J|A|C|S ARTICLES

Published on Web 11/09/2010

Transient Tertiary Structure Formation within the Ribosome Exit Port

Edward P. O'Brien,[†] Shang-Te Danny Hsu,[†] John Christodoulou,[‡] Michele Vendruscolo,[†] and Christopher M. Dobson*,[†]

Department of Chemistry, Lensfield Road, University of Cambridge, Cambridge, CB2 1EW, U.K., Institute of Structural Molecular Biology, University College, London and Birkbeck College, University of London, Gower Street, London WC1E 6BT, U.K.

Received July 31, 2010; E-mail: cmd44@cam.ac.uk

Abstract: The exit tunnel of the ribosome is commonly considered to be sufficiently narrow that co-translational folding can begin only when specific segments of nascent chains are fully extruded from the tunnel. Here we show, on the basis of molecular simulations and comparison with experiment, that the long-range contacts essential for initiating protein folding can form within a nascent chain when it reaches the last 20 Å of the exit tunnel. We further show that, in this "exit port", a significant proportion of native and non-native tertiary structure can form without steric overlap with the ribosome itself, and provide a library of structural elements that our simulations predict can form in the exit tunnel and is amenable to experimental testing. Our results show that these elements of folded tertiary structure form only transiently and are at their midpoints of stability at the boundary region between the inside and the outside of the tunnel. These findings provide a framework for interpreting a range of recent experimental studies of ribosome nascent chain complexes and for understanding key aspects of the nature of co-translational folding.

Introduction

Protein synthesis *in vivo* takes place through the action of the ribosome, which in prokaryotes is a 2.5 MDa complex comprised of three large RNA molecules and over 50 different protein molecules. This large biological machinery translates the genetic information contained in a genomic DNA sequence into a nascent polypeptide chain (NC) through the medium of mRNA.[1] The ribosome complex consists of a large and a small subunit; in the case of the *Escherichia coli* ribosome studied here these are known as the 50S and 30S subunits, and the intact ribosome as the 70S particle. Translation is achieved through catalysis of peptide bond formation at the peptidyl transferase center (PTC), which is located on the 50S subunit.[2]

NCs are synthesized from their N-terminal ends, as amino acids attached to tRNAs are sequentially transferred to a growing NC. As new amino acids are incorporated into the NC, the latter is extruded through a tunnel in the 50S subunit and out into the cellular milieu (Figure 1a).[3−5] Since the N-terminus is the first part of the NC to emerge from the tunnel, any folding prior to release of the full-length NC from the ribosome must occur in the absence of direct interactions with the C-terminal group of residues, which are still in the tunnel.[6] This vectorial process of NC synthesis has been suggested to lead to alterations in the details of folding pathways from those observed for proteins *in vitro*,[7] where the initial stages of folding can involve interactions between all the segments of the polypeptide chain.[8,9] Indeed, there is both experimental and computational evidence to suggest that *in vivo* and *in vitro* folding may differ in their details,[10−14] although at least many of the essential principles must be similar.[9] Very little, however, is known experimentally about the structures that the NC populates during translation. Thus, characterizing the structural intermediates that form as NCs are synthesized is a primary goal in studies of the co-translational process.[7,14−19]

The structure of the *E. coli* ribosome, among others, has been obtained at atomic resolution by cryo-electron microscopy (cryo-

[†] University of Cambridge.
[‡] University of London.

(1) Steitz, J. A.; Jakes, K. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 4734–4738.
(2) Shoji, S.; Walker, S. E.; Fredrick, K. *Chem. Biol.* **2009**, *4*, 93–107.
(3) Lu, J. L.; Deutsch, C. *Nat. Struct. Mol. Biol.* **2005**, *12*, 1123–1129.
(4) Voss, N. R.; Gerstein, M.; Steitz, T. A.; Moore, P. B. *J. Mol. Biol.* **2006**, *360*, 893–906.
(5) Wekselman, I.; Davidovich, C.; Agmon, I.; Zimmerman, E.; Rozenberg, H.; Bashan, A.; Berisio, R.; Yonath, A. *J. Pept. Sci.* **2009**, *15*, 122–130.
(6) Fedorov, A. N.; Baldwin, T. O. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1227–1231.
(7) Komar, A. A. *Trends Biochem. Sci.* **2009**, *34*, 16–24.
(8) Guo, Z. Y.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83–102.
(9) Dobson, C. M. *Nature* **2003**, *426*, 884–890.
(10) Frydman, J.; Erdjument-Bromage, H.; Tempst, P.; Hartl, F. U. *Nat. Struct. Biol.* **1999**, *6*, 697–705.
(11) Wang, P. Y.; Klimov, D. K. *Proteins Struct., Funct. Bioinf.* **2008**, *70*, 925–937.
(12) Morrissey, M. P.; Ahmed, Z.; Shakhnovich, E. I. *Polymer* **2004**, *45*, 557–571.
(13) Elcock, A. H. *PLoS Comput. Biol.* **2006**, *2*, e98.
(14) Evans, M. S.; Sander, I. M.; Clark, P. L. *J. Mol. Biol.* **2008**, *383*, 683–692.
(15) Cabrita, L. D.; Dobson, C. M.; Christodoulou, J. *Curr. Opin. Struct. Biol.* **2010**, *20*, 33–45.
(16) Kosolapov, A.; Tu, L. W.; Wang, J.; Deutsch, C. *Neuron* **2004**, *44*, 295–307.
(17) Woolhead, C. A.; McCormick, P. J.; Johnson, A. E. *Cell* **2004**, *116*, 725–736.

