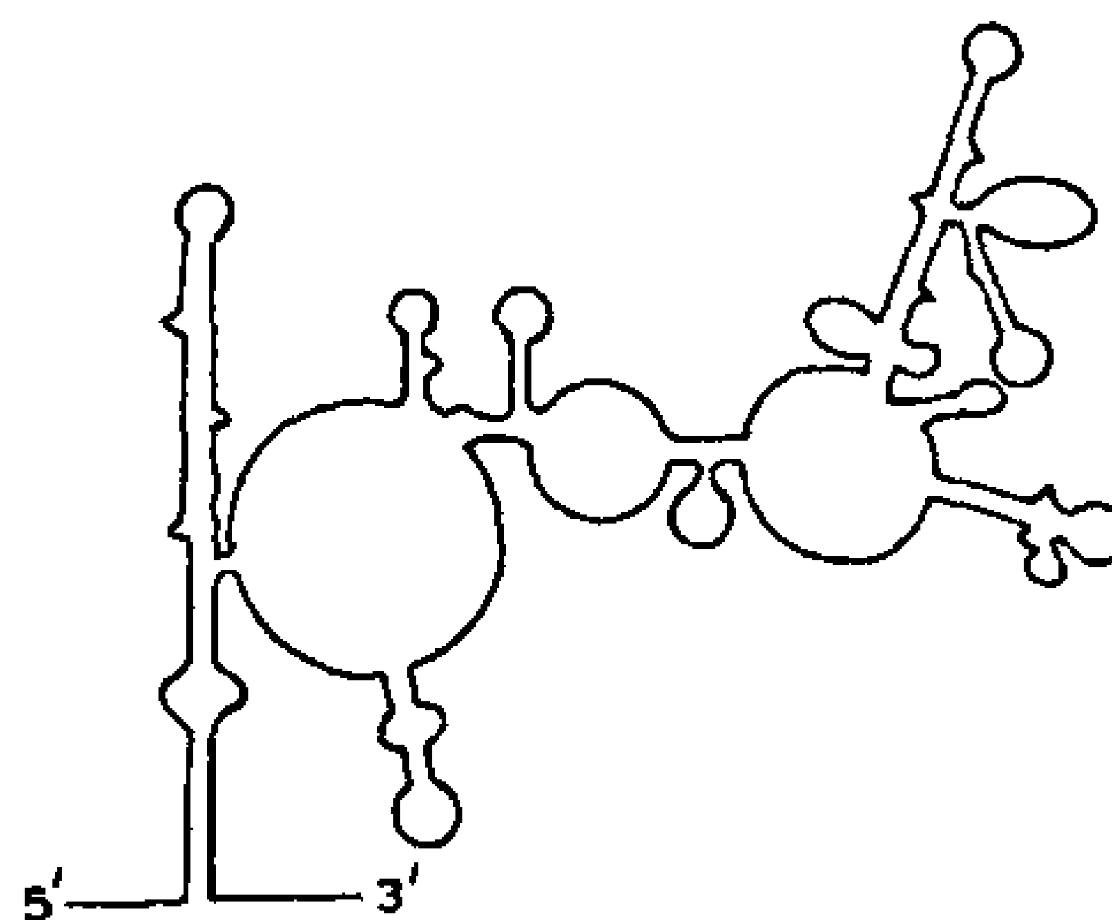

Figure 2. Secondary structure in outline (omitting the bases) of M1 RNA (catalytic component of RNase P) of *S. typhimurium*. From Watson, J. D. *et al.*, *Molecular Biology of the Gene*, The Benjamin/Cummings Publishing Co., Inc.

The other startling observation made in 1981 and 1982 by Cech[15] and his colleagues is a turning point in the history of RNA research. While studying RNA from the protozoan *Tetrahymena thermophila* they found that precursor ribosomal RNA can splice itself and this was a newer system than the widely known splicing process. It was first demonstrated by Philip Sharp and coworkers[16] and another group[17] that messenger RNAs are originally transcribed in a form larger than the actual messages used in translation. The regions which are actually responsible for coding the message are known as exons whereas those which separate exons and are normally expected not to have any message are known as introns. In a complicated process catalysed by enzymes which are part of spliceosomes[18,19] the introns are cut out and the exons are joined together to form the mature messenger RNA which is used for translation at the ribosomal site. It is widely believed and more or less established by working with organelle DNAs (mitochondrial and chloroplast DNAs) that the splicing process involves lariat formation. Actually introns are divided into two classes (introns I and II) depending on the specific sequences present in them which are responsible for organization of the RNA in a proper form to be spliced. One such intermediate is formed by the nucleophilic attack from a base at the 3' region of the intron of the splice junction at the 3'-end of exon I (on the 5' side) (Figure 3*a*). This leads to the cutting at the site and a lariat formation (intron itself) at the 5' end of exon II (on the 3' side). Subsequently the lariat is removed by the nucleophilic attack (transesterification process) by the newly produced 3'-OH group of exon I at the juncture of lariat and exon II. The lariat is removed and exons I and II are joined together by phosphate linkage. The whole process is enzyme-catalysed and dependent on certain sequences in RNAs which in turn are responsible for the organization of RNA before splicing.

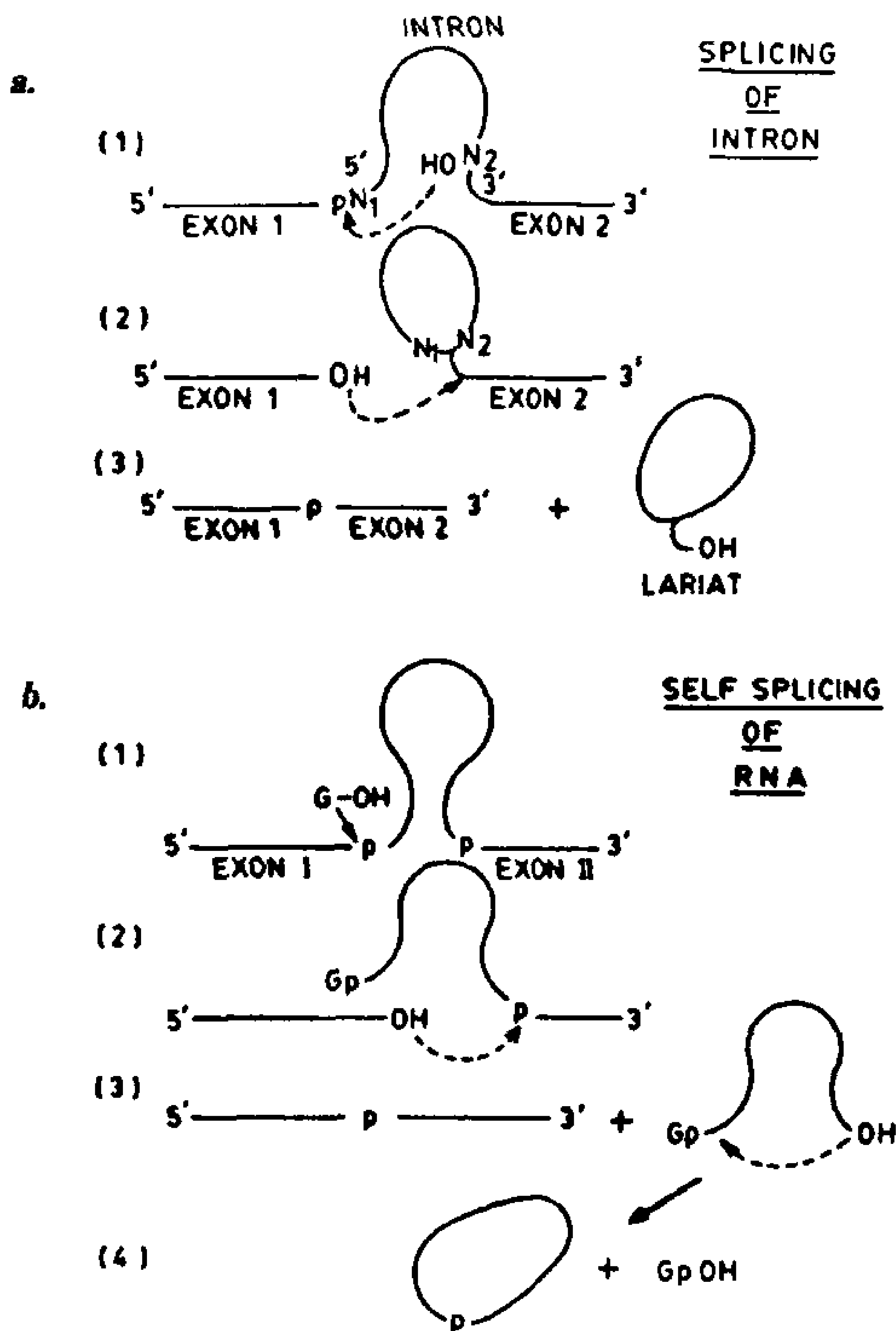

Figure 3. *a*, Mechanism of splicing of precursor messenger RNA by transesterification: *b*, Mechanism of self-splicing of *Tetrahymena* ribosomal RNA.

