Mechanistic Insights on the Folding of a Large Ribozyme during Transcription†

Terrence Wong,‡ Tobin R. Sosnick,*,‡,§ and Tao Pan*,‡

Department of Biochemistry and Molecular Biology and Institute of Biophysical Dynamics, University of Chicago, Chicago, Illinois 60637

Received November 19, 2004; Revised Manuscript Received March 21, 2005

ABSTRACT: RNA folding during transcription resembles folding in a cellular environment. We previously investigated the folding of a large ribozyme derived from a bacterial RNase P RNA during its transcription by the *Escherichia coli* RNA polymerase and the effect of the elongation factor NusA. We found that transcriptional pausing at a specific site induced by NusA significantly altered the folding pathway. In this work, we compare folding during transcription by the *E. coli* RNA polymerase of circularly permuted variants and site-specific mutants of the RNase P ribozyme to elucidate the molecular mechanism of transcriptional pausing and RNA folding. The effect of NusA-induced pausing depends on the order of RNA synthesis and only affects local folding of the RNA. Pausing likely prevents a misfolded structure between the 5′ strand of a helix and its adjacent junction located in the specificity domain and a region known to bind single-stranded RNA located in the catalytic domain. These results lead to a structural model on how transcriptional pausing affects folding of RNase P RNA. Structural rearrangements of a nascent RNA transcript enhanced by transcriptional pausing may be a general feature of RNA folding during transcription.

RNA folding during transcription is dictated by two fundamental aspects of transcription, the 5′ to 3′ polarity of RNA synthesis coupled with the transcription speed and site-specific pausing (*1*−*8*). Local RNA structures can form easily in the absence of the global tertiary structure. Prior to the completion of the transcription, therefore, the nascent RNA transcript already contains many local, nativelike structures. The polarity and transcription speed largely determine which local structure forms during transcription. Transcriptional pausing, in contrast, results from interactions between the elongating RNA structure with the polymerase and its associated factors (*4*, *9*, *10*). This RNA−protein interaction could alter the nascent RNA structure that is present at the time of the pause. We have demonstrated previously that transcriptional pausing at a specific site can alter the folding during transcription for a large ribozyme (*5*).

This ribozyme is a circularly permuted variant of the RNA component from *Bacillus subtilis* RNase P (denoted P RNA).[1] P RNA generates the mature 5′ end of all tRNAs in the cell through an endonucleolytic cleavage reaction (*11*, *12*). The bacterial RNase P is made of two subunits, a large RNA of ∼300−400 nucleotides and a small protein of ∼14 kDa. The RNA component contains the active site of this enzyme and the binding site of the tRNA substrate. These two functions, catalysis and tRNA binding, are separated onto two independently folding domains, the catalytic C-domain and the specificity determining S-domain (*13*, *14*). The circularly permuted ribozyme with its 5′ end at residue 240 (CP240) is constructed such that the entire C-domain is transcribed first, followed by the S-domain (Figure 1).

Folding of CP240 followed two different pathways depending on the presence of a transcriptional pause at residue U55 (the 55th nucleotide of the natural P RNA and the 225th nucleotide of CP240, Figure 1B). Pausing at this site was induced by the addition of an elongation factor, the NusA protein. Pausing did not change the folding rate of the C-domain. In the absence of NusA, this rate was 3.5 times faster than that of the S-domain. In the presence of NusA, the S-domain folded at least 3 times faster than the C-domain. This effect was entirely dependent on pausing at this site, since NusA had no effect on the folding of a ribozyme mutant lacking this pause site or when a nonpausing RNA polymerase mutant was utilized (*5*).

To understand the relationship between site-specific pausing and RNA folding during transcription, we investigate the folding of several circularly permuted P RNA variants and site-specific mutants in this work. Circular permutation indicated that the effect of transcriptional pausing on RNA folding depends on the order of RNA synthesis. We identify local regions and structural motifs of the ribozyme affected by the pausing at U55. The results lead to a molecular model to explain how transcriptional pausing affects the folding pathway of this large ribozyme.

MATERIALS AND METHODS

Sources of Proteins. Escherichia coli RNA polymerase and NusA were obtained as previously described (*15*, *16*). RNase H was purchased from Epicentre.

† Supported by a grant from the NIH (GM57780). T.W. was supported by the Medical Scientist Training Program.

* Corresponding authors: taopan@uchicago.edu and trsosnic@uchicago.edu (e-mail); 773-702-4179 (phone); 773-702-0439 (fax).

‡ Department of Biochemistry and Molecular Biology, University of Chicago.

§ Institute of Biophysical Dynamics, University of Chicago.

[1] Abbreviations: CP240, the circularly permuted P RNA with the 5′ end at residue 240; C-domain, catalytic domain consisting of residues 240−409 plus 1−83; P RNA, the *Bacillus subtilis* RNase P RNA component; S-domain, specificity domain consisting of residues 84−239.

10.1021/bi047560l CCC: $30.25
Published on Web 04/28/2005

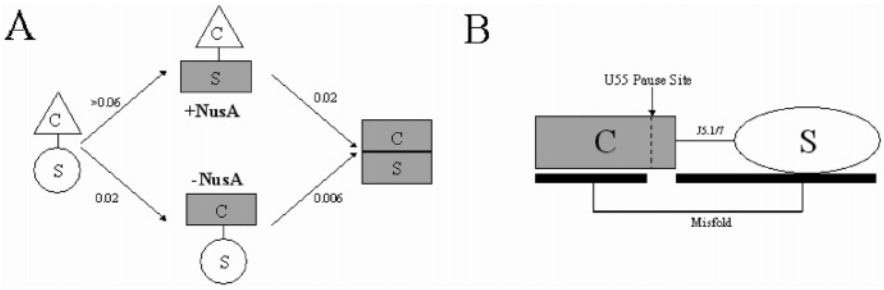

FIGURE 1: (A) Effect of NusA on the folding of CP240 during transcription (results from ref *5*). Shaded boxes indicate folded domains, and open symbols represent not yet folded domains. In the absence of NusA, CP240 folds through an intermediate containing a folded C-domain, and folding of the S-domain is rate-limiting. In the presence of NusA, CP240 folds through a different intermediate containing a folded S-domain, and folding of the C-domain is rate-limiting. (B) The misfolded intermediate in the absence of NusA. A portion of the C-domain transcribed before the pause site forms a non-native structure with a portion of the S-domain, resulting in the slow folding of the S-domain. This misfolded structure is avoided when RNA polymerase pauses at U55 in the presence of NusA.