 10.1021/ja106530y

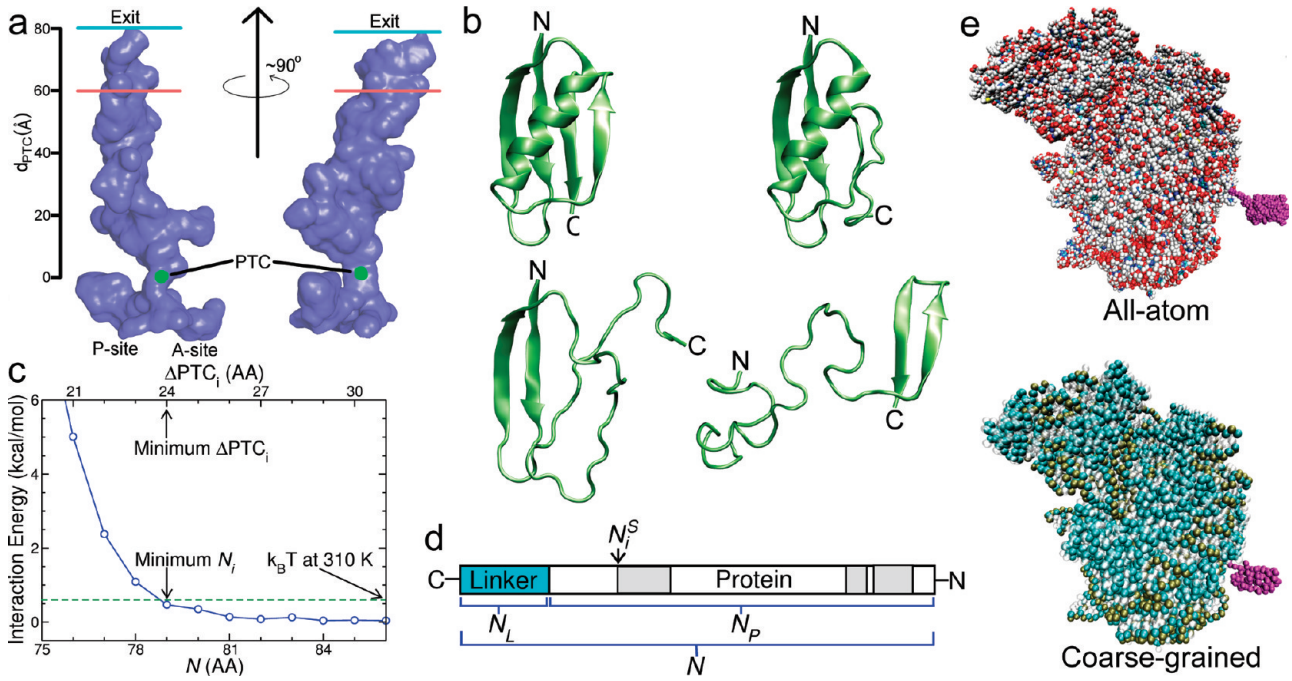

Figure 1. (a) Two views of a space-filling representation of the ribosome exit tunnel (adapted from ref 4). The locations of the peptidyl transferase center (PTC) and A- and P-site tRNA binding sites are indicated. The exit is labeled and indicated by the solid cyan bar. The results in this study indicate that tertiary contacts and native structure can form in the exit port of the tunnel, which starts at the solid red line. (b) Illustration of 4 out of the 23 tertiary structural elements of protein G (see also Figure S2, Supporting Information). In each structure a different folded tertiary structural element (FTE) is restrained to remain folded in the simulation protocol (see Experimental Procedures). From left-to-right and top-to-bottom: FTE *i* corresponds to the full native-state structure, the N-terminal β-hairpin plus helix, the N-terminal β-hairpin, and the C-terminal β-hairpin, respectively. (c) Illustration of how the minimum N_i (or equivalently ΔPTC_i using eqs 1 and 2) is determined. Plotted is the interaction energy between the ribosome and the FTE *i*, corresponding to the fully folded protein G in this case, versus *N*, the number of residues in the nascent chain. As *N* decreases, the interaction energy becomes large and unfavorable due to steric overlap between the FTE *i* and the ribosome. The minimum N_i is the minimum value of *N* at which this interaction energy is equal to k_BT or less (indicated by the dashed green line). (d) Illustration of a nascent polypeptide chain (NC) comprised of an unstructured linker sequence (cyan box), containing N_L residues, attached to the C-terminus of the protein of interest (white box), containing N_P residues. The total number of residues in the NC is $N = N_L + N_P$. Supposing that three secondary structural elements (indicated as gray boxes) along the NC form an FTE, then N_i is the number of the residue closest to the C-terminus that takes part in the FTE where the numbering starts from 1 at the N-terminus. (e) All-atom and coarse-grained representation of protein G-RNC stalled on the 50S subunit of the ribosome with $N_L = 40$ AA. The coarse-grained model captures many of the essential surface features of the all-atom model (Figure S8, Supporting Information).

EM) and X-ray diffraction studies.[20,21] The exit tunnel, which is some 80 Å long and has an average diameter that varies between 10 and 20 Å along its length,[4] appears too narrow to allow tertiary structure formation to occur,[4,5,22,23] suggesting that protein folding involving tertiary structure cannot be initiated until the NC is completely outside the tunnel. In a recently reported cross-linking experiment involving a potassium channel protein bound to the ribosome,[24] however, it was shown that a tertiary structural element comprised of two helices that are in contact in the native state can also form native contacts (i.e., can cross-link) in the NC when one partner in the interaction is only 19 residues (~63 Å) away from the PTC. This experiment suggested that native tertiary contacts can form in the ribosomal exit port, a region corresponding approximately to the last 20 Å of the tunnel.[24] Moreover, it has also been proposed from cryo-EM studies, albeit at low resolution, that some elements of tertiary structure could form inside the exit tunnel.[25] The nature of any such tertiary structural elements in the tunnel is, however, unknown, and it is crucial to address this issue in detail to understand and explain the similarities and differences between the molecular events associated with *in vitro* and *in vivo* folding.