Tetrahymena's precursor ribosomal RNA does the splicing job in a slightly different way[3,15]. It does not require the help of any outside enzyme (protein) and does the job itself (self-splicing). However, it requires the help of an outside small molecule guanosine (or GMP or GDP or GTP). Guanosine initiates the nucleophilic attack at the splice junction between exon I and intron and attaches itself to the 5'-end of the intron by transesterification (Figure 3*b*). In the second step the OH group at the end of exon I attacks at the junction of G-attached intron with exon II (as described for the enzyme-catalysed splicing) and the two exons are joined by transesterification with the release of the intron which cyclises by a further nucleophilic attack from the 3'-OH of the intron at a site near the 5'-end of the intron. A 15-nucleotide stretch (with G at the start) is released and a circular molecule (designated as intervening sequence, IVS) is formed. This also undergoes further modification. The overall splicing mechanism appears to be the same but without the intervention of a protein and therefore the term 'ribozyme' was introduced. The major objection raised at that time was that self-splicing RNA acts only once and as soon as mature RNA is formed the catalytic activity disappears. Later Cech and his coworkers demonstrated that a slightly modified form of the same RNA can catalyse the assembly of RNAs other than splicing itself and therefore acts as true catalyst. In this case also guanosine is required which makes it a G-dependent transesterification process.

The introns, as already mentioned, could be of two types, groups I and II[18,19]. In another way the introns can be classified into four groups, one group made up of the transcripts processed in the nucleus to yield tRNAs. A second is made up of the molecules that are also processed in the nucleus and yield mature mRNA. The other categories include RNAs from diverse sources. Group I includes an assortment of tRNA, rRNA and mRNA introns in mitochondria and chloroplasts (cytoplasmic organelles). Group II introns which mainly exist in fungal mitochondria and plant chloroplasts are less common. Detailed investigations draw attention to the core three-dimensional structure of introns present in all these and its responsibility for their catalytic activity (Figure 4). Apparently the three-dimensional organization is the backbone of the self-splicing and normal introns. As in proteins, this catalytic centre is responsible for the splicing of RNA. Ribozymes in all respects have been found to behave as

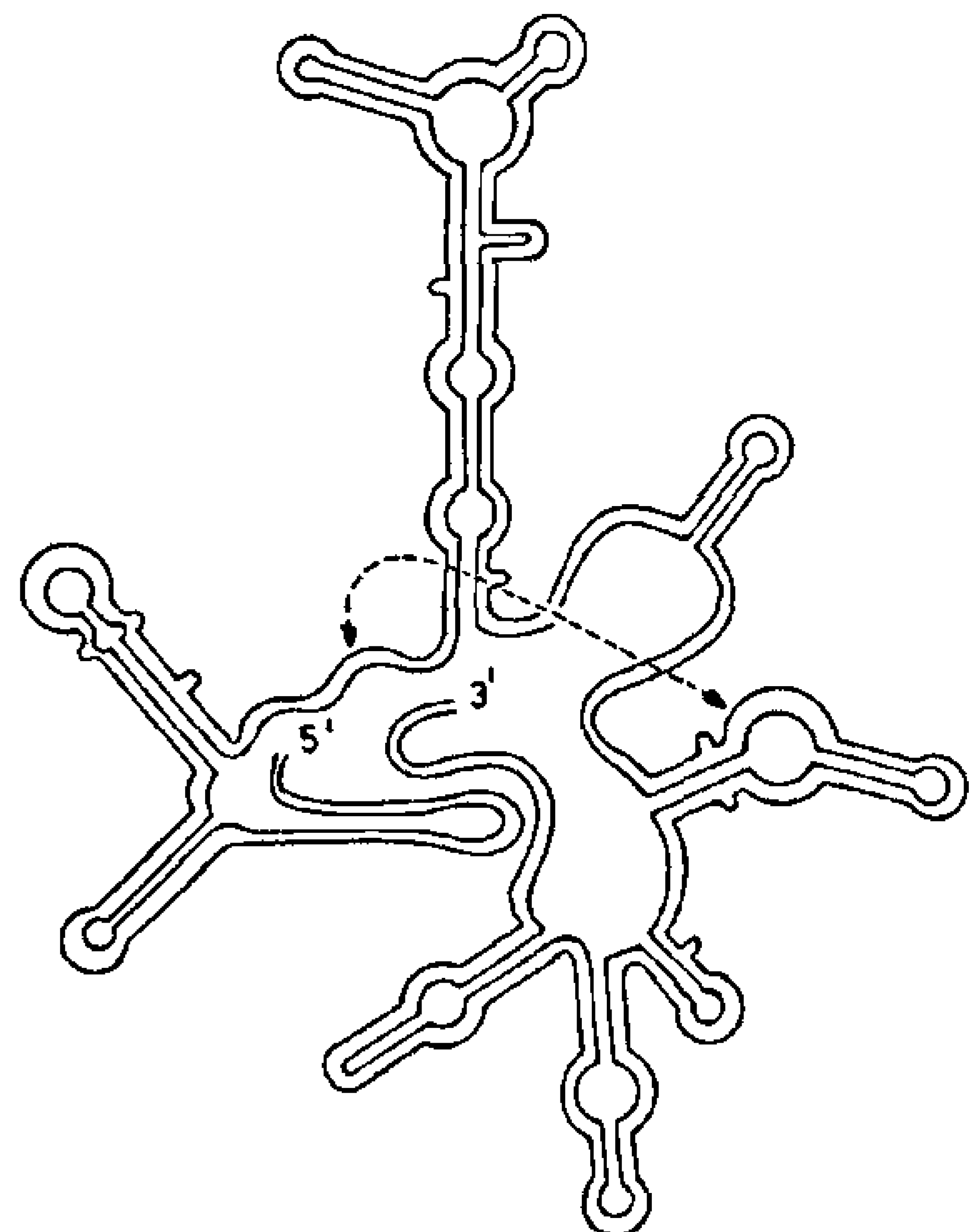

Figure 4. Secondary structure in outline of the catalytic centre of *Tetrahymena* self-splicing ribosomal RNA intron (omitting the bases), adapted from ref. 15. The arrow indicates the bases in close proximity in the three-dimensional organization.

enzymes. The only question which remains unanswered is whether these were the first catalysts (some relics are detectable even today) and diversified their functions by passing the responsibility to proteins which were evolving independently or had actually evolved to take up this job from RNAs. That remains an area of wild speculation.

It has already been mentioned that spliceosomes occurring in the nucleus are responsible for splicing of precursor RNAs. Actually they are very complicated structures and their mechanism of action is largely unknown. Among the components occurring in these organelles the small nuclear RNAs (SnRNAs) designated as U1, U2, U4, U5 and U6 are well-studied. Small nuclear RNAs are stable, relatively abundant RNA molecules found in the nucleus of a wide variety of eukaryotes and range in size from 90 to 220 nucleotides. Further, their sequences are highly conserved throughout evolution. SnRNAs actually exist in the cell associated with proteins as small nuclear ribonucleoproteins (SnRNPs pronounced as 'Snurps'). In the splicing reaction the SnRNPs, other essential protein factors and the premRNA form a macromolecular complex which is known as spliceosome. In the electron microscope they appear as 40 to 60 nm particles with a distinctive morphology.

SnRNPs occur in at least three morphologically and biochemically distinct forms in the amphibian oocyte nucleus[20]. These are referred to as A, B and C snurposomes. The A snurposomes are the simplest and contain only U1 SnRNA and U1-associated proteins. It may act as storage for U1RNP. The B snurposomes contain all five splicing SnRNAs. The snurposomes vary widely in size and usually have one to many B snurposomes embedded in their surface. The storage and recycling of SnRNPs may be the functions of B and C snurposomes as well. Finally SnRNPs, non SnRNP-splicing factors and other proteins associate with preRNAs in the form of heterogeneous nuclear ribonucleoprotein particles (hnRNAs).

Although the actual role of SnRNAs in splicing reaction is not known, some of the preliminary data and parallel observations in other systems have helped to understand the role of RNA in splicing[21] (see Figure 3). It is well-established that there are certain specific base sequences in splice junctions which help precursor RNAs to assume the requisite three-dimensional structure. One function of SnRNAs would be to recognize these sequences and help the precursor RNA to fold into the required structure. The other speculation is that SnRNAs play a direct role in splicing reaction as the RNA component of RNase P does. If this is true, the protein components would have the minor role of keeping SnRNAs in proper three-dimensional structure for the display of enzyme activity as is speculated in case of ribosomes (to be discussed in detail later). If this is found true the emergence of the RNA world prior to the DNA world will be further strengthened.