Transcription in Vitro. Templates for transcription of the mutant constructs and circular permuted P RNAs were generated by PCR and confirmed by sequencing.

Multiple-round transcription was performed as previously described (*5*). Briefly, transcription using *E. coli* RNA polymerase was carried out in 40 mM Tris-HCl, pH 7.9, 10 mM 2-mercaptoethanol, 4 mM spermidine, 1 mM each of ATP, GTP, CTP, and UTP, 5−10 μCi of [α-^{32} P]ATP, 100 μg/mL BSA, 10 mM $MgCl_2$, 120 mM KCl, 0.05−0.1 μM DNA template, 0.2 μM *E. coli* RNA polymerase, and 0 or 0.4 μM NusA protein. Transcription was carried out for 8 min with the removal of aliquots at various time points. When necessary, folding was allowed to proceed for 40 min after stopping the transcription at 8 min by the addition of 2 μM rifampicin.

Single-round transcription was performed under standard conditions with minor modification (*17*). Briefly, ternary complexes were formed at 50 nM in transcription buffer (20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 14 mM $MgCl_2$, 14 mM 2-mercaptoethanol, 0.1 mM EDTA, 25 μg/mL BSA, 2.25% glycerol) with or without 0.4 μM NusA protein. Through the inclusion of 150 μM GpC dinucleotide and 2.5 μM ATP, GTP, and CTP in the transcription buffer, transcription proceeded to A14 of the CP240 construct. After a 10 min incubation, elongation was resumed with the addition of 200 μM (final concentration) ATP, UTP, CTP, and GTP and 5−10 μCi of [α-^{32}P]ATP. Transcription was limited to a single round with the concurrent addition of rifampicin to a final concentration of 2 μM. Transcription was carried out for up to 8 min with removal of aliquots at designated time points.

Folding Monitored by Catalytic Activity. Catalysis assays were performed as previously described (*5*). Equal volumes of the transcription reaction and trace amounts of ^{32}P-labeled pre-tRNA substrate (<2 nM) were mixed together. Both the S-domain and the C-domain must be folded for the RNase P ribozyme to be catalytically active against this substrate. The final reaction condition was 50 mM Tris-HCl, pH 8.1, 100 mM $MgCl_2$, and 0.6 M KCl. Aliquots were taken from the cleavage reaction within 7−21 s and quenched with an excess of EDTA and urea. The time allowed for the cleavage reaction (<21 s) is short compared to the folding time of the C-domain in the RNA transcripts ($t_{1/2}$ > 35 s). Therefore, only a small fraction of RNA would fold during the cleavage reaction at high concentrations of monovalent and divalent ions.

Data analysis was performed as previously described (*5*). Briefly, the relative amount of transcript synthesized [$S(t)$] was determined by detecting the incorporation of [α-^{32}P]-ATP into the RNA. The cleaved substrate and product were separated on denaturing polyacrylamide gels, and the amount of the cleaved substrate was quantified by phosphorimaging. The relative amount of active ribozyme [$A(t)$] was determined from the rate of substrate cleavage. The data were fit to the equation:

$$\frac{A(t)}{S(t)} = 1 - C\left(\frac{1 - e^{-k_f t}}{k_f t}\right) \quad (1)$$

For each construct, the $A(t)/S(t)$ ratio was normalized to the ribozyme's intrinsic activity. Occasionally, the x-intercept of this curve was >0, because the *E. coli* RNA polymerase must first finish synthesizing the transcript before the ribozyme becomes catalytically active. A constant term (C) was added to eq 1 to account for this lag observed in some reactions.

Folding Monitored by Oligonucleotide Hybridization. The oligohybridization assay was a modified version of the method developed by Zarrinkar et al. (*18*). Aliquots of the transcription reaction were taken and added to an equal volume of oligonucleotide−RNase H reaction solution consisting of 20 mM Tris-HCl, pH 7.5, 20 mM KCl, 26 mM $MgCl_2$, 0.1 mM EDTA, 0.1 mM DTT, 4 μM rifampicin, 200 μM DNA probe, and 0.4 unit/μL RNase H. The RNase H cleavage reaction proceeded for 30 s before quenching with an excess of EDTA and urea. The amount of cleaved and uncleaved RNA transcripts was identified by phosphorimaging after separation on denaturing polyacrylamide gels.

RESULTS

A Preliminary Model of Pausing and Folding during Transcription. Our previous results suggested that in the absence of NusA, and hence, no pausing at U55, the folding of the S-domain was rate-limiting in the folding of CP240 during transcription (ref *5*; Figure 1A). NusA-induced pausing at U55 accelerated the folding of the S-domain by at least 10-fold. The pause site at U55 was located toward the 3′ end of the C-domain. Pausing at this site altered the folding of the yet to be synthesized S-domain, while having no effect on the folding of the C-domain. These results led to a folding model in which a misfolded structure formed between the portions of the C-domain transcribed before the

pause site and the downstream S-domain (Figure 1B). This misfolded structure temporarily blocked folding of the S-domain and had to be disrupted in order for folding of the S-domain to proceed. The pausing at U55 permitted the folding and sequestering of an already synthesized region, thereby preventing the formation of the misfolded structure. As a result, the folding of the S-domain could proceed unhindered.

Effect of NusA Examined by Oligonucleotide Hybridization. Our previous study used catalytic activity assays that identified whether the entire C-domain or the S-domain was folded. In the present study, resolution at the level of structural motifs was obtained through the use of the oligonucleotide hybridization strategy employed previously in the Mg^{2+}-initiated refolding studies of the group I intron (*19*, *20*) and RNase P RNA (*18*). This assay relied on the hybridization of DNA oligonucleotides to their complementary regions in the RNA. Hybridization was more rapid when the corresponding RNA region was exposed, for example, prior to its folding. Conversely, hybridization was slower when the corresponding RNA region was buried. The fraction of oligonucleotide hybridized during a fixed amount of time was identified by the RNase H cleavage of the radioactively labeled RNA transcript.