In the present paper, we explore the likely nature of any folding events in the exit tunnel by progressing significantly beyond simplified treatments of the ribosome tunnel and NC structure to more realistic models and address three questions: (1) Can long-range tertiary contacts, which have been shown to be crucial elements in folding *in vitro*,[9,26,27] form in a NC inside the exit tunnel? (2) Can native tertiary structure be accommodated inside the tunnel on steric grounds? [By "native tertiary structure" we refer to the folded tertiary structural elements (FTEs) that comprise a given native state, where FTEs are composed of combinations of secondary structural elements

(18) Hsu, S. T. D.; Fucini, P.; Cabrita, L. D.; Launay, H.; Dobson, C. M.; Christodoulou, J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16516–16521.
(19) Hsu, S. T. D.; Cabrita, L. D.; Fucini, P.; Christodoulou, J.; Dobson, C. M. *J. Am. Chem. Soc.* **2009**, *131*, 8366–8367.
(20) Gabashvili, I. S.; Agrawal, R. K.; Spahn, C. M. T.; Grassucci, R. A.; Svergun, D. I.; Frank, J.; Penczek, P. *Cell* **2000**, *100*, 537–549.
(21) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905–920.
(22) Vila-Sanjujo, A.; Ridgeway, W. K.; Seymaner, V.; Zhang, W.; Santoso, S.; Yu, K.; Cate, J. H. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8682–8687.
(23) Fulle, S.; Gohlke, H. *J. Mol. Biol.* **2009**, *387*, 502–517.
(24) Kosolapov, A.; Deutsch, C. *Nat. Struct. Mol. Biol.* **2009**, *16*, 405–411.
(25) Gilbert, R. J. C.; Fucini, P.; Connell, S.; Fuller, S. D.; Nierhaus, K. H.; Robinson, C. V.; Dobson, C. M.; Stuart, D. I. *Mol. Cell* **2004**, *14*, 57–66.
(26) Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 10869–10873.
(27) Klimov, D. K.; Thirumalai, D. *Proteins Struct., Funct. Genet.* **2001**, *43*, 465–475.

that are in contact (Figure 1b). If an FTE cannot fit in the exit tunnel, then it is energetically unlikely to form within it.] (3) What are the thermodynamic stabilities of FTEs inside the tunnel?

We address these questions using coarse-grained simulations of a representative set of proteins stalled on the ribosome. The validity of the results of these simulations is established by a series of tests, including comparisons with higher resolution and computationally more expensive all-atom simulations, and also with the results reported from cross-linking experiments. Our findings suggest that the common assumption that tertiary structure cannot be accommodated within the exit tunnel, and hence that tertiary protein folding cannot occur until the NC is fully outside the tunnel, considerably oversimplifies the likely events that occur. We find that long-range tertiary contacts can form at least transiently within the exit port in an unstructured NC bound to the ribosome, and we show that the irregular geometry of the exit tunnel and FTEs determines which folded structures can sterically fit in the tunnel and explains why simple geometric arguments fail to predict tertiary folding in the tunnel. We further find in the simulations that many of those FTEs whose steric properties allow them to fit into the exit port do occasionally form within the port at equilibrium, and hence provide an explanation for observations made in cross-linking and other experimental studies.

Results

Determining experimentally whether tertiary structure formation occurs in the exit tunnel, and in which sections of the tunnel it does so, is a formidable task. Cryo-EM and NMR spectroscopic studies are providing the first residue-specific glimpses of the nascent chain within and outside the confines of the exit tunnel.[18,19] Here we use molecular simulations to complement such studies, to describe high-resolution structural details of NC features that could develop during the biosynthesis process, and to probe the behavior of this species on the ribosome.

Defining when a Folded Tertiary Structural Element can Fit inside the Exit Tunnel in the Simulation. The location of tertiary structure formation within the NC during its biosynthesis can be described by the number of residues separating the natively structured portion of the NC (i.e., an FTE) from the PTC, which we denote ΔPTC_i, where i uniquely identifies the FTE under consideration (Figure 1b). More explicitly, $\Delta PTC_i = N - N_i^S$, where N is the number of residues in the NC, and N_i^S is the number of the residue closest to the PTC that is part of an FTE i (Figure 1d). The subscript S in N_i^S emphasizes that we are referring to a natively structured portion of the NC, as opposed to the rest of the NC, which is unstructured (Figure 1b). Experiments on stalled ribosomal NCs (RNCs), i.e., NCs that remain attached to the ribosome, often include a linker peptide, which is not folding competent, that is covalently linked to the C-terminus of the protein of interest,[18,19] allowing ΔPTC_i to be determined when i also corresponds to the entire native structure (Figure 1e). These numbering systems are related to each other and can be easily interconverted:

$$N = N_P + N_L \quad (1)$$

$$\Delta PTC_i = N_P + N_L - N_i^S \quad (2)$$

where N_L is the number of linker residues attached to the fully synthesized protein, and N_P is the number of residues in the protein of interest (Figure 1d).

To determine whether an FTE can fit inside the exit tunnel, we calculated its "minimum ΔPTC_i" (in terms of amino acid number (AA)) and its "minimum $d_{PTC,i}$", the spatial distance between the PTC and FTE in units of Å. These quantities correspond, respectively, to the minimum number of residues and to the minimum distance separating FTE i from the PTC that is required to ensure that the interaction energy between FTE i and the ribosome is less than k_BT (Figure 1c), i.e., to ensure that FTE i has no significant steric clashes with the ribosome. The exit tunnel, being 80 Å in length,[4] requires a minimum of 24 AA (assuming 3.35 Å per AA) to traverse the length of the tunnel.[28] Thus, FTE i is able to fit a portion of its native structure in the exit tunnel if its minimum $\Delta PTC_i \leq 24$ AA and its minimum $d_{PTC,i} \leq 80$ Å. By inserting the minimum ΔPTC_i into eq 2, we also calculate the minimum N (denoted "minimum N_i"), the smallest number of NC residues that must be synthesized for FTE i to avoid steric overlap with the ribosome (Figure 1c).

