Are synthesis and folding of proteins overlapping functions of the ribosomal RNA?

Chanchal Das Gupta

Department of Biophysics and Molecular Biology, University of Calcutta, College of Science, 92 Acharya Prafulla Chandra Roy Road, Calcutta 700 009, India

A general protein folding activity has been detected in the large subunit of the ribosome. This activity seems to be located in the peptidyl transferase domain of the large RNA of this subunit. In contrast to the protein folding activity of the molecular chaperones, this activity is (a) present in the RNA and is (b) universal, not selective for any protein. The overlap of this active site with the peptidyl transferase centre on the ribosomal RNA suggests a functional overlap between protein synthesis and folding by ribosome in the cell.

RIBOSOMES synthesize protein molecules and active protein molecules are released when we break the cells open. Even when messages are translated *in vitro* with ribosomal preparations, our experience tells that most of the polypeptides acquire active conformations[1–3]. This is true even in some cases where the ribosomal extract used for *in vitro* protein synthesis is freed from the known chaperone proteins which are likely to associate with the ribosomes[3]. But whereas all efforts were directed towards working out in commendable detail the mechanism of polypeptide synthesis on the ribosomes in prokaryotic and eukaryotic cells, not much effort was made to examine the ability of ribosomes to fold polypeptides when they were synthesized on them[3,4]. The reason behind this could be the overwhelming appeal of the Anfinsen paradigm[5] that the instructions for folding into the tertiary conformation of a protein are written in its amino acid sequence.

Versatile roles of chaperone proteins in protein folding

When genetic and biochemical data accumulated to describe circumstances in which a number of protein molecules failed to attain the active form within the cell in absence of assisting proteins, collectively called molecular chaperones[6–8], attention of scientists was drawn to the chaperones and data started pouring in on various activities of a large number of chaperone proteins from wide variety of sources. In short, we now have an estimate of what fraction of cellular proteins could depend on the available chaperones to maintain their structures and functions in normal and stressed conditions. But at the same time the chaperones are selective. Not a single chaperone protein has been shown to be able to fold each and every protein from its denatured state. The selectivity, on the other hand, enables the chaperones to play exclusive roles in cellular metabolism, e.g. in protein trafficking, cellular stress alleviation, phage morphogenesis, formation of replication and transcription initiation complexes, to name a few[9].

Ribosomes as the general protein folding modulators

Over the last several years, reports from this and a few other laboratories appeared to have gone towards establishing a general role of ribosomes in protein folding. In what follows, some facts about the role of ribosome in protein folding and the lessons we learn about the versatile role of RNA (ribosomal RNA in this discussion) from such studies are discussed.

It was shown from this laboratory that ribosomes from both prokaryotic and eukaryotic sources could refold a large number of proteins from their denatured states (the nearest to the nascent synthesized polypeptide conformation that we can achieve in the test tube) to active form. The ribosomes were taken from *Escherichia coli*, methanogenic bacterium *Methanosarcina barkeri*, wheat germ and rat liver and the proteins were also chosen randomly. The only criterion for choice of enzyme was that it should be amenable to quick and quantitative assay, because that will give precisely the extent to which the protein could be folded by the ribosomes. A large number of proteins like bacterial alkaline phosphatase, glucose 6-phosphate dehydrogenase, glucose oxidase, lactate dehydrogenase, horse radish peroxidase, malate dehydrogenase, β lactamase, restriction endonucleases like *Eco*R1, *Bam*H1, *Hin*dIII, *Pst*I, β-galactosidase, carbonic anhydrase, etc. could be folded by the ribosomes. Ribosomes did not fail to fold any protein used in such experiments so far. Thus it appears

email: ckdg@cubmb.ernet.in.

that the ribosomes are general protein folding modulators crossing species barrier[10–14].

When the ribosomal particles were split into the larger and smaller subunits, the protein folding activity was found in the large subunit only[12,14–16]. This was not surprising, considering that the large subunit mainly constitutes the workshop for polypeptide synthesis – the formation of peptide bond and growth of the polypeptide chain. The small subunit basically is engaged in mRNA recognition.