As CP240 folded, its structural motifs became more protected from binding to complementary DNA probes. Snapshots taken at various time points identified the degree of protection against the different probes and, consequently, provided site-resolved information about CP240's folding pathway. Seven different probes were used, targeting five areas in the C-domain and two in the S-domain. The probes for the C-domain targeted the 5′ and the 3′ strands of the P2 helix, the 5′ and the 3′ strands of the P4 helix, and J18/2. All three regions are part of the active site of this ribozyme (*14*, *21*). One specificity domain-specific probe targeted J11/12, which was involved in a tertiary structure required for substrate binding (*22*). The second probe targeted the 5′ strand of the P7 helix and its adjacent single-stranded junction, J5.1/7 (P7-J5.1/7 region).

Compared to the multiple-round transcription assays, the single-round assays were more useful in identifying the folding behavior of specific structural motifs (data not shown). To synchronize transcription, halted complexes were formed at the 14th nucleotide of CP240 through the omission of UTP. Transcription resumed upon the addition of UTP and was limited to a single round by the concurrent addition of rifampicin. For all seven probes, we were able to detect both the full-length and cleaved constructs, allowing us to determine the RNA's degree of protection against each probe (Figure 2A,B).

Five of the seven RNA motifs analyzed exhibited a level of protection that was independent of NusA (Figure 2C). Therefore, it appears that CP240's folding pathway in the presence of NusA is globally similar to its pathway in the absence of NusA as the majority of structural motifs were not influenced by the altered transcriptional pausing.

NusA-induced pausing at U55, however, decreased the level of protection against the DNA probe targeting the P7-J5.1/7 region (Figure 2B), making this region a potential candidate for being part of the misfolded structure between the C- and S-domains (Figure 2D). While involved in the misfold, this region would be inaccessible to its DNA probe.

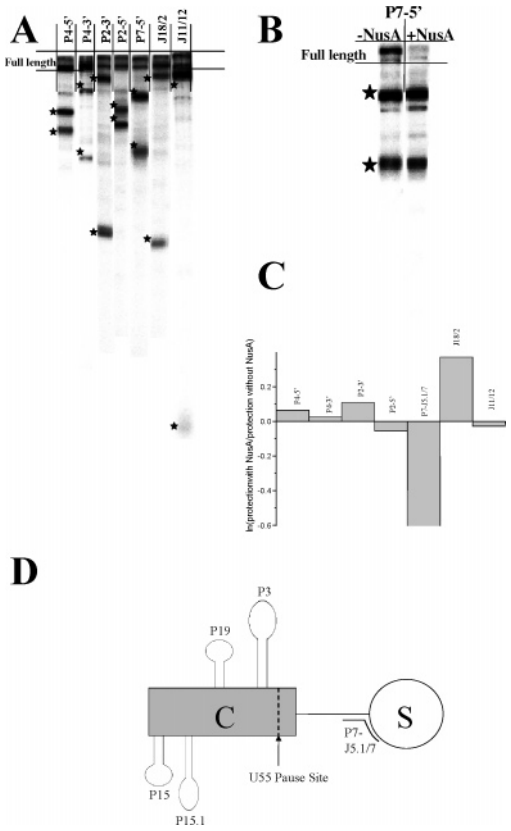

FIGURE 2: (A) An example of oligohybridization analysis of CP240 using seven DNA probes during single-round transcription. The uncleaved (full-length) RNA is at the top. Stars indicate the cleaved products. (B) Protection of CP240 from the P7-J5.1/7 probe at 1.25 min either in the absence or in the presence of NusA. (C) Effect of NusA on the level of protection during single-round transcription. The average ratios of protection in the presence versus the absence of NusA at three time points (1.25−4 min) are shown. In the presence of NusA, the P7-J5.1/7 region has decreased protection while the J18/2 region has increased protection. (D) Candidate structural motifs in both domains that may participate in the misfolded structure.

Transcriptional pausing at U55 would allow CP240 to fold without going through this misfold, and the P7-J5.1/7 region would be more transiently exposed during CP240 folding.

The addition of NusA also increased protection against the DNA probe targeting the J18/2 region in the C-domain. Was the increased protection in J18/2 related to or independent of the interdomain misfold illustrated in Figure 1B? The catalytic activity of the C-domain was examined previously by the cleavage of an in vitro selected substrate (*23*). On the basis of this assay, NusA-induced pausing had no effect on C-domain folding. The J18/2 region, however, was a part of the active site. Therefore, this region needs to be folded in order for there to be catalytic activity. Hence, the null effect of NusA-induced pausing on C-domain activity argues that the increased protection in J18/2 due to pausing was likely to be unrelated to the interdomain misfold.

Generally, to have a properly folded active site, the junctions J5/15, J15/15.1, J18/2, J19/4, and J3/4 and the helices P2 and P4 need to be folded (Figure 3, refs *14* and *21*). The catalytic activity against the selected substrate also required the L1 loop for substrate binding (*23*). What remains outside these elements are the P15, P15.1, P19, and P3 regions. Residues within these regions were therefore