Simulation Results are Converged and Robust to Changes in Model Resolution and Simulation Protocol. To examine tertiary contact and structure formation in the exit tunnel, we have used coarse-grained molecular simulations of the *E. coli* 50S ribosomal subunit and NC in which amino acids are represented by two interaction sites and nucleotide bases by up to four interaction sites (see Experimental Procedures, Figure 1e, and Figure S1 in the Supporting Information). We use this coarse-grained model, as opposed to an all-atom model, because it allows a higher level of precision through better configurational sampling, and its greater speed enables a larger number of RNCs to be examined. In the coarse-grained simulations we immobilized the ribosome interaction sites so that they did not fluctuate, which allowed longer simulation times to be achieved and further enhanced precision.

In part of the analysis presented below, we used a biased simulation protocol (discussed in the Experimental Procedures) in which a given FTE of the NC was restrained to remain folded while the rest of the NC was unstructured; this protocol allowed the minimum N_i and minimum ΔPTC_i to be determined accurately. Figure 2 establishes the validity and precision of this modeling approach by showing excellent agreement between the calculated minimum N_i for the protein G-RNC from coarse-grained and from all-atom implicit solvent simulations (Figure 2a), from simulations in which the ribosome was allowed a degree of flexibility compared to ones in which it was held in a rigid state (Figure 2b), and from restrained compared to unrestrained NC simulations (Figure 2c). A detailed discussion of this validation is provided as Supporting Information.

In the final portion of the analysis presented below, we calculated the equilibrium stability of FTEs of protein G on an arrested ribosome. In the cell, continuous translation occurs, which represents a nonequilibrium situation. To determine whether the equilibrium results from the arrested ribosome studied here are equivalent to nonequilibrium continuous translation, we ran continuous translation simulations of protein G (Experimental Procedures). We find this nascent chain behaves in the same way in both of these situations in terms of size and structure formation (Figure S3, Supporting Information). The reason for this result is that the time scale for domain folding is much faster than that of amino acid incorporation

(28) Seidelt, B.; Innis, C. A.; Wilson, D. N.; Gartmann, M.; Armache, J.; Villa, E.; Trabuco, L. G.; Becker, T.; Mielke, T.; Schulten, K.; Steitz, T. A.; Beckmann, R. *Science* **2009**, *326*, 1412–1415.

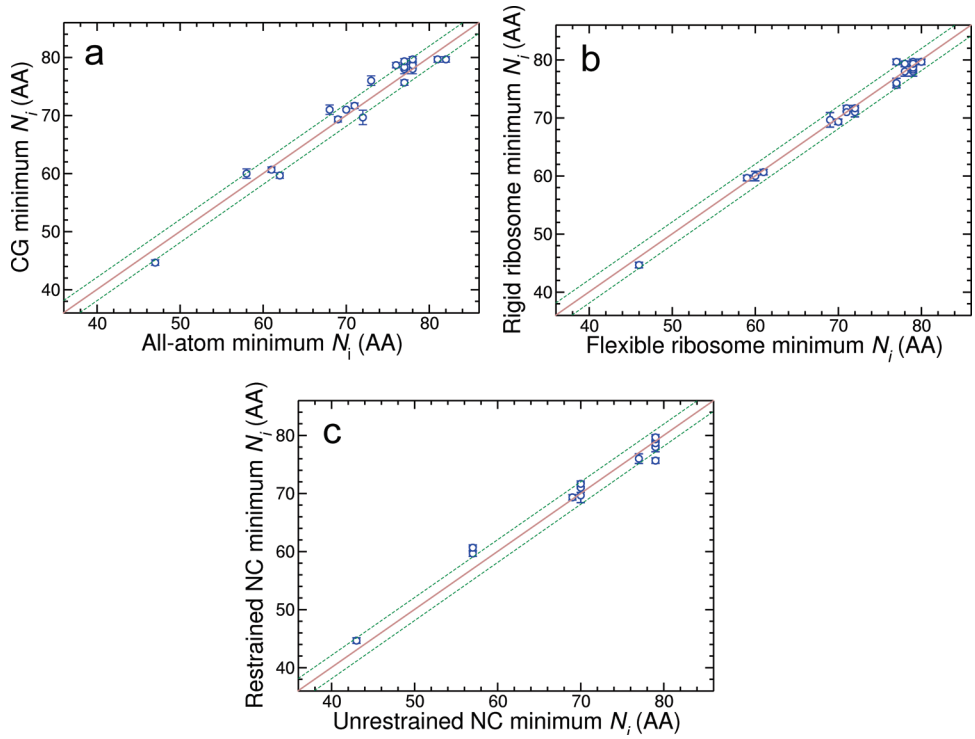

Figure 2. Comparison of the calculated minimum N_i at 310 K for the 23 FTEs of protein G (Table S1 and Figure S2, Supporting Information) from (a) coarse-grained and all-atom simulations in which FTEs of the NC are restrained to remain folded and ribosome interaction sites are held fixed (rigid); (b) flexible ribosome (in which the ribosome interaction sites fluctuate with a variance of 3 Å^2) and rigid ribosome simulations using the coarse-grained model in which FTEs of the NC are restrained to remain folded; and (c) restrained and unrestrained NC simulations using the coarse-grained rigid ribosome model in which ribosome interaction sites are held rigid. The solid gray line corresponds to the identity line, and the dashed green lines are offsets of ±2 AA from this line. The correlation coefficient $R^2 > 0.98$ for graphs (a) through (c). Due to the large computational expense of running the simulation protocol in triplicate, the standard error was only calculated for the coarse-grained rigid ribosome restrained NC simulations (*y*-axis error bars in graphs (a) through (c)). The convergence and robustness of these results to changes in model resolution, ribosome structural fluctuations, and simulation protocol justify the use of the restrained NC coarse-grained rigid ribosome model.

into the growing nascent chain, resulting in a quasi-equilibrium folding process during continuous translation.