In order to see which part(s) of the large ribosomal subunit are necessary for protein folding, the *E. coli* 50S ribosomal subunit was chosen and its proteins were stripped off with the usual high salt wash. Increasing concentrations of lithium chloride were used to gradually dissociate proteins from the surface of the 50S subunit. But after each salt wash, the protein folding activity stayed in the core particle which was resistant to the wash. Finally the activity could be traced in the 23S ribosomal RNA. This RNA was shown to carry out peptidyl transferase reaction when the proteins were largely or completely stripped off from the 50S subunit[12]. The 23S rRNA was also found to fold denatured proteins almost completely[12,14,15]. Since the peptidyl transferase activity, including amino acyl and peptidyl tRNA binding, etc. appeared to be associated with the domain V of 23S rRNA[17,18], it appears attractive to think that this domain could have some role to play in the protein folding reaction as well.

The protein synthesis inhibitory antibiotics which are known to bind to the large loop of domain V could also inhibit protein folding. Also, oligonucleotides complementary to the single-stranded segments in the central loop could inhibit protein folding when annealed to the 23S rRNA[14]. Similar results were reported from the laboratory of Boyd Hardesty who worked with some other proteins[16].

Role of ribosome in protein folding in the cell

The 23S rRNA-mediated protein folding is inhibited by antibiotics like chloramphenicol, lincomycin and erythromycin which are known to bind to the central loop of its domain V and stop protein synthesis[19]. But it is insensitive to the antibiotics like streptomycin, kasugamycin, etc. which bind to the 30S subunit and stop protein synthesis. When any of these antibiotics was added to the growing *E. coli* cells in which β galactosidase was induced, the enzyme synthesis stopped immediately; but the enzyme activity continued to rise for about 10 min before reaching a plateau in cells grown in presence of 30S specific antibiotics whereas it did not go up in cells growing in presence of 50S specific antibiotics. The increase in enzyme activity in cells growing in presence of 30S specific antibiotics was significant and this could only be due to the enzyme molecules which took some time to fold to active form after their synthesis was stopped with antibiotics. At any point during this rise in activity, addition of 50S specific antibiotics would arrest the process. In other words, the enzyme molecules acquired their active state post translationally and the large ribosomal subunit directed that process[20]. This activity of ribosome had not been investigated so far, although it is coupled with protein synthesis in as much as the peptdiyl transferase activity overlapped with it and both the processes are sensitive to the same set of antibiotics. It appears that the ribosomes not only synthesize the polypeptide chains, but also work on it to see that the chains fold to meaningful proteins to carry on cellular activities. Part of this process takes place on 50S subunit since it is inhibited by 50S specific antibiotics only up to several minutes after the stoppage of protein synthesis[20]. Movement of the newly-synthesized polypeptide on the 50S subunit has also been reported by Choi and Brimacombe[21] as shown in Figure 1.

Is domain V RNA a unique ribozyme that acts on polypeptide chains?

The fact that the domain V of 23S rRNA is responsible for protein folding necessitates special properties for this RNA segment. We are familiar with many activities of RNA of which RNA processing, RNA splicing, RNA editing, etc. are the most important. But these are ribozymes which act on RNA only in *cis* or in *trans*. A general responsibility of the domain V RNA also appears to be to fold the polypeptides being synthesized on the ribosome. This activity does not appear to be a simple prevention of aggregation of polypeptides since only one polypeptide chain remains associated at a time with the ribosome[21] for several minutes during which its folding remains sensitive to the action of domain V RNA specific antibiotics[20]. A more detailed study of the involvement of domain V RNA in protein folding has provided better understanding of this function.