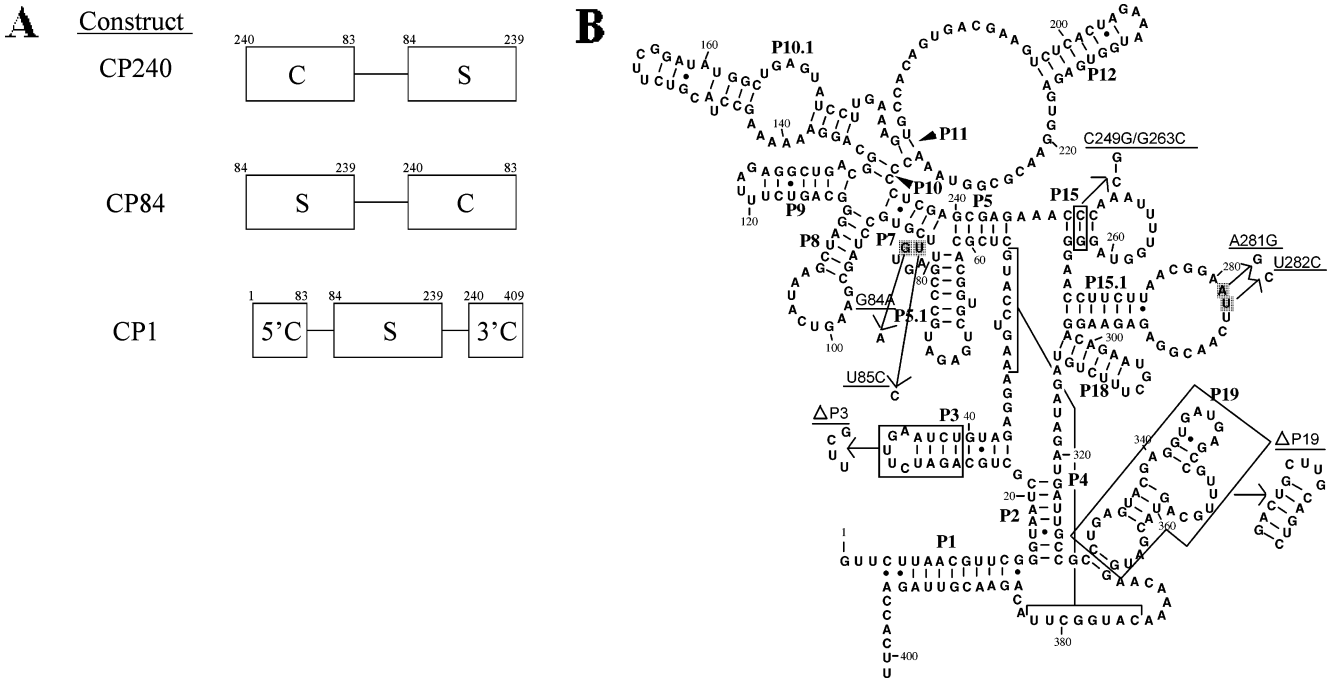

FIGURE 3: (A) Order of synthesis in the circular permutants studied. (B) Secondary structure of the *B. subtilis* RNase P RNA with sequence changes made in the mutant constructs indicated.

candidates for the formation of the interdomain misfold in the absence of NusA (Figure 2C).

The NusA Effect on Folding Depends on the Order of Transcription. The polarity of transcription suggested that folding of a tertiary RNA structure during transcription may depend on the order of the RNA synthesis. In the CP240 construct used so far, the order of synthesis is C-domain → S-domain. A common way to change the order of RNA synthesis is through circular permutation in which the normal 5′ and 3′ ends are covalently joined and new 5′ and 3′ termini introduced elsewhere (*24*). We made two circularly permuted RNA constructs to discover the role transcription order plays in determining the effect NusA-induced pausing has on folding (Figure 3A). The first construct had the 5′ end at nucleotide 84 (denoted CP84), and the order of synthesis was S-domain → C-domain. The second construct was the natural *B. subtilis* P RNA (CP1) in which one-third of the C-domain (5′C) was transcribed first, the S-domain second, and the remaining two-thirds of the C-domain (3′C) last. The critical U55 pause site had a very different context in these constructs. In CP84, this pause site was near the end of the transcript. In CP1, this pause site was present when less than one-third of the C-domain had been transcribed (55 out of 255 residues). In CP240, this pause site was present when almost the entire C-domain had been transcribed (225 out of 255 residues).

Folding of all circularly permuted constructs was analyzed by measuring the catalytic activity, normalized to the amount of the ribozyme synthesized, using a pre-tRNA substrate (*5*). In this assay, the observed catalytic efficiency was derived from two effects: the amount of the folded ribozyme transcript and the intrinsic catalytic activity of the CP RNAs. We determined the intrinsic activities by allowing these CP ribozymes to fold for an additional 40 min after stopping the transcription at 8 min. Because the folding of the RNA transcript is complete after this 40 min incubation, the relative activity at the 40 min time point identified the intrinsic activity of the CP ribozymes. This information allowed us to fit the data for each CP ribozyme according to their own intrinsic catalytic activity.

Because of the low intrinsic catalytic activity of the CP1 construct and CP84 toward the in vitro selected substrate, folding of the catalytic domain alone within these two CP constructs could not be reliably analyzed for the selected substrate (data not shown).

Folding of both constructs proceeded at a rate similar to that of CP240 in the absence and the presence of NusA (Table 1). This result is consistent with these circular permutants having the same rate-limiting step in folding during transcription. For example, a change in the rate-limiting step of CP240 in Mg^{2+}-induced refolding altered the folding rate by as much as 15-fold (*25*). The addition of NusA did not affect the folding rate of either construct (Table 1, Figure 4A), even though NusA still induced pausing at U55 during CP1 transcription (ref *26* and data not shown). Thus, the effect on folding of NusA-induced pausing was highly context dependent. Two scenarios may explain this context dependence, assuming the conservation of the rate-limiting step. In the first scenario, applicable to CP84, regions involved in the misfolded structure already were transcribed before the pause at U55, which was located near the end of the transcript (381 out of 409 residues). In the second scenario, applicable to CP1, none of the regions involved in the misfolded structure were transcribed before pausing at U55, which was located near the beginning of the transcript (55 out of 409 residues). In CP240, pausing at U55 occurred between the two (or more) regions involved in the misfolded structure.

