Protein folding activity and the central dogma of molecular biology

Pallavi Ghosh and Dipankar Chatterji*
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India and Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560 064, India

Biological systems, in general, can function effectively when the products of the system are in proper configuration and harmful effects due to misaggregation are avoided. Folding of proteins and their functional consequences have been a subject of active research since several years now. However, it is not clear whether during protein synthesis from genetic message, the same set of rules are employed or participation of new effectors take place. In this review we show that at least in the case of transcription and translation, the part of the machinery act on ‘self’ for the folding activity.

The central dogma of molecular biology as enunciated by Crick in 1958 forms the keystone of modern biology and deals with the detailed residue-by-residue transfer of sequential information from DNA to protein. Close to fifty years of work on each step of this central dogma, i.e., replication, transcription and translation led us to some very interesting observations. It appears now that the machineries involved in transcribing and translating genetic messages themselves have associated protein folding activity for self whereas several proteins that take part in DNA replications complex are well known chaperones for other cellular processes. The reasons for such special benefit thrust onto a particular metabolic path is not clear. The present review aims just to consolidate several such observations in the background of gene expression.

Molecular chaperones in DNA replication

The first implication of chaperones in DNA replication came from the identification of three mutations in dnak, dnaJ and grpE in E. coli that led to an impairment of λDNA replication in their host.3 These mutations have also been shown to lead to similar phenotypes in bacterial physiology and regulation implicating the role of their gene products in E. coli chromosomal replication. dnaK, dnaJ and grpE were later shown to code for molecular chaperones that can assist the folding of many proteins other than those involved in lambda replication. The available literature on the involvement of chaperones in the replication machinery in different organisms is discussed in the following sections.

λDNA replication

The involvement of chaperones in DNA replication has been most extensively studied in bacteriophage λ replication. The key molecule in the replication of both phage and bacterial DNA is the DnaB protein. The steps involved in directing the DnaB protein on Oriλ are briefly outlined below.3

Step 1: The λO protein recognizes and binds a specific sequence in the λ origin sequence (Oriλ) causing a destabilization of an AT-rich region of Oriλ leading to the formation of an open complex.

Step 2: The λP recruits the DnaB protein to the Oriλ by the formation of specific protein–protein interactions between λP and λO. Pre. Biochemical evidences indicate that λP has a higher affinity for DnaB than DnaC does. This property helps λP to sequester DnaB from the host replication pathway and direct it to its own origin (Oriλ). The strong λP–DnaB interactions, which assured its recruitment to Oriλ, now prevents the helicase activity of DnaB; for DnaB to be active, the λP has to dissociate.6,7

Step 3: The λ-DNA replication can proceed up to this stage without the involvement of chaperones. However, the activation of DnaB by expulsion of λP requires the concerted action of the DnaK, DnaJ and the GrpE proteins.8,9 Using 14C-labelled λP protein, Liberek et al.10 have shown that (i) the presence of both DnaK and DnaJ are required for the release of λP from the preprimosomal complex; neither of the proteins alone can carry out this function, (ii) the release is dependent on the hydrolysis of nucleoside triphosphate, (iii) mutants in DnaK cannot cause the release of 14C labelled λP. Subsequently it has been shown that the DnaJ protein first binds to the Oriλ–λO–λP–DnaB complex by virtue of its interaction with λP, λO and DnaB.11,12 The DnaJ protein, by additional interaction with DnaK, helps in positioning DnaK close to the λP protein.5 The DnaJ protein also stimulates the weak ATPase activity of DnaK by 50-fold in this complex.6 ATP hydrolysis results in the dissociation of λP from the preprimosomal complex allowing DnaB to start unwinding the DNA at the Oriλ (refs 9, 12). However, the released λP can rebind to DnaB at Oriλ and inhibit
REVIEW ARTICLES

replication. It is believed that the DnaK protein is released along with λP upon ATP hydrolysis; DnaK remains bound to λP and sequesters it from re-binding to DnaB. Further studies showed that the presence of the nucleotide exchange factor GrpE reduces the requirement of DnaK in the DNA replication reaction. The presence of GrpE has also been shown to enhance the ATPase activity of DnaK and to stimulate the release of ADP from DnaK suggesting that GrpE might facilitate the intracellular cycling of the chaperone DnaK. A hypothesis to explain the activity of GrpE to appreciably lower the concentration of DnaK required for λ replication is that GrpE itself sequesters the λP protein following DnaK action, thus alleviating the requirement for high concentration of DnaK. Purified GrpE also causes the release of λP protein from DnaK (ref. 13).