Complementary role of two fragments of domain V RNA in protein folding

The entire domain V of the 23S rRNA from *B. subtilis* and its central loop were cloned in the laboratory of B. Weisblum[22] next to SP6 promoter to permit *in vitro* transcription of the entire domain V (600 nt long RNA) or of the central loop (337 nt RNA). The 600 nt long RNA could refold denatured proteins[23]. The cloned 660 nt long insert could be cut with restriction endonuclease Sma I and transcribed *in vitro* to get a 425 nt long runoff RNA. The 337 nt (RNA 1) and 425 nt (RNA 2) long

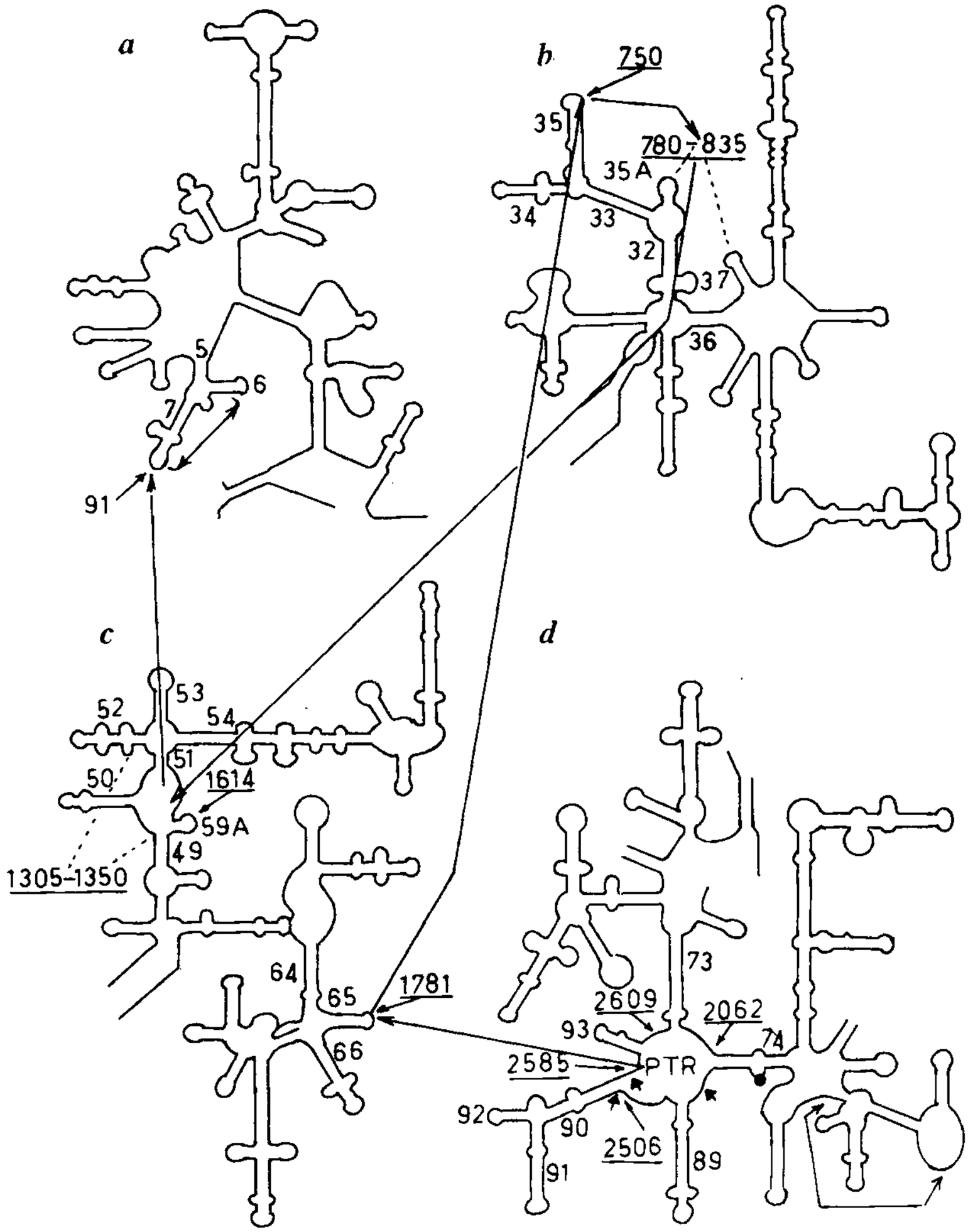

Figure 1. The path of the nascent peptide chain through the 23S rRNA. The complete secondary structure of the 23S rRNA is shown with helices in the neighbourhood of the cross link sites numbered[31]. The cross-linked nucleotides (or regions) are underlined and the overall path of the peptide is indicated by the heavy arrowed lines. PTR stands for peptidyl transferase ring[32] and the blunt arrows in this area are cross link sites for amino acyl tRNA[33]. The black circle under helix 74 is also a cross-link from amino acyl tRNA[34]. Photo reactivated cross-links of the peptide to 23S rRNA was used to identify the points of contact between the RNA and the peptide labelled at N-terminal with diazirine (reproduced with permission from ref. 21).