The above are not the only examples of diverse functions of RNAs. The more the complex physiological assemblies are looked into, more are the different types of functions come across. One of the early examples is the 7S RNA (ref. 22) present in signal recognition particles which are responsible for protein translocation across the lipid bilayer membrane of the endoplasmic reticulum[23]. SRP recognizes the leader sequence of the nascent polypeptide chain as it emerges from the ribosome. The interaction leads to a translational arrest that is released only if the complex is targeted to a receptor in the membrane. The SRP was originally isolated as an 11S complex consisting of six different polypeptide chains from a salt-wash of membrane residues. Eventually this complex was found to contain 7S RNA. The structure of 7S RNA containing ~250 nucleotides has been worked out. It is not known whether 7S RNA is simply a structural element of SRP or whether it plays an important functional role. Only future can tell whether 7S RNA is a product of the RNA world. Similarly an RNA-containing protein has been reported to modulate the expression of *c-myc* gene[23a].

The specialized structures present at the ends of the eukaryotic linear chromosomes are known as telomeres[24]. It was known for quite sometime that telomeres undergo rapid turnover. The understanding of the structure of DNA at the telomeric ends has helped a lot to understand the mechanism of their turnover. The telomeric sequences stand out from the internal DNA sequences and their primary sequences and functions are entirely different. Telomeres are actually DNA-protein complexes, protein being involved in protection of the end. Telomeric DNA is highly conserved and has a unique tandem repeat sequence with one strand containing clusters of G residues. This repeat may vary from 100 to 1000 base pairs. G-rich strand is oriented 5′ to 3′ toward chromosomal terminus and protrudes 12–16 nucleotides beyond the C-rich strand. Actually a protecting protein binds to the protruded chain. With the discovery of the enzyme telomerase the mechanism of this turnover has become known. This is a novel mechanism due to the built-in RNA template in the enzyme itself (for general discussion see ref. 25). The template has a complicated RNA structure but a sequence complementarity to the G-rich overhanging strand. Whenever there is a damage to this strand the enzyme fixes the template according to the complementary base sequence of the damaged part and extends the DNA chain guided by the template (Figure 5). This is a reverse transcriptase activity and quite different from regular DNA polymerase activity or even reverse transcriptase activity both of which use template which

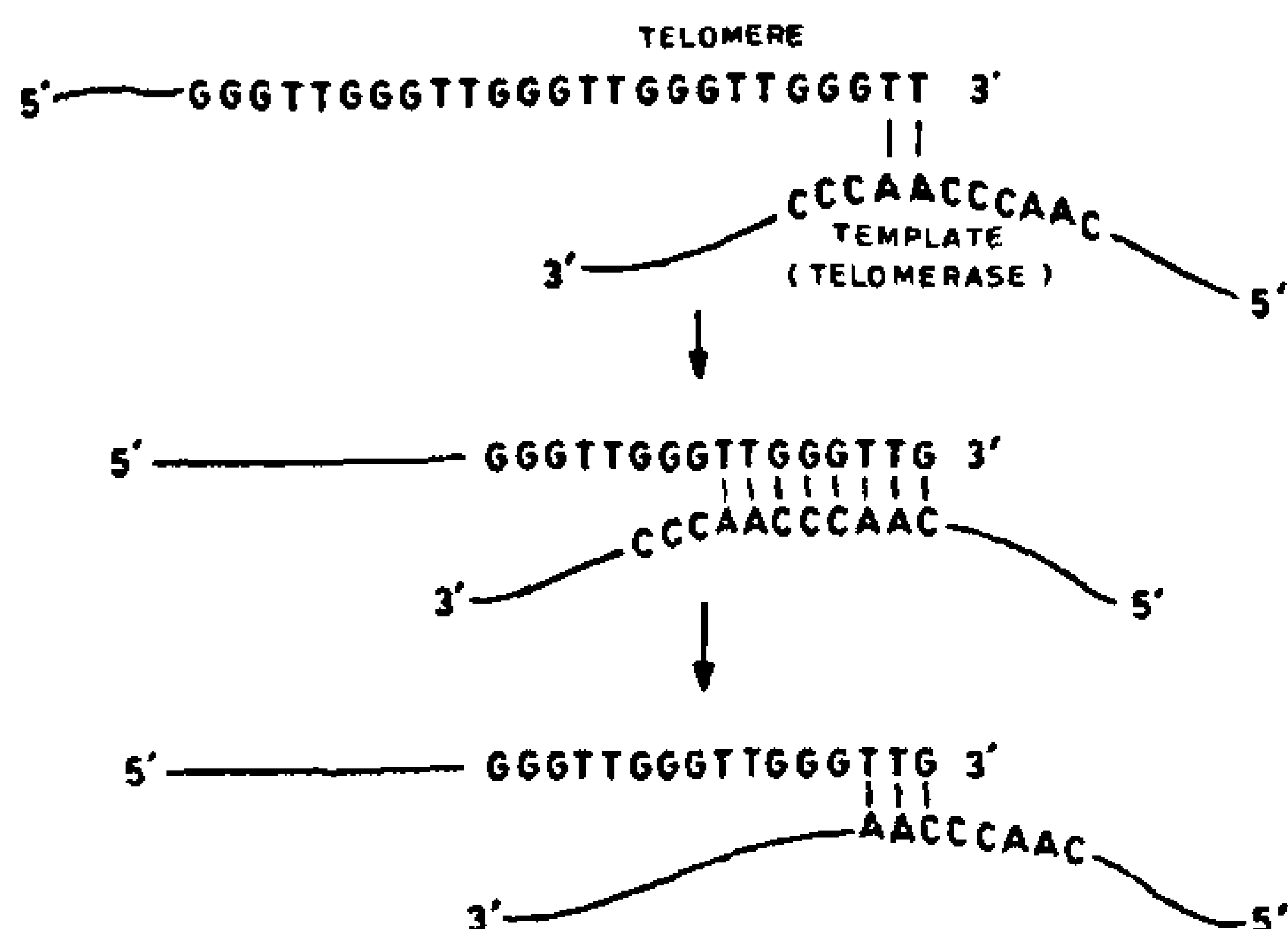

Figure 5. Mode of action of telomerase in repairing of telomeres and the role of built-in primer RNA.

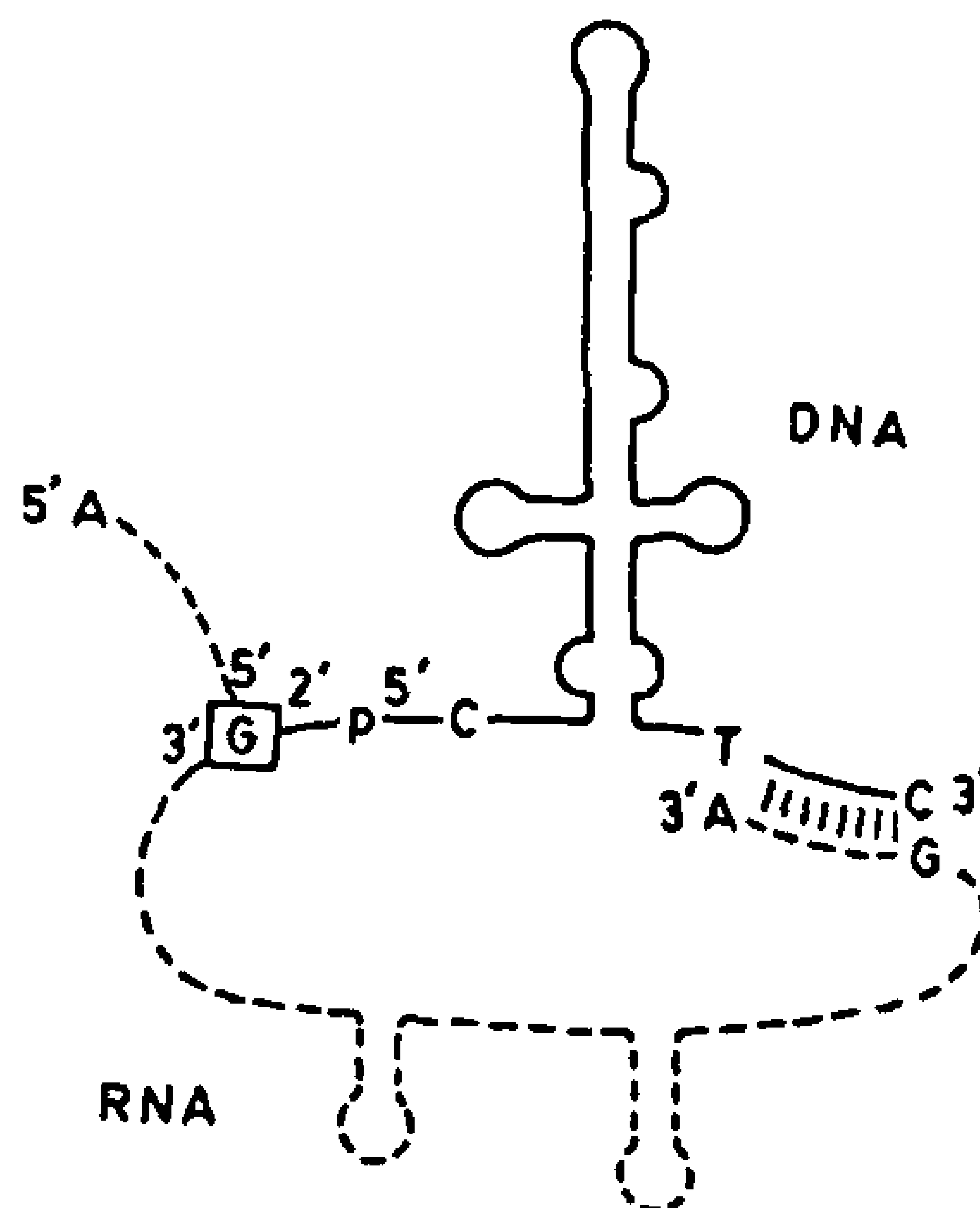

Figure 6. Structure of multiple-copy single stranded (ms) DNA (redrawn from cover of *Cell* **48**, No. 1).

is not built in its own structure. Telomerase can make the repeats by translocating the template. The C-rich strand can be synthesized in the normal way by DNA polymerase but the G-rich strand is left with an overhang which is subsequently protected by a protein. Although this mechanism is primarily to protect the ends of the chromosomes which have no message, the major object is not to expose the internal message to external damaging agents.