Tertiary Contacts in a Disordered NC can Form inside the Exit Port. Tertiary contact formation is essential for the initiation of protein folding from a highly denatured state.[26,27] For this reason, we first explored the position along the tunnel where the formation of tertiary contacts can occur by simulating a 100-residue polyalanine chain (A_{100}), bound to the ribosome, that undergoes a coil-to-globule transition as the temperature is lowered (Figure 3a). The globular state is not folded, but rather it populates a compact heterogeneous structural ensemble without persistent secondary or tertiary structure; such collapse transitions have been observed experimentally in denatured states of proteins under renaturing conditions.[29,30] By studying this A_{100}-RNC at 310 K (Figure 3b), we have determined (1) the probability distribution $P(N_R,d_{PTC})$ of the number of NC residues, N_R, that can fit in the tunnel at a distance d_{PTC} from the PTC (Figure 3c), and (2) the probability distribution $P(\Delta S,d_{PTC})$ of the sequence separation ΔS ($= |j - k|$) between residues j and k that form tertiary contacts (defined as $\Delta S \geq 7$ AA) at a distance d_{PTC} from the PTC (Figure 3d).

The results of this approach reveal that, when $d_{PTC} \leq 65$ Å, the width (i.e., the variance) of these probability distributions as a function of d_{PTC} is small, but when $d_{PTC} > 65$ Å, the distributions are much broader. Therefore, the final 15 Å of the tunnel prior to the exit (Figure 1a) can fit more NC residues inside it and form more long-range tertiary contacts between segments of the NC than can the first 65 Å of the tunnel. Plotting the maximum values of N_R and ΔS observed in the simulations as a function of d_{PTC} (Figure 3e,f), we see that, for $d_{PTC} \leq 60$ Å, no more than ~5 residues can fit in any 3.8 Å window of the tunnel, and no significant tertiary contact formation can occur. For $d_{PTC} > 60$ Å, however, a rapid increase in the number of NC residues that can fit in the tunnel and in the sequence separation between residues that can form tertiary contacts is seen. For example, it is possible to fit seven residues of the NC (not necessarily nearest neighbors in sequence space) in the 3.8 Å window centered at $d_{PTC} = 67$ Å (Figure 3e). More remarkable, however, is that, within this section of the tunnel, the nascent chain can form tertiary contacts that are separated by $\Delta S = 28$ residues along the NC sequence (Figure 3f). This number increases greatly to $\Delta S = 68$ AA a little further along the tunnel at $d_{PTC} = 70$ Å (Figure 3f), corresponding to a structure in which the N-terminus of the NC threads its way back through the exit port to interact with a residue significantly closer to the C-terminus of the NC (Figure 3g). Indeed, ~15 residues can fit in the tunnel in the 3.8 Å window centered at $d_{PTC} = 77$ Å and the maximum $\Delta S \approx 72$ AA. These results clearly demonstrate that tertiary contacts between NC residues can, in principle, form inside the exit port starting a mere 60 Å from the PTC.

Elements of Native and Non-native Tertiary Structure can Fit inside the Exit Port. Although tertiary contacts can form in the exit tunnel for a structurally disordered NC, this does not necessarily mean ordered tertiary structure can also form.

(29) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359–374.
(30) Ziv, G.; Thirumalai, D.; Haran, G. *Phys. Chem. Chem. Phys.* **2009**, *11*, 83–93.