Step 4: DnaB helicase, liberated from λP begins to unwind the DNA at Ori2. DnaG primase recognizes DnaB single-stranded DNA complex and synthesizes several RNA primers, which are extended into DNA by DNA polymerase III holoenzyme.

E. coli chromosomal DNA replication

The involvement of chaperones in chromosomal DNA replication is not as well defined as in the λ replication. However, there are scattered reports implicating the involvement of chaperones, which are summarized here.

Initiation of replication at the E. coli chromosome begins with the binding of the DnaA protein to a unique sequence called the OriC (refs 14, 15). Like the λO initiator, the DnaA also binds to four DnaA boxes on OriC to form a large complex of 20 or more monomers around which the DNA is wrapped. The DnaA protein then causes an unwinding of an AT-rich region near the OriC, thereby allowing the entry of the DnaB helicase followed by the primase (DnaG) and the DNA polymerase III (refs 17, 18). However, unlike the λ-replication system, chaperones are not required for the activation of DnaB helicase activity as these replications use DnaC to transfer DnaB to the origin.

One of the first indications that the DnaK/I system is involved in chromosomal replication comes from genetic studies in which dnaK, dnaJ and grpE mutants show abnormalities in cell division, chromosome segregation and DNA synthesis. Subsequent involvement of the DnaK/DnaJ was implicated when a dnaK mutant of E. coli containing the mutation dnaK III was found to be conditionally defective in the initiation of DNA replication. The dnaKIII mutant was defective in replication initiation when the cells were transferred to high temperatures implicating the dnaK gene in the initiation of E. coli replication at high temperatures. Replication became temperature resistant upon inactivation of the rhs gene (encoding Rnsh) suggesting that the DnaK function is involved in the dnaA-dependent initiation at the OriC at least at high temperatures. Subsequently, it was demonstrated that the DnaA initiator protein is very prone to aggregation. Only the DnaA monomer has the ability to initiate replication whereas the aggregated form does not. It had been proposed earlier that DnaA protein is regulated in the cell by conversion of a membrane-bound aggregated form of DnaA to a monomer form. DnaK has been shown to be able to activate the DnaA protein in an ATP-dependent manner by disassembly of aggregated DnaA from phospholipid complexes where phospholipids have similar composition as those in the cell membrane. Conversion of inactive DnaA–phospholipid complexes to the active form may contribute to the regulation of initiation of chromosomal replication of E. coli.

The function of GrpE and DnaJ in this process has not been unequivocally demonstrated till date but has been proposed to be similar to the mechanism used by DnaK to disassemble λP from Ori2. DnaK has also been proposed to interact with nascent chains of DnaA during its synthesis on the ribosome and prevent it from aggregation until it is properly folded into its native replication competent conformation.

In addition to DnaK, GroEL and GroES also have the ability to inhibit the aggregation of DnaA in the presence of ATP. GroEL and GroES have also been shown to overcome the block in E. coli replication initiation caused by the dnaA46 mutation which is defective in binding to OriC. Furthermore, mutations in GroEL also drastically reduce the rate of DNA synthesis at elevated temperature. Besides these implications, the exact role of these chaperones is still not elucidated.

Another protein of the replication apparatus that has been shown to be affected by chaperones is the epsilon (ε) subunit of the DNA polymerase III holoenzyme, encoded by dnaQ. Mutations in dnaK, dnaJ and grpE result in a dramatic decrease in the levels of the epsilon subunit. Biochemical studies suggest that these chaperones interact with the nascent chain of the epsilon and protect it from degradation.

Molecular chaperones in transcription

Transcription is the first step in the expression of genetic information involving the synthesis of RNA from the DNA template and carried out by a single multi-subunit protein called the DNA-dependent RNA polymerase (RNAP). The subunit composition and mechanism of action of the E. coli enzyme, which forms a paradigm for the study of all RNA polymerases, are described in greater detail in several reviews.