The use of RNA primer in normal DNA synthesis has already been mentioned. This RNA primer synthesized as complementary to the DNA strand is removed after DNA synthesis proceeds beyond. That way gaps are left at the ends. Telomerase fills up the gaps without a primer but requires a template to extend one strand only. Is this mechanism indicative of the process through which RNA had passed on the genetic message to DNA in case RNA acted as the original genetic elements and eventually passed on the messages to DNA by synthesizing it from its own ingredients. The most striking example of such an RNA fossil is the msDNA (multiple copy single-stranded DNA) known to occur in some microorganisms[26,27]. This single-stranded DNA is a hybrid fusion between a DNA strand and an RNA strand (Figure 6). At one end the two are linked through a 2′–5′ phosphodiester bridge, at the other end the two are joined through hydrogen bonds between complementary base sequences, giving rise to a circular hybrid structure with a protruding tail of an RNA strand. Other than its hybrid structure its synthesis is of great interest. The two components, DNA and RNA, are synthesized from two opposite strands (partially overlapping) of DNA which are transcribed most probably by a retro element (like reverse transposonase). The 2′–5′ linkage between the two is a mystery although there are abundant examples of such linkages in nature (for example, lariat formation in processing of mRNA) and enzymes responsible for synthesis of such unusual linkages are not uncommon. One wonders whether such hybrid molecules are the fossils of the period during which DNA evolved to act as a depository of genetic message contained in RNA, the instability of RNA being the compelling reason.

RNA-editing with the help of RNA is another example of diverse functions of RNA[28,29]. Three different laboratories while studying the mitochondrial genes in the parasite *Trypanosoma brucei* and *Leishmania tarentolae* ran into a puzzle. The coding in the gene did not correspond strictly to the sequence of the protein which it synthesizes. More puzzling was that the protein was synthesized in spite of the stop signals present within the reading frame. The obvious approach was to isolate the mRNA which actually directs the synthesis of protein and its sequence was found to be in variance with the genetic code. It appeared that the genes are being corrected, not at the genetic level but at the transcription level. A closer look revealed the fact that the addition or deletion of one or more U residues was responsible for this puzzling situation. The RNA-editing was the obvious term coined for this phenomenon. The actual mechanism of editing is, however, not so far known but the discovery of a species of RNA referred to as 'guide' RNA has provided clues to the understanding of the process. The guide RNA is rich in U and has some bases complementary to the region of mRNA which is to be edited. It is believed that guide RNA associates with mRNA at this site and its poly U tail is involved in editing (Figure 7). Just like in splicing,

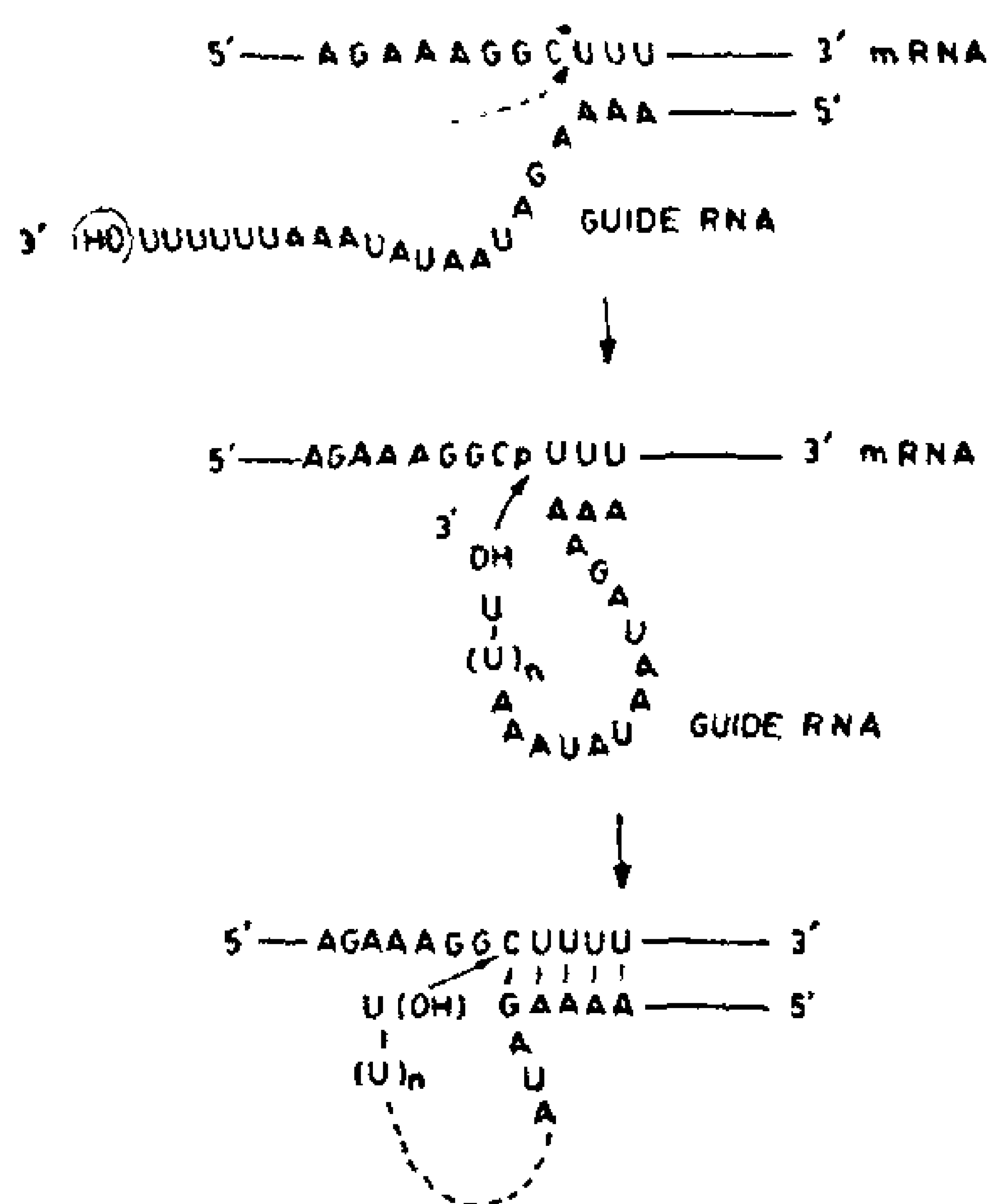

Figure 7. Editing of messenger RNA with the help of guide RNA.