RNA molecules cover the entire length of 600 nt long RNA with some overlap (Figure 2). RNA 1 and RNA 2 complemented in the protein folding reaction which went through the following steps: (1) strong non-covalent binding of denatured enzyme to RNA 1; this RNA bound enzyme was not active; (2) release of the enzyme from RNA 1 by RNA 2 in a folding competent state and (3) slow attainment of active form by this folding competent protein free from both RNA 1 and RNA 2 (Pal *et al.*, communicated). The last step was rate limiting and took about ten minutes to complete. This timing agreed with the slow folding of β galactosidase present in the cell when further synthesis of the enzyme was blocked by the 50S subunit specific antibiotics[20]. Here also, there was slow increase in the activity of the already-synthesized enzyme which took about ten minutes to complete. It appears then that the ribosome releases the newly-synthesized enzyme in a folding competent state that completes the rest of the process post-translationally. Since this is true for all proteins and not selective like the chaperone-mediated folding, the domain V RNA could have very specific

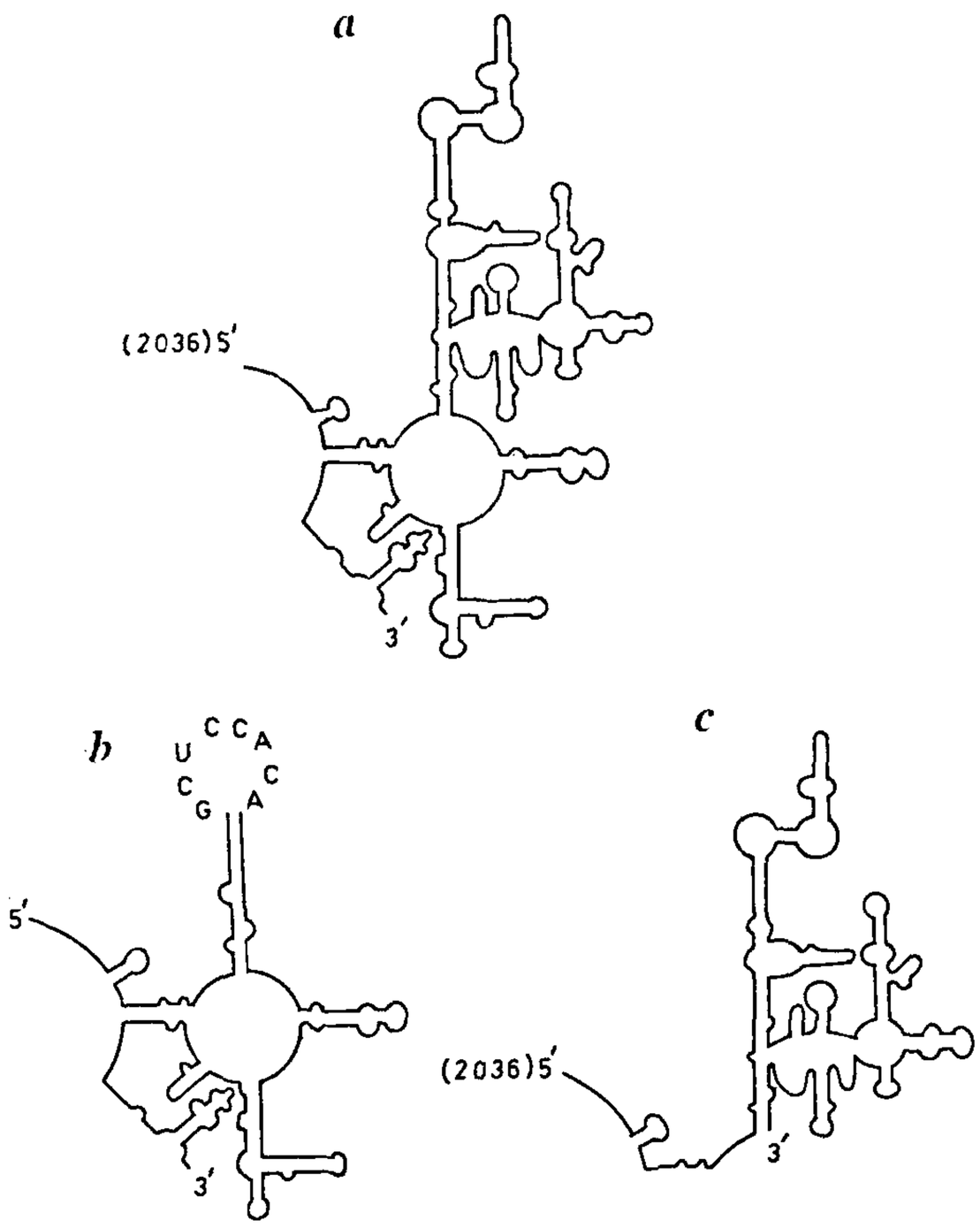

Figure 2. Secondary structures of different segments of domain V of *B. subtilis* 23S rRNA. *a*, 666 nt complete domain V; *b*, 337 nt; *c*, 425 nt long segments of domain V. The 337 nt long RNA was transcribed from the clone in which the nucleotides mentioned in upper stem were added to join the two arms of the stem. All the RNA molecules were transcribed by SP6/T7 RNA polymerase as run off transcripts from appropriate clones linearized by restriction endonucleases.

reaction with all the protein molecules in the cell. It gives a deterministic power to the domain V RNA which would only fold proteins which can engage themselves in this specific reaction with it. This RNA has very similar secondary structure all over the living world and the single-stranded region in its central loop (RNA 1) is most conserved throughout the process of evolution from the mitochondria, the chloroplast, the smallest prokaryotic cells to the higher eukaryotes[18]. Thus the proteins might have evolved under the strict rule of interaction set by the central loop of the domain V RNA. All natural proteins are then likely to have common motif in some intermediate state of folding at which they are