The above results suggest additional information on the location of the region in the C-domain involved in the misfolded structure. In CP1, the 5′ portion of the C-domain was synthesized before pausing at U55; hence the misfolded structure likely involved the 3′ portion of the C-domain (Figure 3A). In CP84, the misfolded structure formed before

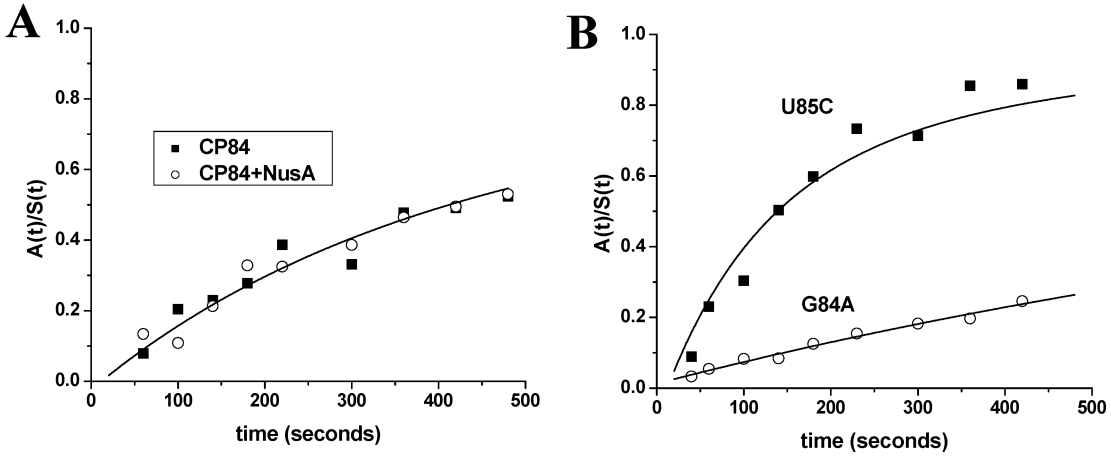

FIGURE 4: (A) Folding of CP84 in the absence and presence of NusA. CP84 folds at the same rate [$(4.0 \pm 0.6) \times 10^{-3}$ s^{-1}] both in the presence and in the absence of NusA. (B) Effects of mutations near the P7 helix on the folding rate. The U85C mutation accelerates folding [$(13 \pm 3) \times 10^{-3}$ s^{-1}] while the G84A mutation significantly slows folding [$(1.3 \pm 0.1) \times 10^{-3}$ s^{-1}].

Table 1: Relative Folding Rates of CP RNAs and CP240 Mutants during Transcription at 37 °C

region of interest	construct	relative folding rate
pseudo-wild-type	CP240	1.0 (3.4)[a]
circular permutants	CP84	0.98 (0.98)
	CP1	0.85 (0.85)
P7-J5.1/7 region	G84A	0.32
	U85C	3.2
P3 and P19	ΔP3	1.0 (6.3)
	ΔP19	0.88
P15 stem	C249G/G263C	1.0
P15.1 loop	A281G	0.95
	U282C	1.0
compensatory mutations	U282C/G84A	0.32
	A281G/U85C	2.8

[a] Ratios in parentheses are from transcription in the presence of NusA.

the S-domain was fully folded. Folding of the S-domain during transcription occurred quickly, probably in less than 5 s (*5*). Hence, the region in the 3′ C-domain involved in the misfolded structure likely was transcribed directly after the S-domain.

S-Domain and C-Domain Regions Involved in the Misfolded Structure. To further define the region(s) in the S-domain and the C-domain involved in the misfolded structure, site-specific and deletion mutants in the context of CP240 were analyzed (Figures 3B and 4B and Table 1). These mutations belong to three groups. The first group contained two separate single point mutations in the region immediately adjacent to the 5′ strand of P7. The second group contained two different deletions of large portions of the P3 and P19 helix. The third group had five variants, with either single or multiple mutations in the P15 stem, P15.1 loop, or in both the P7-J5.1/7 region and the P15.1 loop. As with our circularly permuted constructs, the folding of these mutants was analyzed by monitoring the appearance of their catalytic activity toward the pre-tRNA substrate.

Guided by the oligohybridization data, our search for the S-domain region focused on the 5′ strand of the P7 helix and its neighboring residues in J5.1/7 (Figure 3B). In the native structure, the J5.1/7 region was not involved in tertiary interactions (*14*). Therefore, mutations of these nucleotides in J5.1/7 were expected to have little effect on the native structure or the activity of the ribozyme, which made them suitable candidates to probe the origins of the misfolded structure.

A significant change in the folding rate during transcription was observed for both the G84A and U85C mutants (Figure 4B). Compared to CP240, the folding rate of the G84A mutant was ~3-fold slower, while the rate of the U85C mutant was ~3-fold faster (Table 1). These results were consistent with the oligohybridization data. As mutating these two nucleotides caused the folding rate to vary by an order of magnitude, this region likely was involved in the misfolded structure. The 5′ strand of the P7 helix also was implicated in the misfolded structure, which would slow the formation of the P7 helix and, consequently, the S-domain tertiary structure. The G84A mutation exacerbated this misfolded structure while the U85C mutation alleviated it.

Searching for the C-domain region involved in the misfolded structure was more challenging. The folding intermediate containing the misfolded structure could cleave the selected substrate, implying that the active site and the L1 substrate binding site were properly folded. Hence, regions involved in the active site and the L1 loop were excluded from the search. The P18 region is a short, local hairpin which likely folded immediately upon its transcription, so this region was excluded as well. Thus, the ensuing mutational analysis focused on the four remaining C-domain motifs: P3, P19, P15, and P15.1/L15.1 (Figures 2C and 3B).

Both P3 and P19 have little sequence conservation among B-type P RNA (*27*). In addition, P19 has a high degree of structural variability. The deletion of a portion of both helical structures was therefore expected to have little effect on catalytic activity. P3 was of particular interest as it is the structural element immediately preceding the U55 pause site. Pausing at U55 could have an impact on the folding of P3. P19 has two internal loops and short helical stretches. These loops and a portion of the helical regions may interact temporarily with the P7-J5.1/7 region to generate the misfolded intermediate. We analyzed folding during transcription of two mutants with either P3 or P19 replaced with a short stem−loop (ΔP3 and ΔP19). The folding rates of both mutants were comparable to that of the wild-type CP240. These results suggested that neither P3 nor P19 was involved in the interdomain misfold.

In contrast to P3 and P19, the P15 and P15.1 regions have significant sequence conservation among B-type P RNA (*27*). Hence, major substitutions in these regions could be detrimental to the native structure and catalytic activity of this ribozyme.