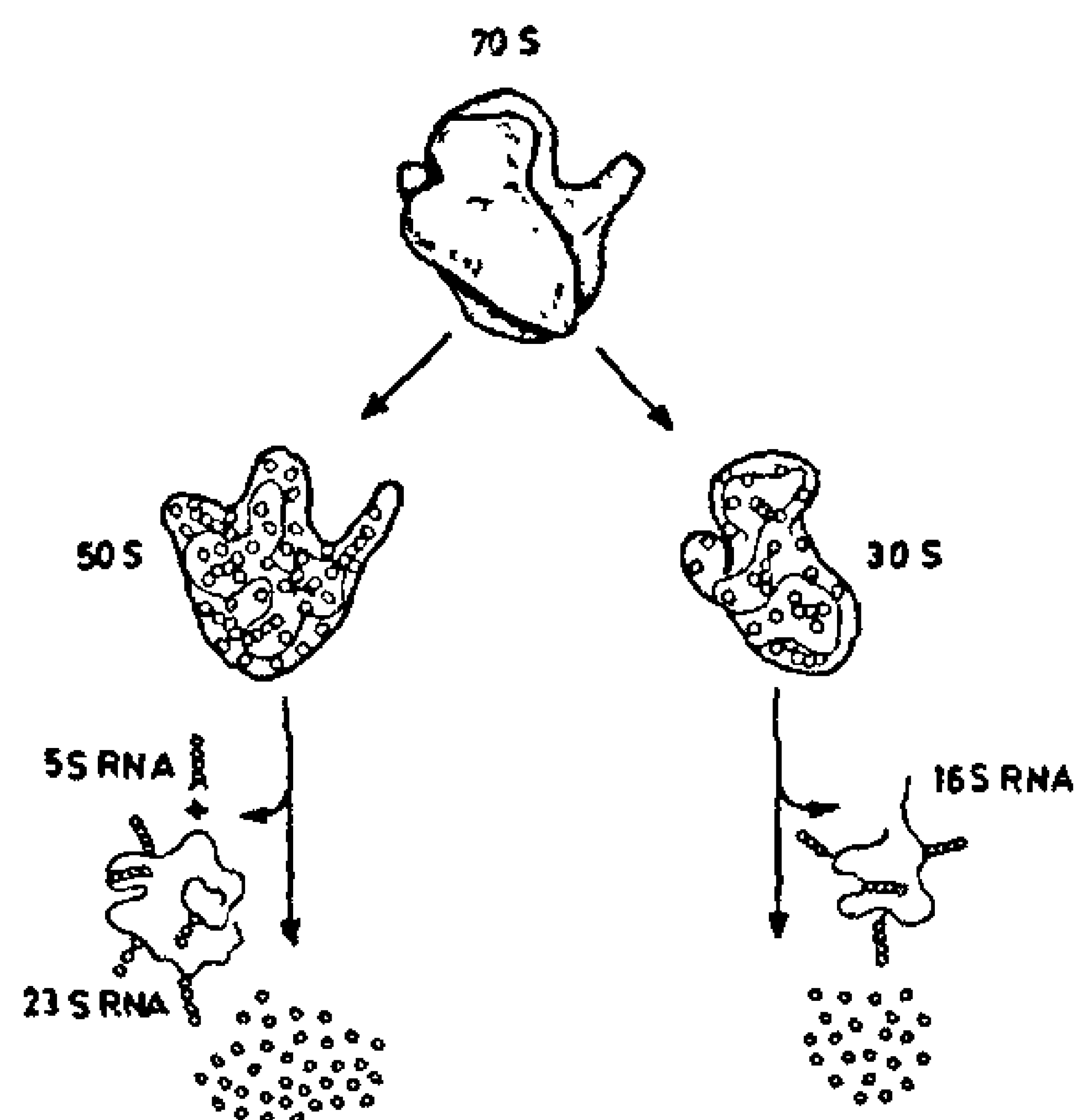

Figure 8. Structure of *E. coli* 70S ribosome and its subunits, 50S and 30S ribosomes.

nucleophilic attack of 3' terminal U at the site of insertion leads to the splitting of RNA. After insertion of U nucleophilic attack from the resulting 3' end leads to transesterification and joining of the split strand with U inserted at the site. Several such U's can be inserted by repetition of the process. The deletion of U may take place by reversal of the process. Although this mechanism is not fully established, the transesterification is the key to the editing process[30]. Such reactions might not be the attributes of only catalytic RNA but also of replicating RNA and thus may throw light on the early process of evolution where RNA acted not only as a genetic message but also as an enzyme and the editing process may be looked upon as one of the relics of the RNA world.

Since the recognition of ribosomes as the translational apparatus extensive work has been carried out to understand the structure and function of ribosomes[31-33]. *E. coli* ribosome (sedimentation constant of 70S) has two subunits like ribosomes from any other source (Figure 8). These subunits in turn have sedimentation constants, 50S and 30S. The smaller subunit has one RNA (16S) having approximately 1500 bases and the larger subunit contains two RNAs, one large (23S) with approximately 3000 bases and another small (5S) with 120 bases. There are 21 proteins present in the small subunit which are numbered S1–S21 and 35 in large subunit numbered as L1–L35. Two proteins (L7 and L12) are identical, L7 being *N*-acetylated L12. Further, L8 is actually a combination of three proteins (L10, L7 and L12). All the proteins occur in one copy per ribosome except L7/L12, four copies of which are present in 50S ribosomes. One protein of 30S subunit (S20) and one protein of large subunit (L26) are identical. Each of these proteins has been isolated in pure form and their sequences are known. Many of them are highly basic proteins and their molecular weights range from 10,000 to 60,000. Some of them have been crystallized and their three-dimensional structures have been worked out. As already mentioned, the sequences of all the three RNAs (5S, 16S and 23S) have been completely worked out; their secondary structures based on base-complementarity have been built up and three-dimensional structures are in the offing. As a whole, all the components of the *E. coli* ribosomes have been structurally elucidated. Further, the three-dimensional organization of the ribosome as a whole has been fairly well-established by electron microscopy, neutron diffraction as well as X-ray diffraction to some extent. This has been possible because the translational machinery built up with both RNAs and proteins has drawn the attention of a large number of physicists, chemists and biologists.

Ribosomes interact with 100 or odd factors during the various steps of protein synthesis, initiation elongation and termination. Naturally it is of great interest to understand the interaction of the components of the ribosomes with these factors and their involvement in protein synthesis. As expected, ribosomal proteins were implicated in all of these functions and the involvement of individual proteins in each of the functions was thought to be worked out. One of the best examples is the action of antibiotics[34]. The sensitivity of the microorganisms to a large number of the antibiotics was traced to ribosomal proteins and in many cases to a single one. For example, sensitivity to

streptomycin was linked to ribosomal protein S12 (ref. 35). The direct proof is the mutation in the protein which leads to the resistance to the antibiotic. Besides the above, the binding of tRNAs to the ribosomes at the decoding site, the peptide bond formation, the release of the finished protein chain, etc. were rightly thought to be mediated by ribosomal proteins through the help of external factors like initiation factors IF-1, IF-2 and IF-3, elongation factors EF-Tu, EF-Ts, EF-G and release factors RF-1, RF-2, and RF-3. Every step in protein synthesis fitted in nicely with the functional properties of proteins in general which became well-established since their identification about a century ago. However, the situation started to look different from a number of observations made comparatively recently.

Some of the external factors were found to interact directly with ribosomal RNAs rather than with proteins, for example, EF-G was shown to bind directly to 23S RNA rather than to any protein[36]. Similarly it was observed that some of the antibiotics bind to ribosomal RNAs than to proteins (review by Zimmermann *et al.*[37]). Actually their direct binding to proteins was searched for without any success. Streptomycin is the first antibiotic which was shown to bind directly to 16S RNA without mediation of any protein. This led to rethinking of ribosomal proteins as functional entities and vigorous attempts were made to find out whether ribosomal RNAs played the functional role in place of proteins. To cite a specific example, the substitution of the base within the binding site of 16S RNA leads to the development of resistance of the organism to streptomycin[38]. Similarly, base change in RNA resulted in streptomycin dependence (*ram* mutation) of the organism. Similar observations were made with other antibiotics as well, for example, thiostrepton[39].

There cannot be better examples of direct involvement of ribosomal RNAs in functions of ribosomes. However, the dilemma arose, whether the earlier observations regarding the involvement of proteins and recent demonstration of involvement of ribosomal RNAs could be reconciled with. This has been solved by demonstrating that mutations in ribosomal proteins lead to conformational changes of the regions of ribosomal RNAs to which the proteins bind[40]. This has not only solved the basic problem but also assigned an important role to ribosomal proteins, tuning the structure of ribosomal RNAs, so far as their functions are concerned. This has considerable evolutionary significance which will be discussed later.

This laboratory has been engaged in studies of the structure and function of *E. coli* ribosomes for quite some time. Some important observations made in this connection led to the new mode of thinking about the direct involvement of ribosomal RNAs in protein synthesis. After the initiation complex is formed by positioning formylmethionine tRNA at 30S ribosomal site with the help of the initiation factors IF-1, IF-2, and IF-3 50S subunit joins with the 30S subunit to form the 70S ribosome following the dissociation of the initiation factors. The two remain associated for as long as the elongation of protein chain continues with the help of elongation factors EF-T and EF-G. On the completion of protein chain the two subunits dissociate to initiate a second round of protein synthesis. The factor IF-3 actually regulates the association and dissociation. In the test tube such association and dissociation can be controlled by the concentration of Mg^{++}. At high Mg^{++} concentration (5–10 mM) the two associate whereas at lower Mg^{++} concentration (say 1 mM) they dissociate. It was shown in this laboratory[41] that naked 16S and 23S RNAs isolated from 30S and 50S ribosomes and free from detectable amount of proteins are capable of associating with each other in presence of high salt and Mg^{++} concentrations (or in the presence of alcohol). These two conditions were known to fold the structures of ribosomal RNAs to produce ribosome-like particles. Further, the former condition is required for the assembly of ribosomes from naked RNAs and proteins. The detailed studies had shown that the association is stoichiometric. Even under ordinary conditions the two RNAs are found to associate with each other, although partially.