recognized by the central loop of domain V. This recognition motif could be spread over in small segments on the linear polypeptides and may not be discernible by simply looking at the amino acid sequence or secondary/tertiary structure of the proteins as shown in the cartoon (Figure 3). Such a core recognition motif could then have incorporated peptide segments which could make local folds like small α helices without disturbing the residues interacting with the central loop. This would lead to increase in length of the protein with concomitant rise in its information content – the evolution of protein molecules. The central loop (RNA 1) is the only region necessary for protein folding. Instead of RNA 2, the denatured protein bound to RNA 1 could also be dissociated with low concentrations of detergent or ethanol which counteract hydrophobic interactions that hold the protein–RNA 1 complex together and the freed protein could fold equally efficiently (Pal *et al.*, communicated).

The central loop of domain V which has been conserved through evolution thus presents a 'molecular mould' where a specific three-dimensional motif of a protein-folding intermediate would fit properly to give rise to a productive-folding intermediate which can complete the remaining steps of the process as soon as it is released from the mould. When released from the mould, the molecule will go through compaction without entangling into steric and topological problems.

Recently the peptidyl transferase activity has been assigned to only the domain V of 23S rRNA. In *in vitro* reaction, a protein free preparation of 23S rRNA could exhibit peptidyl transferase activity[24]. The contribution of the regions of the 23S rRNA other than domain V in this reaction could be to maintain proper conformation of the domain V.

Could the presence of both the peptidyl transferase and the protein-folding activities in the small domain V RNA segment suggest that we are looking at two aspects of the same process? Could this domain scan the polypeptide after it adds every single amino acid to see if it fits the folding mould[21]? Would it reject the polypeptide (fail to recognize) if the added amino acid does not conform to the requirement of folding and the misfolded polypeptide meets natural death by protease degradation?

Is ribosome-mediated protein folding co-translation or post-translational?