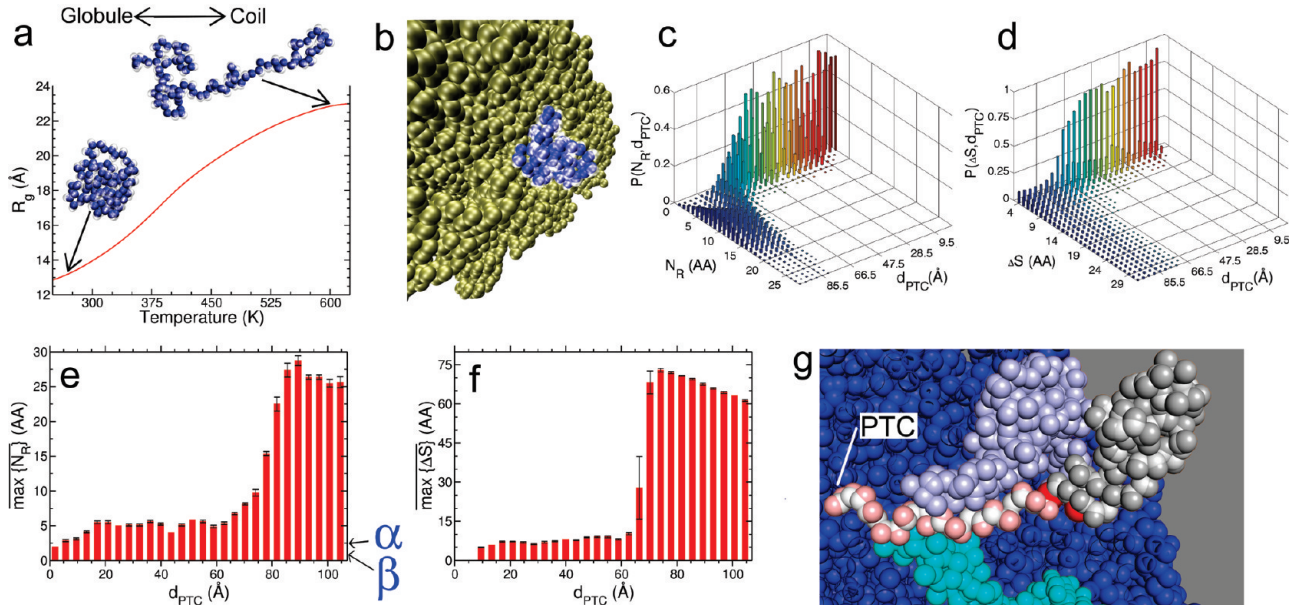

Figure 3. Behavior of the disordered polyalanine NC (A_{100}). (a) In bulk solution, A_{100} undergoes a coil (large R_g)-to-globule (small R_g) transition, as characterized by its R_g versus temperature profile computed from unrestrained coarse-grained simulations. In the structures shown here, the C_α and side-chain interaction sites are represented as blue and white spheres, respectively. (b) A structure from the simulations of A_{100} bound to the ribosome (tan spheres). (c) The probability distribution $P(N_R,d_{PTC})$, computed from coarse-grained A_{100}-RNC simulations at 310 K, of the number of A_{100} residues N_R a distance $d_{PTC} \pm \delta_{PTC}$ from the PTC, where $\delta_{PTC} = 1.9$ Å. This value of δ_{PTC} was chosen because the $C_\alpha - C_\alpha$ distance between neighboring residues is 3.8 Å. (d) The probability distribution $P(\Delta S,d_{PTC})$ represents the probability that if two NC residues (with index numbers j and k) are in contact, they will be separated by ΔS (= $|j - k|$) residues and at a distance $d_{PTC} \pm \delta_{PTC}$ from the PTC at 310 K. (e) The maximum number of A_{100} residues found at a distance $d_{PTC} \pm \delta_{PTC}$ from the PTC, averaged over eight independent coarse-grained simulations of the A_{100}-RNC at 310 K. The standard errors about the mean are shown as black bars. The arrows labeled α and β denote the N_R value expected if the NC is in an unbent α-helix or an extended β-strand conformation, respectively. (f) As (e) except that the maximum sequence separation ΔS between residues in contact at a distance of $d_{PTC} \pm \delta_{PTC}$ from the PTC is plotted. The monotonic decrease in $max\{\Delta S\}$ for $d_{PTC} \geq 75$ Å is due to the finite length of the A_{100} NC. (g) Structure from the simulation of A_{100} bound to the ribosome (blue spheres). Residues 1 (the N-terminus) and 76, shown in red, are in contact a mere 66 Å from the PTC. This tertiary contact is separated by $\Delta S = 75$ residues along the NC (shown in gray). The rest of the NC (residues 76−100) is shown as white and pink spheres corresponding to the C_α and the side-chain interaction sites, respectively.

Therefore, we next examined whether any elements of native tertiary structure can form in the exit tunnel, given the steric constraints its dimensions appear to impose. We approached this problem by determining whether FTEs can fit sterically in the tunnel. An FTE is defined to have the following properties: (1) it is comprised of β-strands and/or α-helices, and these secondary structural elements (SSEs) are arranged as in the native structure (i.e., they are folded), and (2) it forms one spatially contiguous structural unit, with each folded SSE in contact with at least one other folded SSE (see Experimental Procedures). Consider, for example, the native structure of protein G (Figure 1b). It has five SSEs corresponding to four β-strands and one α-helix that yield a total of 23 FTEs including its native state (Figure 1b and Figure S2 and Table S1 in the Supporting Information), each of which represents a possible structural element in a co-translational intermediate that could form during the biosynthesis of this protein.

To examine whether any of these FTEs can fit inside the tunnel, we used a simulation protocol (Experimental Procedures) in which a particular FTE, i, of the native structure of the protein is restrained to remain folded during the simulation while the rest of the NC, which is bound to the ribosome, is unfolded. By starting with an initial configuration in which the FTE is outside the exit tunnel (using a linker with $N_L = 50$ AA, Figure 1e) and slowly pulling the NC toward the PTC with concomitant residue deletions at the C-terminus, we were able to identify the minimum N_i value below which it would be energetically unfavorable for FTE i to stay folded due to steric clashes with the ribosome (Figure 1c). This "reverse synthesis" process should yield results equivalent to those of the "forward synthesis" process (i.e., when the NC is growing), provided the NC achieves quasi-equilibrium during forward synthesis. We tested this hypothesis by performing simulations of stalled and continuous syntheses (see Experimental Procedures) for the protein G-RNC and find they indeed yield similar results (Figure S3, Supporting Information). Thus, at least for small proteins, this protocol of reverse synthesis yields results that should be directly applicable to understanding continuous forward synthesis.

Using this simulation protocol, we have examined nine proteins, whose native structures represent a wide variety of different folds and contain a total of 795 FTEs, where each is bound to the ribosome (Figure 4a and Table 1). Of these FTEs, the numbers that are all-α, mixed α/β, and all-β are respectively 198, 162, and 435. Out of the 435 all-β FTEs, 300 come from GFP alone. We find the histogram of minimum ΔPTC_i values (Figure 4b) calculated from the simulations has minimum and most probable values of 19 and 24 AA, respectively. The simulations, therefore, indicate that native tertiary structure can form in portions of the NC that are separated by a mere 19 AA from the PTC, rather than the 30 AA necessary for the NC to traverse the tunnel if no tertiary structural elements are formed,[4,5,22,23] which places such structure well within the exit port (Figure 5d).