This complex (16S.23S) was found to have limited ribosome-like activity[42]. It was demonstrated that it can synthesize very small amounts of polyphenylalanine with poly U as messenger RNA with the addition of requisite factors, L10 and L7/L12 proteins (ribosomal components). The synthetic capacity was about 1% of intact ribosomes. In a very recent observation Noller and his coworkers[43] have shown that *Tetrahymena* ribosomes retain their peptidyl transferase activity even when 80–90% of the proteins are removed. These results indicate that ribosomal proteins may not be absolutely essential for the biological activity of the ribosomes. Another interesting observation made by Cech, Noller and others[44] with modified ribozyme which can hydrolyse an amino acyl ester beyond the making and breaking of phosphodiester bonds supports the above results.

The elongation of the protein chain involves translocation at the ribosomal site of transfer RNAs which act as carriers of amino acids. In the classical model two sites (peptidyl and aminoacyl) are indicated (Figure 9). The transfer RNA carrying the amino acid to be added to the growing chain is positioned on the ribosome at the A site while the growing chain associated with the transfer RNA is positioned at peptidyl site (P site). With the transfer of the growing peptidyl chain to the incoming amino acid attached to tRNA at the A site the P site becomes free and leaves the ribosome. The elongated (by only one amino acid)

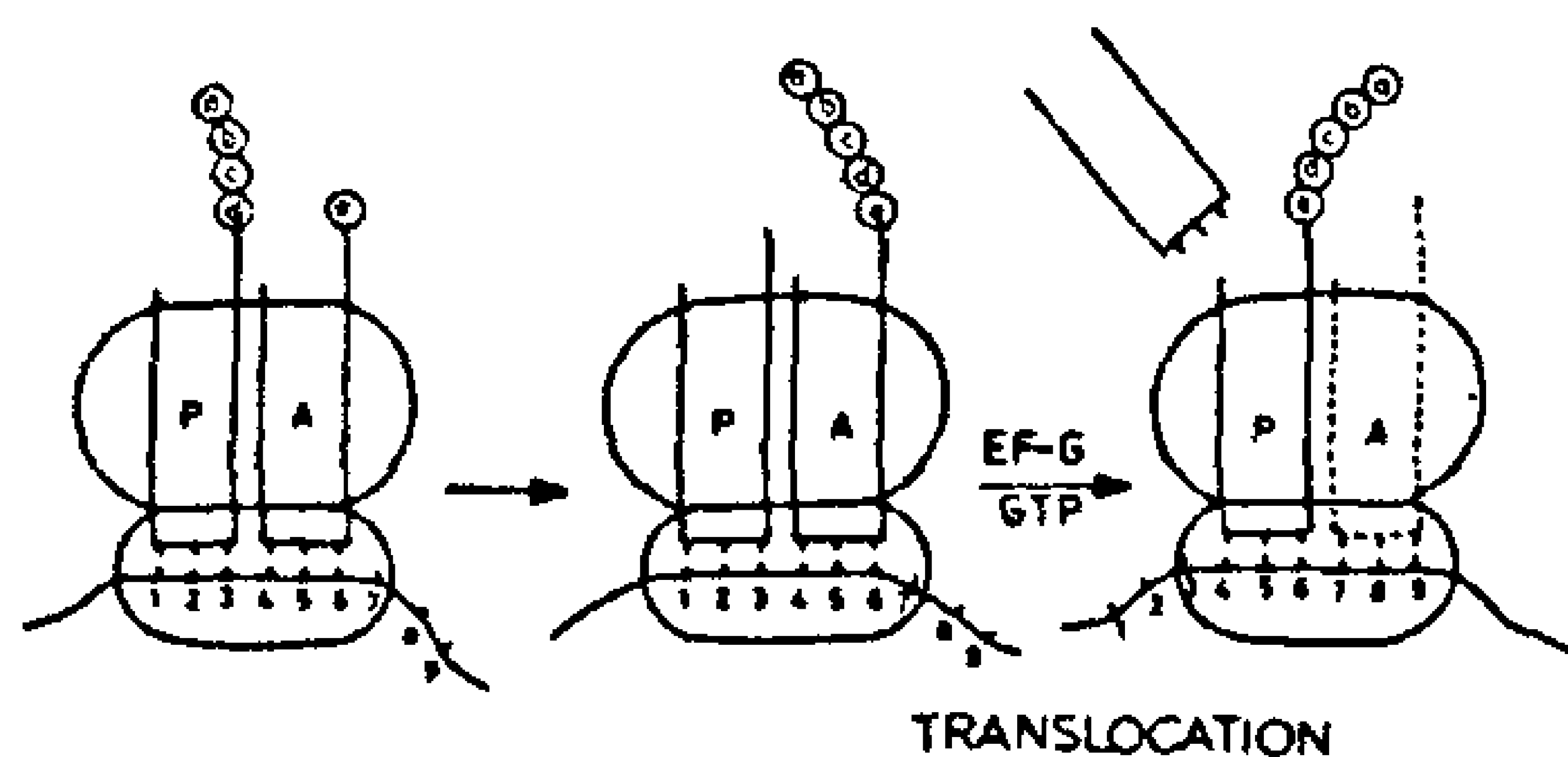

Figure 9. Translocation of peptidyl tRNA from A (aminoacyl) site to P (peptidyl) site during protein synthesis.

peptide chain associated with the transfer RNA at the A site subsequently moves to the P site, making space for the next amino acid to enter along with its transfer RNA at the vacant A site. The cycle is repeated again and again as the chain is elongated. The translocation of peptidyl tRNA from A site to P site (along with messenger RNA) which dictates the exact sequence of the amino acids to be added is a complicated process. The L7/L12 stalk region (composed of 4 L7/L12 molecules, L7 being *N*-acetylated L12) of 50S ribosomes is thought to be involved in this process. The site at which the peptide bond is formed is referred to as 'peptidyl transferase' centre. This is believed to be created at the interface of the two subunits on joining. The elongation factors EF-T and EF-G which are primarily involved in the process were shown to directly interact with 23S RNA at the base of the L7/L12 stalk region. Several evidences from this laboratory indicated that 23S RNA of 50S ribosomes is directly involved in translocation. Actually a model[45,45a,46] has been proposed from this laboratory on basis of these (Figure 10). According to this model 50S ribosome undergoes conformational change during the elongation cycle. The L7/L12 stalk

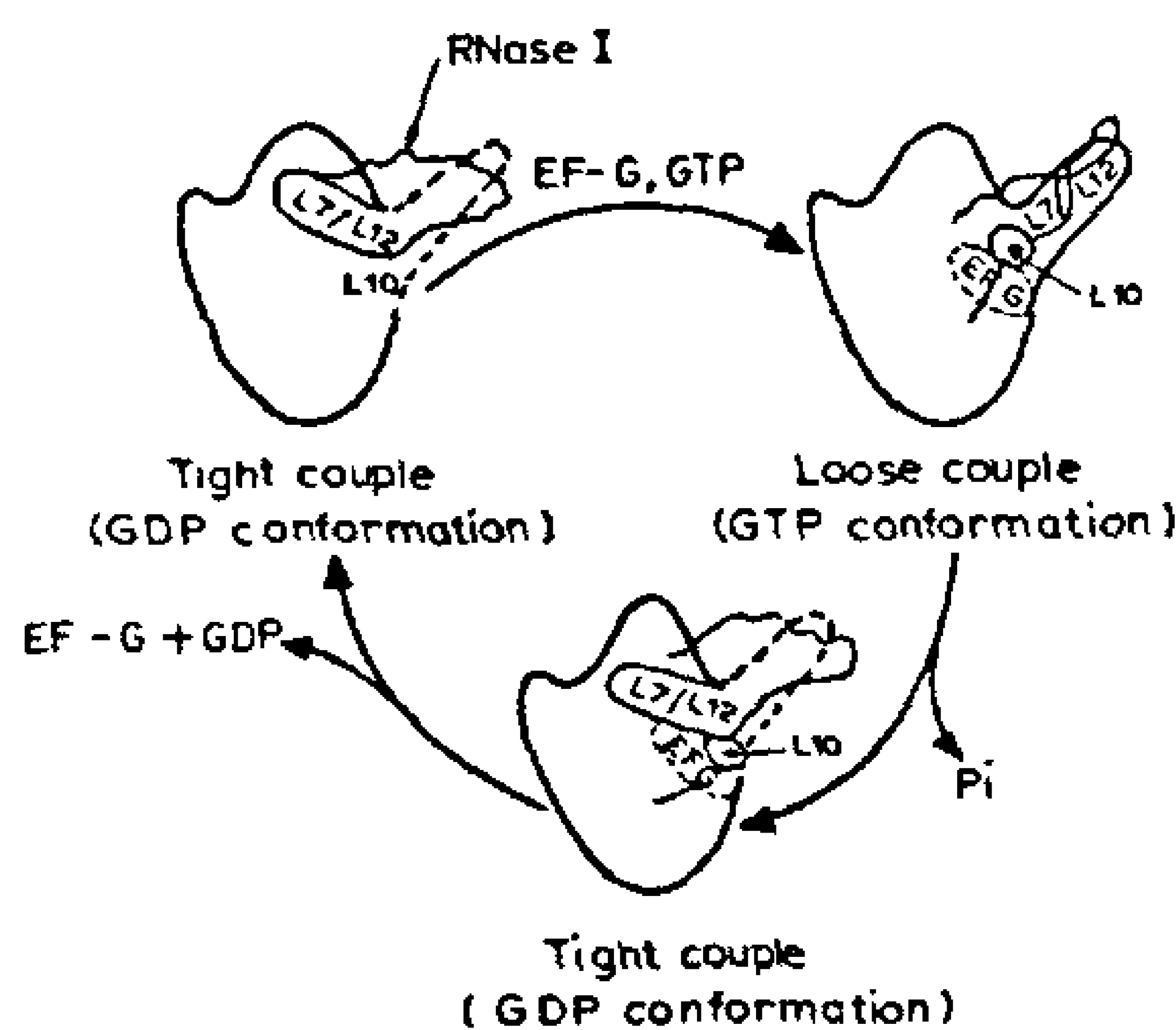

Figure 10. Model of translocation proposed from this laboratory[45,45a,46].