In light of the significant effect of the G84A and U85C mutants, we searched for sequences within P15 and P15.1 that could have base-pair complementarity to the 5′ strand of the P7 helix and its adjacent J5.1/7 junction or the 3′ strand of the helix. Such complementary regions necessary to the formation of the misfolded structure could be eliminated by single site mutations. Four consecutive C's near or within the 3′ strand of P7 (C232−C235) could complement a region of four G's in the P15 region (G262-G265). A C249G/G263C mutant would maintain the base pairs in the P15 stem, but it could disrupt a potentially misfolded helix involving these two complementary regions. However, the folding rate of this mutant did not differ from that of CP240, suggesting that the P15 stem was not involved in the misfolded structure.

A complementarity region between P7-J5.1/7 and P15.1/L15.1 involves nucleotides 276−284 of L15.1 and nucleotides 82−89 of the P7-J5.1/7 region. The level of complementarity increases with the G84A mutant and decreases with the U85C mutant. To test for the possible presence of this non-native helix, two single nucleotide mutations within the L15.1 loop were designed: A281G and U282C. Despite the sequence conservation and the requirement of the L15.1 loop in the tertiary structure of the C-domain, these two possible mutants were not expected to have major effects on the native structure (*14*). In any case, folding during transcription of these two mutants was very similar to that of the wild-type CP240 (Table 1). When double mutants in both L15.1 and the P7-J5.1/7 region were tested for folding during transcription, no additional effects were observed (Table 1), suggesting that these two regions did not interact during transcription.

The above results lead us to conclude that the misfolded structure involves the 5′ strand of P7 and its adjacent residues. Additionally, by the process of elimination, we believe that the L15 region likely is the C-domain structure involved in the misfold. Further support for the involvement of L15 comes from its known interactions with other RNA motifs. It binds to the 3′ CCA of a pre-tRNA substrate (*28*, *29*) and a single-stranded RNA oligonucleotide (*30*). This loop therefore has the capacity to bind single-stranded RNA.

However, a plausible Watson−Crick complementarity leading to the formation of a stable helix between P7-J5.1/7 and L15 could not be identified. As L15 is essential for P RNA catalysis, we were unable to determine the effect of mutations in this region on the folding rate of CP240. Interactions between these two regions may involve a single-stranded RNA binding loop and a region that is transiently single-stranded prior to adopting its native structure.

DISCUSSION

This work aims to elucidate the molecular mechanism of RNA folding during transcription by the *E. coli* RNA polymerase. In a previous work, we found that folding of a large ribozyme is affected significantly by the transcriptional pause at a specific site.

In this study, we show that this effect likely is due to the prevention of a misfolded structure between a region upstream and another region downstream of the pause site at U55. Pausing at this site has no effect on folding when both regions involved in the misfolded structure are synthesized either before or after the pause site. The effect of pausing likely is derived from a structural reorganization of the nascent transcript upstream of the pause site. Pausing has to occur at a suitable location for it to affect RNA folding by resolving or preventing the formation of the misfolded structure.

Two regions in the CP240 RNA are involved in the misfolded structure. The downstream region involved in the misfolded structure is the 5′ strand of P7 and the adjacent single-stranded junction, J5.1/7. Nativelike structures of these regions are necessary for cleaving a pre-tRNA substrate, but they are not needed for catalysis against the selected substrate. Hence, an RNA transcript containing an aberrant structure involving these regions could conceivably cleave the selected substrate but work poorly in cleaving the pre-tRNA substrate.

The upstream region is the P15 hairpin loop identified through the process of elimination and structural considerations. This loop can bind single-stranded RNA, including the conserved C74C75 of a pre-tRNA substrate using two consecutive, conserved G258G259, forming two Watson−Crick base pairs (*28*, *29*). These base pairs significantly improve the catalytic efficiency of the P RNA against the pre-tRNA substrate. On the other hand, binding of the selective substrate used to assay C-domain folding does not result in additional protection of any L15 residues against dimethyl sulfate modification (*31*). Although the L15 loop is part of the binding site for the pre-tRNA substrate, we believe it can be involved in non-native interactions while the remainder of the C-domain is catalytically competent for other substrates. C-domain constructs with significant alterations to L15 have similar catalytic activity against the selected substrate as the wild-type ribozyme (unpublished results and ref *32*). These observations suggest that the involvement of L15 in the misfolded structure should not markedly interfere with ribozyme catalysis for the selected substrate.

There is no obvious extensive base-pair complementarity between L15 and the P7-J5.1/7 region. A short stretch of two A-U and one G-U base pairs may be envisioned between the conserved A252A253U254 in L15 and G84U85U86 in P7-J5.1/7. Formation of this short helix is consistent with the folding results of the G84 and U85 mutants. The G84A mutation should strengthen this helix and slow folding, whereas the U85C mutation should weaken this helix and accelerate folding. Perhaps the particular context of the nascent RNA structure during transcription only allows for the formation of this short, non-native helix and excludes the formation of other non-native helices with better Watson−Crick complementarities. However, these AAU residues in the L15 loop are important in the formation of a specific structure in this loop (*33*). It may be difficult to mutate these AAU residues without affecting the catalytic efficiency of P RNA.

How does NusA-induced transcriptional pausing at U55 prevent the formation of such a misfolded structure? When the P15 motif is initially transcribed, it may exist in one of two possible conformations (Figure 5). The first is an "open" conformation, in which P15 can interact with the single-