We next calculated the distribution of minimum $d_{PTC,i}$ values (Figure 4c) to determine where exactly along the tunnel tertiary structure can fit. In agreement with the analysis of tertiary contact formation described above (Figure 3e,f), we find that native tertiary structure can be accommodated in the tunnel just 61 Å from the PTC, with the most probable location to fit FTEs at d_{PTC} =80 Å; the latter corresponds to the exit port interface

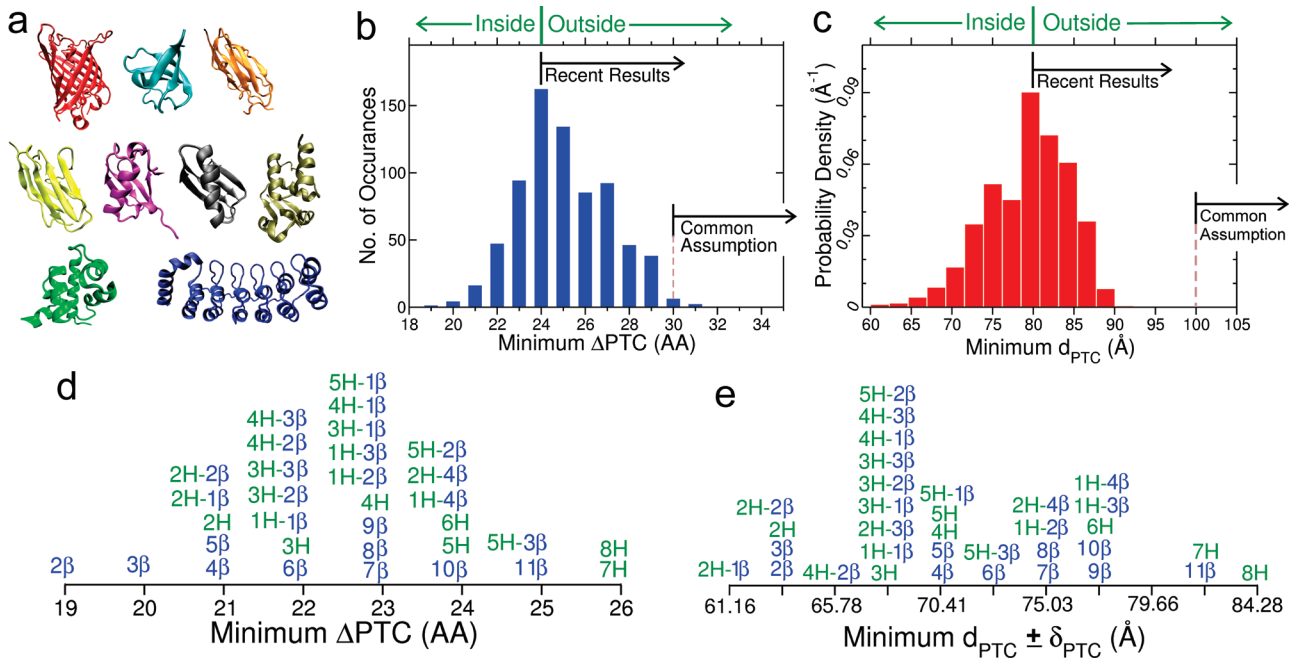

Figure 4. Native tertiary structure in the ribosome exit tunnel. (a) Native structures of the nine proteins examined: from left-to-right and top-to-bottom, green florescence protein (GFP, red), SH3 (cyan), domain 5 of filamen (orange), immunoglobulin domain 27 (I27, yellow), ubiquitin (magenta), protein G (gray), Kv1.2 (tan), λ_{6-85} repressor (green), and notch ankyrin 1−7 (blue), respectively. (b) Histogram of the minimum ΔPTC_i values and (c) the probability density of the minimum d_{PTC} values for the 795 FTEs calculated from the restrained coarse-grained simulations of the nine proteins shown in (a). In (b) the number of occurrences (denoted "No. of Occurrences") indicates the number of FTEs that have a certain minimum ΔPTC_i value. "Common Assumption" denotes the previously reported estimates that that tunnel is anywhere from 100 to 121 Å in length and a minimum of 30 AA is necessary to traverse the tunnel and allow tertiary folding to occur.[5,21,23] "Recent Results" denotes a recent structural analysis[4] indicating the tunnel is actually 80 Å in length and that the midpoint of tertiary structure cross-linking occurs near 24 AA for an α-hairpin.[24] "Inside" and "Outside" indicate the nominal boundary at d_{PTC} = 80 Å that separates the inside and outside of the exit tunnel. In this study we define FTE *i* to be inside the exit tunnel if its minimum $\Delta PTC_i \leq 24$ AA and minimum $d_{PTC,i} \leq 80$ Å. (d) FTE "inventory" of the first occurrence of an FTE containing exactly *X* α-helices (H) and *Y* β-strands (β) as a function of minimum ΔPTC_i. For example, there are many FTEs within the nine native structures shown in (a) that have exactly three β-strands. However, the earliest occurrence of an FTE containing exactly three β-strands fitting in the tunnel occurs at a minimum $\Delta PTC_i = 20$ AA (denoted 3β). FTEs that are comprised of both α-helices and β-strands are denoted with a dash between the helix count and β-strand count of that FTE. (e) Same as (d) except it is plotted as a function of the minimum d_{PTC} and placed in bins of width $2\delta_{PTC}$, where $\delta_{PTC} = 1.16$ Å.