region is known to be mobile and its mobility is responsible for its having two extreme positions, one in the extended form towards the medium and the other in the folded form (more towards the body). The positions are thought to be guided by EF-G along with GTP interacting with 23S RNA at the base of the L7/L12 stalk region. It has been shown in this laboratory that ribosomal RNA in this region can undergo conformational change as its sensitivity to RNase varies under different conditions. Two populations of 70S ribosomes have been known for quite some time. These are designated as tight and loose couples. Tight couple subunits associate at lower Mg^{++} concentration (4–5 mM) whereas loose couples need higher Mg^{++} concentration (10 mM or more) for association. According to our model loose couple with the extended stalk region (and having GTP conformation) is produced from tight couple on reaction with EF-G-GTP. On hydrolysis of GTP, GDP conformation is produced in which the stalk region is folded towards the body. The inter-conversion of tight and loose couples leads to the conformational change of RNA not only at the base of the stalk region but also at the peptidyl transferase centre and this is responsible for the translocation phenomenon.

Some very recent observations made in this laboratory strongly support the above-mentioned model. Antibody raised against the stalk composed of L7/L12 protein reacts to different extents to the TC and LC ribosomes; the LC ribosomes react more than the TC[47]. This is a clear indication of greater accessibility of the stalk in LC. Moreover, according to this model, TC and LC ribosomes are supposed to utilize different regions of RNA for the association of the two subunits during elongation of protein chain. This has been unequivocally demonstrated with the use of synthetic oligodeoxyribonucleotide. A number of oligomers complementary to the regions of RNA suspected to be involved in the association have been synthesized and their effects on the association of the subunits have been studied[48] (also unpublished results from this laboratory). Those which block the association indicate the involvement of the corresponding region. These sites can be classified into three categories—one or two sites are found to be commonly used by both TC and LC ribosomes, for example, 787–795 of 16S RNA and 2304–2314 region of 23S RNA. Some sites are used by only TC for example, 2750–2757 of 23S RNA and some sites by LC only (1491–1500 of 16S RNA). These results indicate the spatial movement of the two subunits with respect to each other which is the basis of translocation. It was shown earlier in this laboratory that the portion of RNA present at the base of the stalk in TC (Figure 10) is sensitive to the action of an enzyme RNase I, whereas LC ribosomes as well as core 50S ribosomes (which are deficient in L7/L12) are resistant to this enzyme. The

change in conformation of RNA in this region was thus strongly supported. Thiostrepton, an antibiotic which has been known to inhibit translocation in TC ribosomes, fails to do that in LC ribosomes (unpublished results from this laboratory). It is well-established that the antibiotic binds directly to RNA in the same region (base of the stalk) where EF–G binds[36]. Taking all these facts into account it may be surmised that the conformation of RNA in this region is dependent on the binding of EF–G and GTP. The conformation of RNA at this stage is referred to as GTP conformation (Figure 10). Subsequently hydrolysis of GTP to GDP takes place and EF–G–GDP complex changes the conformation of RNA to GDP conformation. EF–G–GDP complex due to its low affinity to altered RNA conformation dissociates from the site, leaving the ribosome in GDP conformation. Thus TC and LC 50S ribosomes represent GTP and GDP conformations respectively. It is further assumed that RNA in this region acts as a switch and the change in conformation in this region does not remain localized but is propagated to the peptidyl transferase centre to effect translocation. If this model is eventually found to be correct not only will the functional importance of RNA in ribosomes be established but the textbook information that proteins are the functional elements in ribosomes has to be modified. The role of ribosomal proteins may be reduced to tuning the conformations of RNAs in different regions and thus to be indirectly involved in protein synthesis. Ribosomes may be the modern version of ribosomal RNAs responsible for protein synthesis in the early period.

The above raises an interesting issue on the evolution of ribosomes. Structures of ribosomal RNAs from a large number of sources, both prokaryotic and eukaryotic organisms, have been worked out and a remarkable observation has been made. Most of the regions of RNAs which are found to be involved in some kind of function are highly conserved. The variation is observed in those regions which do not seem to be involved in any kind of function. This is strongly in favour of the fact that ribosomal RNAs were the primitive ribosomes. The counterargument that the protein-binding sites are the conserved ones is not quite convincing as the factor-binding sites (for initiation, elongation and termination factors) which are highly conserved do not belong to proteins but to rRNAs. In this connection Noller[49] proposed an interesting model correlating splicing reaction of RNAs and mode of protein synthesis by ribosomes both of which are found to be sensitive to the same antibiotics[50]. The splicing process in which distant regions of RNAs are brought together to loop out the intron region which is eventually removed might have led to the evolution of the translational machinery. The latter is strongly analogous to two tRNAs bound to

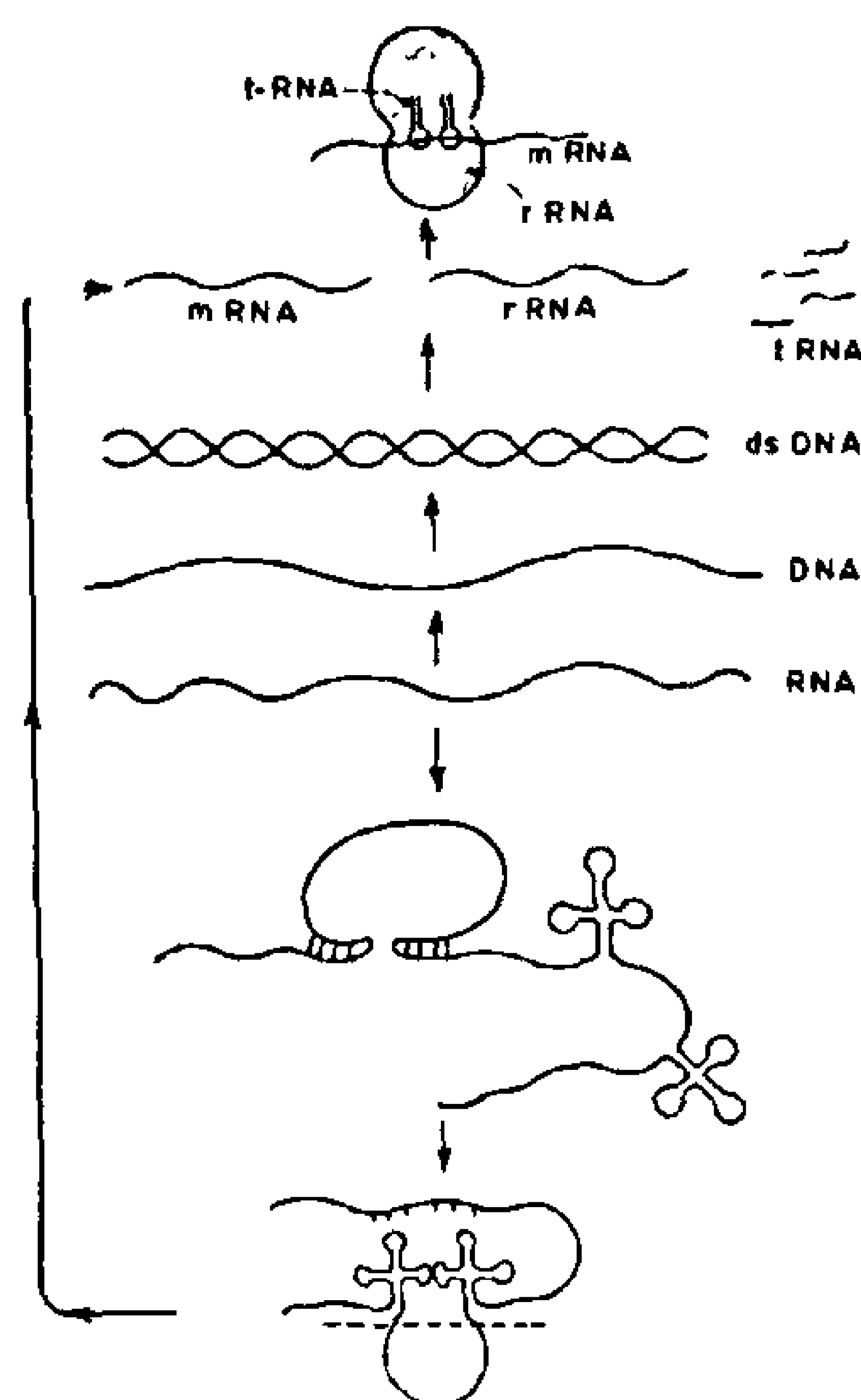

Figure 11. Evolution of ribosome (hypothetical model), including the concept of Noller[49], based on the action of antibiotics on splicing of ribosomal RNA[50].

adjacent codons on a messenger RNA. Transfer RNAs are thus evolutionary products of ribosomal RNAs. It is well known that ribosomal RNAs and some transfer RNAs are synthesized together from DNA and remain linked to each other till processing takes place. A consolidated picture is presented in Figure 11. Further insight into this aspect may throw light on the evolution of the translational apparatus and its present day capability as deployed in association with ribosomal proteins. That will lead to final justification of the RNA world, a concept which is gradually taking shape.