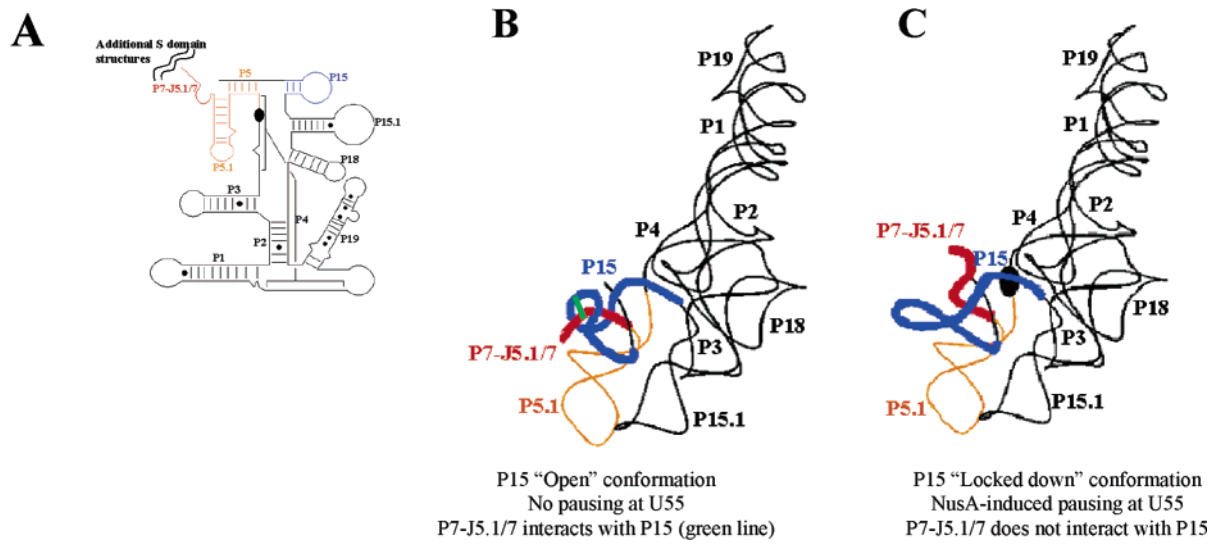

FIGURE 5: (A) Secondary structure of the C-domain highlighting the U55 pause site (●), the 5′ strand of the P7 helix and J5.1/7 (red), and P15 (blue). The nucleotides 5′ to the pause site are shown in black. Nucleotides in the C-domain not yet transcribed when *E. coli* RNA polymerase has reached the pause site are labeled in orange. (B) A C-domain model in the open conformation, adapted from the structural model by Westhof and co-workers (*14*), showing a possible interaction (green) between the P7-J5.1/7 region and the P15 motif in the proposed misfolded intermediate. (C) A C-domain model in the locked down conformation. The proposed misfolded intermediate is avoided through NusA-induced transcriptional pausing at U55.

stranded regions in the vicinity, such as J5.1/7 (Figure 5B). The second is a "locked down" conformation, in which P15 can only interact with the 3′ CCA of the pre-tRNA substrate (Figure 5C). During transcription of CP240 without pausing at U55, P15 exists in the open conformation. In this conformation, L15 can interact with P7-J5.1/7, once it is transcribed, to form the misfolded intermediate. Once this intermediate forms, the RNA must undergo a slow structural rearrangement for folding to proceed to the native state. Upon pausing, P15 adopts the locked down conformation. L15 in this conformation cannot interact with the P7-J5.1/7 region, and the misfolded structure is avoided.

The difference between the open and locked down conformation may simply be the spatial position of the P15 motif, which is adjacent to the pause site at U55. By pausing here, the NusA protein and its associated RNA polymerase complex may force the P15 motif into the locked down conformation. The precise mode of this interaction is unclear. Potentially, the NusA−RNA polymerase complex may accelerate the formation of J5/15 and J15/15.1 structures right at the base of the P15 helix. The lock down of these residues in their respective native structures restrains the spatial position of P15, making it sterically inaccessible to the P7-J5.1/7 region. In CP240, the U55 pause site is transcribed just prior to the J5.1/7-P7 region. This NusA-induced pause site may be in the precise location that leads to lock down of the P15 motif.

Transcription also alters the folding pathways of the group I intron ribozyme and its circular permutants, compared to Mg^{2+}-induced refolding (*2*, *6*, *8*). The act of transcription often increased the ribozymes' rate of folding or, in those RNAs showing biphasic folding kinetics, allowed a greater proportion of the transcripts to fold along the faster pathway. Woodson and co-workers found that folding during transcription allows the group I intron RNA to avoid stable, long-lived intermediates with non-native interactions that form during Mg^{2+}-induced refolding. The increased folding rate during transcription is thought to be derived from avoiding the trapping of the nascent RNA transcript in metastable intermediates or favoring highly branched secondary structures that resemble native conformations.

Transcription can alter the folding pathway of the HDV ribozyme where the transcriptional speed of the RNA polymerase proved critical (*7*). The *E. coli* RNA polymerase, which has a relatively slow transcription speed, allowed the HDV ribozyme to self-cleave before misfold-inducing regions were even synthesized. The T7 RNA polymerase with faster transcription speeds was less effective in preventing these misfolds from occurring.

Our studies of folding during transcription suggest another mode by which transcription can alter the folding pathway. RNA polymerase may play an active role during the search for native conformation by the nascent RNA transcript. Instead of being an indirect participant, the RNA polymerase complex and/or associated transcription factors through site-specific pausing can be located in the right place at the right time to interact with the nascent RNA transcript, thus directly influencing its folding pathway.

The fact that site-specific transcriptional pausing can alter the folding pathway of a large ribozyme may have profound implications on RNA folding in vivo. It was shown previously that RNA polymerases from *E. coli* and *B. subtilis* pause at different locations while transcribing the exact same sequence (*26*). This differential pausing pattern could result in the same RNA transcript folding along different pathways, depending on which polymerase transcribes it. The sequence and pausing sites for particular RNAs may therefore have coevolved to allow for efficient folding when transcribed by its cognate polymerase.

ACKNOWLEDGMENT

We thank the reviewers for insightful comments.

REFERENCES

1. Lewicki, B. T., Margus, T., Remme, J., and Nierhaus, K. H. (1993) Coupling of rRNA transcription and ribosomal assembly in vivo.