During protein synthesis, the growing polypeptide chain remains fairly flexible in the tunnels and channels on the 50S ribosomal subunit. Cross linking of growing polypeptide chain with the 50S particle showed many contacts, especially between the nucleotides in the domain V and the growing polypeptide chain[21]. In fact, an unusually large number of nucleotides in this domain remain exposed from the protein coat of the 50S subunit, presumably because the two major activities, polypeptide synthesis and its folding into active form, are largely due to their contacts with the charged tRNAs (the A and P sites), the growing polypeptide chain, etc.[17]. The polypeptide chain can move on the 50S subunit, making contacts with these nucleotides[21], so that widely separated parts of it can come close together

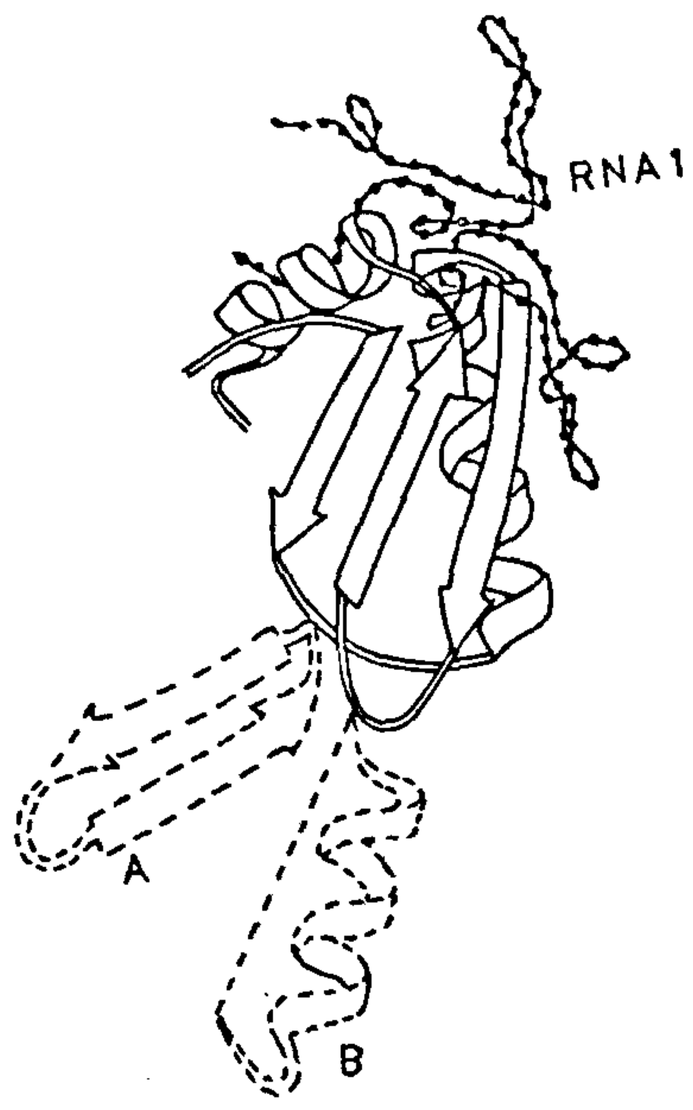

Figure 3. Cartoon showing a polypeptide folding intermediate interacting with domain V RNA. The dashed β-turn(A) and α-helix(B) are regions which could possibly be inserted in the regions of the polypeptide which would not interfere with its interaction with the domain V RNA.

to form the folding intermediates having large part of its secondary structures formed and could even engage in tertiary structure formation[25]. Yet, the final level of folding takes place outside the ribosome[20,26], so the process is called 'post translational'. However, in this case, the released polypeptide chain received the instructions for folding from the ribosome[27].

Similar *in vivo* and *in vitro* experiments have not been done in eukaryotic systems. These need to be done. The existing data suggest that most part of folding takes place on the ribosomes in eucaryotic cells[28,29]. The proteins which are larger on the average than those from bacterial cells and consisting of larger number of domains are folded 'co-translationally', that is, before they are released from the ribosome. It is suggested that the domains can fold independently in co-translational process so that the chance of larger length of polypeptides getting wrongly entangled is reduced. In fact, genes of H Ras and DHFR linked through an oligopeptide linker were shown to make the fused protein having both the H Ras and DHFR domains active in eucaryotic cell whereas in *E. coli*, the fused construct failed to make the protein which would fold both the domains independently. The protein was misfolded forming aggregation[28].