1. Mirsky, A. E., *Sci. Am.*, 1868, June, 78.
2. Watson, J. D. and Crick, F. H. C., *Nature*, 1953, **171**, 737.
3. Cech, T. R. and Bass, B. L., *Annu. Rev. Biochem.*, 1986, **55**, 599.
4. Joyce, G. F., *Nature*, 1989, **338**, 217.
5. Kornberg, A., *DNA Replication*, 1980; Supplement to *DNA Replication*, 1982, W. H. Freeman, San Francisco, USA.
6. Zamecnick, P. C., in *Symposium on Quantitative Biology*, Cold Spring Harbor Laboratory, USA, 1969, vol. 34, p. 1.
7. Holley, R. W., *Sci. Am.*, 1966, February, 30.
8. Suddath, F. L., Quigley, G. J., McPherson, A., Sneden, D., Kim, J. J., Kim, S. H. and Rich, A., *Science*, 1974, **248**, 20.
9. Brosius, J., Palmer, M. L., Kennedy, P. J. and Noller, H. F., *Proc. Natl. Acad. Sci. USA*, 1978, **75**, 4801.
10. Brosius, J., Dull, T. J. and Noller, H. F., *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 201.
11. Brimacombe, R., Atmadja, J., Stiege, W. and Schuler, D., *J. Mol. Biol.*, 1988, **199**, 115.
12. Doring, T., Greuer, B. and Brimacombe, R., *Nucleic Acids Res.*, 1991, **19**, 3517.

13. Jacob, F. and Monod, J., *J. Mol. Biol.*, 1961, **3**, 318.
14. Altman, S., in *Advances in Enzymology* (ed. Meister, A.), John Wiley and Sons Inc., 1989, p. 1.
15. Cech, T. R., *Sci. Am.*, 1986, November, 76.
16. Berget, S. M., Moore, C. and Sharp, P. A., *Proc. Natl. Acad. Sci. USA*, 1977, **74**, 3171.
17. Chow, L. T., Gelinas, R. E., Broker, T. R. and Roberts, R. J., *Cell*, 1977, **12**, 1.
18. Brody, E. and Abelson, J., *Science*, 1985, **228**, 963.
19. Maniatis, T. and Reed, R., *Nature*, 1987, **325**, 673.
20. Gall, J. G., *Science*, 1991, **252**, 1499.
21. Cech, T. R., *Cell*, 1986, **44**, 207.
22. Walter, P. and Blobel, G., *Nature*, 1982, **299**, 691.
23. Walter, P. and Blobel, G., *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 7112.
23a. Davis, T. L., Firulli, A. B. and Kinniburgh, A. J., *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 9682.
24. Blackburn, E. H., *Nature*, 1991, **350**, 569.
25. Blackburn, E. H., *Trends Biochem. Sci.*, 1991, **16**, 379.
26. Furuichi, T., Dhundale, A., Inouye, M. and Inouye, S., *Cell*, 1987, **48**, 47.
27. Inouye, M. and Inouye, S., *Trends Biochem. Sci.*, 1991, **16**, 18.
28. Banne, R., *Biochim. Biophys. Acta*, 1989, **1007**, 131.
29. Simpson, L., *Science*, 1990, **250**, 512.
30. Blum, B., Sturn, N. R., Simpson, A. M. and Simpson, L., *Cell*, 1991, **65**, 543.
31. Hansen, H. A. S., Volkman, N., Piefke, J., Glotz, C., Weinstein, S., Makowski, S., Meyer, S., Wittmann, H. G. and Yonath, A., *Biochim. Biophys. Acta*, 1990, **1050**, 1.
32. Wittmann, H. G., in *Structure, Function and Genetics of Ribosomes* (eds. Hardesty, B. and Kramer, G.), Springer Verlag, New York, 1985, p. 1.
33. Burma, D. P., *J. Sci. Ind. Res.*, 1984, **43**, 388.
34. Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmand, M. H. and Waring, J. J., *The Molecular Basis of Antibiotic Action*, J. P. Wiley and Sons, New York, 1981.
35. Davis, B., *Microbiol. Rev.*, 1987, **51**, 431.
36. Girshovich, A. S., Bochkareva, E. S. and Gudkov, A. T., *FEBS Lett.*, 1982, **150**, 99.
37. Zimmermann, R. A., Thomas, C. L. and Wower, J., in *The Ribosomes, Structure, Function and Evolution* (eds. Hill, W. E., Dahlberg, A. E., Garrett, R. A., Moore, P. B., Schlessinger, D., Warner, J. R.), American Society for Microbiology, Washington DC, 1990, p. 331.
38. Montandon, P. E., Wagner, R. and Stutz, E., *EMBO J.*, 1986, **5**, 3705.
39. Ryan, P. P. and Draper, D. E., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 6308.
40. Allen, P. M. and Noller, H. F., *J. Mol. Biol.*, 1989, **208**, 457.
41. Burma, D. P., Nag, B. and Tewari, D. S., *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 4875.
42. Burma, D. P., Tewari, D. S. and Srivastava, A. K., *Arch. Biochem. Biophys.*, 1985, **239**, 427.
43. Noller, H. F., Vernita, H. and Zimniak, L., *Science*, 1992, **256**, 1416.
44. Piccirilli, J. A., McConnell, T. S., Zaug, A. J., Noller, H. F. and Cech, T. R., *Science*, 1992, **256**, 1420.
45. Burma, D. P., *J. Biosci.*, 1984, **6**, 419.
45a. Burma, D. P., Srivastava, A. K., Srivastava, S., Dash, D., Tewari, D. S. and Nag, B., *J. Biosci.*, 1985, **8**, 757.
46. Burma, D. P., Srivastava, S., Srivastava, A. K., Mahanti, S. and Dash, D., in *Ribosomes, Structure, Function and Genetics* (ed. Hardesty, B. and Kramer, G.), Springer Verlag, New York, 1986, p. 438.
47. Dey, D. N. and Burma, D. P., *Indian J. Biochem. Biophys.*, 1991, **28**, 369.
48. Agrawal, R. K., De, A. and Burma, D. P., *Indian J. Biochem. Biophys.*, 1992, **29**, 148.
49. Noller, H. F., *Nature*, 1991, **353**, 302.
50. Von Ahsert, U., Davies, J. and Schroeder, R., *Nature*, 1991, **353**, 386.

RESEARCH COMMUNICATIONS

Global search for optimal biomolecular structures using mutually orthogonal Latin squares—A novel algorithm

N. Gautham and Z. A. Rafi

Department of Crystallography and Biophysics, University of Madras, Guindy Campus, Madras 600 025, India

We propose a modified grid search technique to find global starting points for the optimization of biomolecular conformations, in which the multidimensional search space is reduced to a two-dimensional one using mutually orthogonal Latin squares (MOLS). If there are m dimensions in the conformational space, each of size n, the method involves the computation of the energy values at n^2 of the n^m possible points, these n^2 points being chosen using MOLS. Subsequent analysis of the n^2 energy values using simple statistics allows identification of the optimal conformation. The computation time compared to conventional grid search methods could be reduced by several orders of magnitude. We have succesfully applied the method to arrive at the gross optimal conformation for a mononucleotide, a dinucleoside monophosphate, and for a tetrapeptide.

THEORETICAL studies of biomolecular conformation often use optimization techniques to minimize the energy and arrive at a stable geometry. Some of the most frequently used techniques[1-3] are essentially local minimization methods and require an appropriate initial conformation. In the absence of previous structural knowledge about the system, these initial points in the multidimensional configuration space may be arrived at by a sampling of the entire space. If the sampled points are closely spaced, such a search may become prohibitive in terms of computation time. We describe here a method to drastically decrease the computation time required to identify possible starting conformations. The method uses mutually orthogonal Latin squares (MOLS).