The first indication of chaperones in the transcription machinery came from studies on the dnaK and dnaJ mutants of E. coli which showed a block in RNA synthesis at non-permissive temperatures. Skelly et al. then
showed that the DnaK protein is often associated with purified E. coli RNA preparations. However, the first systematic study on the involvement of DnaK and DnaJ proteins in the transcription machinery came in 1990 when it was proposed that DnaK can protect RNA from heat inactivation in an ATP-dependent manner. In addition, it can disaggregate preformed aggregates of RNA and reanimate enzyme activity. This was followed by another report which showed that the chaperone GroEL could also protect RNA polymerase from heat inactivation in an ATP-dependent manner, better than any other protein like Dnak. However, these studies do not elucidate the biological significance of the chaperones on RNA polymerase. Their relevance in vivo was first suggested in 1999 by a pulse chase experiment carried out on E. coli cells in order to determine the set of newly synthesized proteins interacting with GroEL. At different times of chase, the cells were lysed, GroEL–substrate complexes were isolated by immunoprecipitation, separated on 2-D PAGE and identified using mass-spectrometry. The α, β and ω subunits were shown, by this method, to be substrates of GroEL in vivo. The β and ω subunits interact with GroEL and are released during the chase period. However, α remains associated with GroEL even after the chase, implying the requirement for GroEL in its initial folding as well as for its subsequent maintenance. This implies that GroEL remains associated with RNA polymerase even after the initial folding of individual subunits and their assembly into a functional enzyme. This is in agreement with the observation made decades back that an ATPase, later identified to be GroEL, co-purifies with RNA preparations from E. coli.

However, while dissecting the structure–function relationship in case of individual subunits of RNA polymerase we noticed that no clear-cut functional role had been ascribed to ω-subunit, a small 10 kDa protein coded by rpoZ gene. Deletion of ω-subunit only showed slow growth phenotype in E. coli which, however, could also be due to the polar effect of a neighbouring gene. Later, we observed that ω-subunit helps in the recovery of the activity of denatured RNA polymerase and thus we proposed that the major role of ω is the chaperone-like activity for RNA polymerase alone. Crystal structure determination of T. aquaticus RNA polymerase and sequence homology of rpoZ gene across species supported our initial observation further.

Subsequently, an omega-less mutant of RNA polymerase was shown to accumulate large amounts of GroEL, significantly higher than the levels associated with wild-type E. coli RNA preparations. Removal of GroEL from ω-less polymerase led to a completely inactive enzyme that even lacked the ability to bind sigma. Addition of ω and/or GroEL to the ω-less enzyme could restore the activity to that of the wild type. These observations indicated that omega was carrying out a GroEL-like function in the folding/maintenance of the structure of RNA polymerase. Further studies showed that ω is required to prevent the aggregation of the β′ subunit (incidentally the only subunit that is not folded by GroEL in vivo), maintain its structure and recruit it to the αβ subassembly to form a functional core. Figure 1 shows a schematic representation of the reconstitution of RNA polymerase with or without ω-subunit.

**Molecular chaperones in translation**

The final step in gene expression is the translation of nucleotide sequences in mRNA to a protein sequence. Ribosomes are macromolecular assemblies made up of proteins and RNA involved in the process of translation. Bacterial ribosomes which are the most extensively studied are composed of a small (30S) and a large (50S) subunit which associate together to form a complete 70S ribosome. The ribosome accomplishes the process of protein synthesis by three distinct steps of initiation, elongation and termination, each of which require accessory factors. The 3OS subunit (composed of a 16S rRNA and 20 ribosomal proteins) is responsible for binding mRNA and the anticodon arm of tRNA whereas the 50S subunit (consisting of 5S RNA, 23S RNA and 31 proteins) contains the peptidyl-transferase centre that catalyses the formation of peptide bonds.

The broad scheme of translation was worked out many decades back. The precise mechanistic details have, however, been revealed only in the last decade from high-resolution structures of 30S, 50S subunits and 70S ribosome. The most startling discovery from these structures is that the decoding of information on mRNA as well as catalysis of peptide bond formation is the province solely of ribosomal RNA.

Reconstitution of completely active ribosomal subunits in vitro has been achieved with purified ribosomal proteins and RNA indicating that the assembly of a ribosome is a self-sufficient process and does not require the presence of external factors. Both the ribosomal constituents, however, have very complicated structures. The 16S and 23S RNA have a very intricate secondary structure composed of regular double helices connected by single stranded loops with numerous stacking interactions between the helices. The ribosomal proteins contain one or more globular domains along with characteristic long extensions in the form of α-hairpins, β-hairpins or loops and extended carboxy- or amino-terminal tails; some are entirely extended lacking tertiary and in many cases even secondary structures.