In this context, we have to consider a number of remarkable differences in the process of translation in procaryotic and eucaryotic cells. The most important difference is in the rate of polypeptide chain growth. Whereas in bacterial cells, fifteen amino acids are added to the growing polypeptide chain per second, the number is as low as two to three amino acids in eucaryotic cell. A five hundred amino acid-long polypeptide will, therefore, be synthesized in *E. coli* in about thirty seconds, but will take about three to four minutes in eucaryotic cells. Obviously, even if the mechanism of folding is the same in the two systems, the discrepancy between the rate of protein synthesis and folding could explain the above noted difference. In bacterial cell, the protein will be released before completion of folding, but in eucaryotic cell, folding can go on simultaneously with polypeptide chain growth because the latter is slow. There are other differences also, e.g. (1) transcription and translation are coupled in procaryotes, but are physically separated events in eucaryotes and (2) foreign proteins in a cell could always face an element of uncertainty; even when one bacterial gene is overexpressed by cloning in another bacterial cell, it could run the risk of misfolding and aggregate formation. The problem of misfolding and aggregation appears to be mainly due to the failure of the ribosome to fold the protein when it is produced at a faster rate. It is a general practice to grow *E. coli* at a low temperature (~ 22°C) when cloned genes are expressed so that the rate of protein synthesis remains lower than what would produce inclusion bodies. To confirm that protein folding in eucaryotic cell is co-translational, experiments on *in vivo* folding of a resident gene in eucaryotic cell, along the same line as *in vivo* folding of β galactosidase in *E. coli*[20], are required.

The RNA world and some speculations

The 'RNA world'[30] thus seems to be not mere imagination and theorization by scientists, but a reality. Not only did RNA evolve very early in the process of biological evolution, but we still have this molecule playing crucial enzymic role in principal biological processes. If the large loop of domain V RNA could put the amino acids together in the polypeptide chain and dictate the folding pattern of the protein thus formed, the process of translation in the central dogma of molecular biology could be entirely controlled by such RNA and the precursors of tRNA. What we have here now is RNA-coded message decoded by RNA charged with amino acid and amino acids then joined together and the polypeptide finally folded with the help of a small RNA segment about 300 nt long within the large loop of domain V. The genetic material was RNA before retrogressing into DNA which is less reactive and more stable compared to RNA. The elucidation of the whole process of translation with RNA leaves us with the impression that the central dogma could have evolved in the following di-

rection: Translation and protein folding → Transcription/RNA replication → DNA replication.

The following arguments can be put forth in support of the above scheme. The process of translation and protein folding could be organized by the domain V of ribosomal RNA. The activity has been thoroughly preserved throughout evolution and remains fully valid today. Some components of translation are involved in the process of transcription but not the reverse. Ribosome itself controls transcription in the process of attenuation in tryptophan operon, etc. Ribosomal protein like S10 (NusE) is involved in transcription anti-termination in λ phage gene expression. RNA replication requires translation components like in the $Q\beta$ RNA replicase; EF-Tu and EF-Ts happen to be its components. Transcription components are required further down the line in DNA replication and not the reverse, for example, RNA polymerase synthesizes the RNA primer in DNA replication. Thus the central dogma could have evolved in a direction opposite the way we present it. The biological macromolecule that is at the helm of this process is RNA.