Formation of active ribosomal subunits in *Escherichia coli* requires transcription of rRNA genes by host RNA polymerase which cannot be replaced by bacteriophage T7 RNA polymerase, *J. Mol. Biol. 231*, 581−593.
2. Emerick, V. L., and Woodson, S. A. (1993) Self-splicing of the Tetrahymena pre-rRNA is decreased by misfolding during transcription, *Biochemistry 32*, 14062−14067.
3. Lewin, A. S., Thomas, J., Jr., and Tirupati, H. K. (1995) Cotranscriptional splicing of a group I intron is facilitated by the Cbp2 protein, *Mol. Cell. Biol. 15*, 6971−6978.
4. Uptain, S. M., Kane, C. M., and Chamberlin, M. J. (1997) Basic mechanisms of transcript elongation and its regulation, *Annu. Rev. Biochem. 66*, 117−172.
5. Pan, T., Artsimovitch, I., Fang, X., Landick, R., and Sosnick, T. R. (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA, *Proc. Natl. Acad. Sci. U.S.A. 96*, 9545−9550.
6. Woodson, S. A. (2000) Recent insights on RNA folding mechanisms from catalytic RNA, *Cell. Mol. Life Sci. 57*, 796−808.
7. Diegelman-Parente, A., and Bevilacqua, P. C. (2002) A mechanistic framework for co-transcriptional folding of the HDV genomic ribozyme in the presence of downstream sequence, *J. Mol. Biol. 324*, 1−16.
8. Heilman-Miller, S. L., and Woodson, S. A. (2003) Effect of transcription on folding of the Tetrahymena ribozyme, *RNA 9*, 722−733.
9. Landick, R. (1997) RNA polymerase slides home: pause and termination site recognition, *Cell 88*, 741−744.
10. Artsimovitch, I., and Landick, R. (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals, *Proc. Natl. Acad. Sci. U.S.A. 97*, 7090−7095.
11. Frank, D. N., and Pace, N. R. (1998) Ribonuclease P: unity and diversity in a tRNA processing ribozyme, *Annu. Rev. Biochem. 67*, 153−180.
12. Altman, S., and Kirsebom, L. (1999) Ribonuclease P, in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., pp 351−380, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
13. Loria, A., and Pan, T. (1996) Domain structure of the ribozyme from eubacterial ribonuclease P, *RNA 2*, 551−563.
14. Massire, C., Jaeger, L., and Westhof, E. (1998) Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis, *J. Mol. Biol. 279*, 773−793.
15. Schmidt, M. C., and Chamberlin, M. J. (1984) Amplification and isolation of *Escherichia coli* nusA protein and studies of its effects on in vitro RNA chain elongation, *Biochemistry 23*, 197−203.
16. Landick, R., Stewart, J., and Lee, D. N. (1990) Amino acid changes in conserved regions of the beta-subunit of *Escherichia coli* RNA polymerase alter transcription pausing and termination, *Genes Dev. 4*, 1623−1636.
17. Landick, R., Wang, D., and Chan, C. L. (1996) Quantitative analysis of transcriptional pausing by *Escherichia coli* RNA polymerase: his leader pause site as paradigm, *Methods Enzymol. 274*, 334−353.
18. Zarrinkar, P. P., Wang, J., and Williamson, J. R. (1996) Slow folding kinetics of RNase P RNA, *RNA 2*, 564−573.
19. Zarrinkar, P. P., and Williamson, J. R. (1994) Kinetic intermediates in RNA folding, *Science 265*, 918−924.
20. Treiber, D. K., and Williamson, J. R. (2000) Kinetic oligonucleotide hybridization for monitoring kinetic folding of large RNAs, *Methods Enzymol. 317*, 330−353.
21. Chen, J. L., Nolan, J. M., Harris, M. E., and Pace, N. R. (1998) Comparative photocross-linking analysis of the tertiary structures of *Escherichia coli* and *Bacillus subtilis* RNase P RNAs, *EMBO J. 17*, 1515−1525.
22. Krasilnikov, A. S., Yang, X., Pan, T., and Mondragon, A. (2003) Crystal structure of the specificity domain of ribonuclease P, *Nature 421*, 760−764.
23. Pan, T., and Jakacka, M. (1996) Multiple substrate binding sites in the ribozyme from *Bacillus subtilis* RNase P, *EMBO J. 15*, 2249−2255.
24. Pan, T. (2000) Probing RNA structure and function, by circular permutation, *Methods Enzymol. 317*, 313−330.
25. Pan, T., Fang, X., and Sosnick, T. R. (1999) Pathway modulation, circular permutation and rapid RNA folding under kinetic control, *J. Mol. Biol. 286*, 721−731.
26. Artsimovitch, I., Svetlov, V., Anthony, L., Burgess, R. R., and Landick, R. (2000) RNA polymerases from *Bacillus subtilis* and *Escherichia coli* differ in recognition of regulatory signals in vitro, *J. Bacteriol. 182*, 6027−6035.
27. Brown, J. W. (1999) The Ribonuclease P Database, *Nucleic Acids Res. 27*, 314.
28. Kirsebom, L. A., and Svard, S. G. (1994) Base pairing between *Escherichia coli* RNase P RNA and its substrate, *EMBO J. 13*, 4870−4876.
29. Busch, S., Kirsebom, L. A., Notbohm, H., and Hartmann, R. K. (2000) Differential role of the intermolecular base-pairs G292-C(75) and G293-C(74) in the reaction catalyzed by *Escherichia coli* RNase P RNA, *J. Mol. Biol. 299*, 941−951.
30. Willkomm, D. K., Gruegelsiepe, H., Goudinakis, O., Kretschmer-Kazemi Far, R., Bald, R., Erdmann, V. A., and Hartmann, R. K. (2003) Evaluation of bacterial RNase P RNA as a drug target, *ChemBioChem. 4*, 1041−1048.
31. Odell, L., Huang, V., Jakacka, M., and Pan, T. (1998) Interaction of structural modules in substrate binding by the ribozyme from *Bacillus subtilis* RNase P, *Nucleic Acids Res. 26*, 3717−3723.
32. Smith, G. J., Sosnick, T. R., Scherer, N. F., and Pan, T. (2005) Efficient fluorescence labeling of a large RNA through oligonucleotide hybridization, *RNA 11*, 234−239.
33. Glemarec, C., Kufel, J., Foldesi, A., Maltseva, T., Sandstrom, A., Kirsebom, L. A., and Chattopadhyaya, J. (1996) The NMR structure of 31mer RNA domain of *Escherichia coli* RNase P RNA using its nonuniformly deuterium labeled counterpart [the "NMR-window" concept], *Nucleic Acids Res. 24*, 2022−2035.

BI047560L