Although the requirement for additional factors/chaperones has not been identified in the assembly of the subunits, it may well be imagined that chaperones could be involved in preventing the aggregation of ribosomal proteins after their synthesis as these are disor-
Figure 1. Depiction of the folding pathway of E. coli RNA polymerase in the presence and absence of omega subunit.

Figure 2. A generalized scheme for the ribosomal RNA mediated protein folding.
dered in the absence of RNA and thereby prone to aggregation. A preliminary analysis also shows that ribosomal proteins L7 and L12 require GroEL for their folding\textsuperscript{39}. The presence of chaperones have also been implied in the case of eukaryotic ribosomal proteins that are assembled in the nucleolus, a highly crowded organelle\textsuperscript{45}. The assembly process requires the presence of a number of non-ribosomal proteins\textsuperscript{49}. One such protein, B23 protein has been shown to have properties of molecular chaperones and the B23-ribosomal protein complexes have been implicated to serve as a transition phase in the assembly of these proteins into ribosomes\textsuperscript{50}.

The ribosome as a chaperone

Proteins, synthesized vectorially from their N-termini to their C-termini on ribosomes, undergo co-translational folding as they extend from the peptidyl-transferase centre through the ribosome exit tunnel of \(\sim 100\) Å (ref. 51). The nascent peptide is sheltered from external proteins and proteases and offered a conducive environment for folding in this exit tunnel\textsuperscript{51}. Structures higher than a helix cannot form inside this tunnel as it has a width of only 15 Å.

In addition to providing a sheltered environment, there is evidence to suggest that the ribosome itself can chaperone protein folding. In a series of papers, Dasgupta et al.\textsuperscript{22,53}, showed ribosome can assist folding of proteins in a wide variety of systems very efficiently. This activity has finally been attributed to domain V of the 23S RNA, which has been shown to assist the renaturation of denatured proteins\textsuperscript{54,55} suggesting that the peptidyl-transferase centre may be acting as a folding modulator as it simultaneously catalyses the formation of a peptide bond. Accessory factors of the ribosome like IF-2, EF-Tu and EF-G have also been shown to possess chaperone activities\textsuperscript{56,57} and may further assist co-translational folding. It has been suggested that translation factors may have been ancestral chaperones dedicated to folding of nascent polypeptides before the evolution of general chaperones. Figure 2 depicts a general scheme for the ribosomal RNA-mediated folding of proteins.

In summary, not much literature is available on the folding of individual subunits of the replication apparatus. Only two proteins, DnaA and the \(\varepsilon\) subunit, have been shown to require chaperones for their folding in the absence of which they aggregate. Although DnaK/J were first discovered in replication, they are not dedicated to this machinery but are required to assist the folding of a variety of proteins. However, there are a few limitations in the chaperone system, such as the DnaK and DnaJ deletion strains of \(E.\ coli\) still growing slowly within a narrow range of temperature around 30°C (ref. 58). GroEL assists folding only for a minority of cytosolic proteins (5%). Interestingly, there is another protein known as trigger factor (TF) which is ribosome-bound\textsuperscript{59} and is associated with nascent polypeptides. This is a ribosome-associated chaperone and displays functional cooperativity with DnaK\textsuperscript{60}, and prolyl isomerase activity\textsuperscript{61}.

The components of the transcription and translation machineries seem to have evolved both dedicated chaperones in addition to utilizing the general chaperones. One of the subunits of RNA polymerase, \(\alpha\), helps in maintaining the structure of another subunit \(\beta\) whereas the other components require general chaperones for their folding. The ribosome also uses general chaperones for preventing the aggregation of the unstable ribosomal proteins until they become associated with RNA which then take over the function of these general chaperones in maintaining their structure.

REVIEW ARTICLES


ACKNOWLEDGMENT. We thank the Council of Scientific and Industrial Research (CSIR) for financial support through NMITLI grant. P.G. was a recipient of CSIR fellowship. We also thank the anonymous reviewer for bringing our attention to the role of Trigger factor and DnaK in protein folding.

Received 6 March 2003; revised accepted 3 June 2003