1. Askones, B. A. and Williumson, A. R., *Cold Spring Harbor Symp. Quant. Biol.*, 1967, **32** 223–231.
2. Zubey, G., *Annu. Rev. Genet.*, 1973, **7**, 267–287.
3. Kudlicki, W., Mount, M., Waterscheid, J. P., Kramer, G. and Hardesty, B., *Anal. Biochem.*, 1994, **227**, 12–19.
4. Zipser, D. and Perrin, D., *Cold Spring Harbor Symp. Quant. Biol.*, 1983, **28**, 533–537.
5. Anfinsen, C. B., *Science*, 1973, **181**, 223–230.
6. Laskey, R. A., Honda, B. M., Mills, A. D. and Finch, J. T., *Nature*, 1978, **275**, 416–420.
7. Ellis, R. J., *Nature*, 1987, **328**, 378–379.
8. Ellis, R. J., *Biochem. Biophys. Res. Commun.*, 1997, **238**, 687–692.
9. Hartl, F. U., *Nature*, 1996, **381**, 571–580.
10. Das, B., Chattopadhyay, S. and Das Gupta, C., *Biochem. Biophys. Res. Commun.*, 1992, **183**, 774–780.
11. Chattopadhyay, S., Das, B., Bera, A. K., Dais Gupta, D. and Das Gupta, C., *Biochem. J.*, 1994, **300**, 717–721.
12. Bera, A. K., Das, B., Chattopadhyay, S. and Das Gupa, C., *Biochem. Mol. Biol. Int.*, 1994, **32** 215–223.
13. Das, B., Chattopadhyay, C., Bera, A. K. and Das Gupta, C., *Eur. J. Biochem.*, 1996, **235** 613–621.
14. Chattopadhyay, S., Das, B. and Das Gupta, C., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 8284–8287.
15. Beera, A. K., Das, B., Chattopadhyay, S. and Das Gupta, C., *Curr. Sci.*, 1994, **66**, 230–232.
16. Kudlicki, W., Coffman, A., Kramer, G. and Hardesty, B., *Folding Design*, 1997, **2**, 101–108.
17. Noller, H. F., *Annu. Rev. Biochem.*, 1991, **60**, 191–227.
18. Douthwaite, S., *J. Bacteriol.*, 1992, **174**, 1333–1338.
19. Moazed, D. and Noller, H. F., *Biochimie*, 1987, **69**, 879–884.
20. Chattopadhyay, S., Pal, S., Pal, D., Sarkar, D., Chandra, S. and Das Gupta, C., *Biochim. Biophys. Acta*, 1999, **1429**, 293–298.
21. Choi, K. M. and Brimacombe, R., *Nucleic Acids Res.*, 1998, **26**, 887–895.
22. Kovalic, D., Giannatasio, R. B. and Weisblum, B., *Biochemistry*, 1995, **34**, 15838–15844.
23. Pal, D., Chattopadhyay, S., Chandra, S., Sarkar, D., Chakrabarti, A. and Das Gupta, C., *Nucleic Acids Res.*, 1997, **25**, 5047–5051.
24. Noller, H. F., Hoffarth, V. and Zimniak, Z., *Science*, 1992, **256**, 1416–1419.
25. Fedorov, A. N., Friguet, B., Djavadi-Ohaniance, L., Alakhov, Yu, B. and Goldberg, M. E., *J. Mol. Biol.*, 1992, **228**, 351–358.
26. Netzer, W. J. and Hartl, F. U., *Trends Biochem. Sci.*, 1998, **23**, 68–73.
27. Fedorov, A. N. and Baldwin, T., *J. Biol. Chem.*, 1997, **272**, 32715–32718.
28. Netzer, W. J. and Hartl, F. U., *Nature*, 1997, **388**, 343–349.
29. Komar, A. A., Kommer, A., Krasheninnikov, I. A. and Spirin, A. S., *J. Biol. Chem.*, 1997, **272**, 10646–10651.
30. Gilbert, W., *Nature*, 1986, **319**, 618.
31. Brimacombe, R., *Eur. J. Biochem.*, 1995, **230**, 365–383.
32. Vester, B. and Garrett, R. A., *EMBO J.*, 1988, **7**, 3577–3587.
33. Steiner, G., Kuechler, E. and Barta, A., *EMBO J.*, 1988, **7**, 3949–3955.
34. Mitchell, P., Stade, K., Osswald, M. and Brimacombe, R., *Nucleic Acids Res.*, 1993, **21**, 887–896.

ACKNOWLEDGEMENTS. Research in my laboratory was supported by grants from DAE-BRNS, CSIR and DBT, Govt. of India. The experiments mentioned from the author's laboratory were done by B. Das, S. Chattopadhyay, A. Bera, D. Pal, D. Sarkar, S. Chandra, S. Pal and S. Chaudhuri.