Table 1. Number of FTEs for the Proteins Examined in this Study

protein	PDB ID	structure	N_P[a]	N_{FTE}[b]	N_{fit}[c]	ref[d]
I27	1tit	β	89	17	16	51
protein G	1gb1	α/β	56	23	12	52
λ_{6-85} repressor	1lmb	α	80	23	11	53
SH3	1shg	β	57	25	14	54
ubiquitin	1ubq	α/β	76	39	17	55
filamen[e]	1qfh	β	105	73	54	56
notch ankyrin$_{1-7}$	1ot8	α	209	125[f]	9	57
Kv1.2	1qdv	α/β	99	170	73	58
GFP	1w7s	β	230	300[g]	68	59

[a] Number of residues in the fully synthesized protein. [b] Number of unique FTEs that comprise the crystal structure. [c] Number of FTEs that sterically fit inside the exit port. [d] Reference to crystal structure used. [e] Only domain 5 was used in the simulations (residues 646−750). [f] Only domains 5−7 (residues 121−237) were included in the FTE count since the *N* values we study affect only these residues. [g] The helices of GFP were not included in the FTE count due to the combinatorial explosion that occurs when they are included.

separating the inside and outside of the tunnel (Figure 1a). Figure 5d displays one of the restrained simulation structures found at a minimum ΔPTC_i value of 19 AA (d_{PTC} = 63.5 Å). These results show that, on average, 35% (= $N_{fit}/N_{FTE} \times 100\%$, Table 1) of FTEs can fit a portion of their native tertiary structure inside the exit port of the tunnel and therefore have the opportunity, at least from a steric perspective, to begin to fold. For protein G, 12 out of its 23 FTEs can fit into the exit port; the statistics for other RNCs are listed in Table 1. Moreover, by creating an FTE "inventory" listing the secondary structural elements that make up an FTE as a function of ΔPTC_i and d_{PTC} (Figure 4d,e), we find that the first occurrence in which two β-strands form folded tertiary structure occurs at $\Delta PTC_i = 19$ AA (d_{PTC} = 63.5 Å and Figure 5d). Three β-strands can form tertiary structure starting at $\Delta PTC_i = 20$ AA (d_{PTC} = 63.5 Å), and two helices can assemble tertiary structure starting at $\Delta PTC_i = 21$ AA (d_{PTC} = 61.2 Å). Closer to the exit opening, greater combinations of helices and β-strands can be found (Figure 4d,e). Thus, perhaps quite unexpectedly, our results indicate that FTEs containing a large number of SSEs can, from a steric point of view, fit a portion of their folded structure inside the exit port.

To determine whether non-native structures can also fit inside the ribosome exit tunnel, we examined an IgG binding domain protein that has been shown to adopt two different native folds upon the mutation of a single residue.[31] The native structure of one sequence, referred to as G_A98 in ref 31, is a three-helix bundle, while the other sequence, G_B98, forms a native structure comprised of a β-sheet and α-helix. Given the single-residue difference between the primary structures of these two proteins, we expect that the native structure of G_A98 represents a possible non-native structure populated by G_B98 and vice versa (Figure 6). Using the same simulation protocol as above, we examined whether these non-native structures can fit in the exit tunnel. We find that approximately one-half of G_B98's non-native tertiary structure can sterically fit in the exit tunnel, while for

(31) Alexander, P. A.; He, Y.; Chen, Y.; Orban, J.; Bryan, P. N. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 21149–21154.

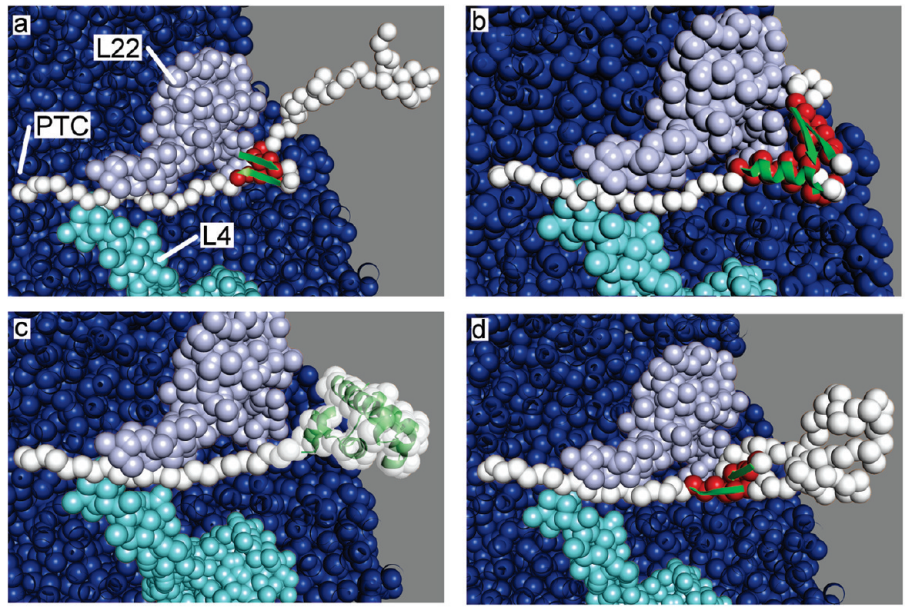

Figure 5. Folded FTEs sampled during the unrestrained RNC simulations, where NC folding and unfolding occur at equilibrium. The ribosome is shown as blue spheres, except for the ribosomal proteins L4 and L22 are shown in cyan and light blue colors, respectively. All NC structures shown have a C_α root-mean-squared-deviation of less than 3 Å from their respective crystal structure. For the sake of clarity only the C_α interaction sites of the NC are shown. (a) A FTE of protein G containing natively structured C-terminal β-strands at $\Delta PTC_i = 22$ AA. The corresponding β-strands from the crystal structure are superimposed on the FTE as a green secondary structure representation. (b) A FTE of protein G containing N-terminal β-strands and an α-helix (red spheres) at $\Delta PTC_i = 21$ AA. The corresponding FTE from the crystal structure is superimposed (green). (c) Fully folded λ_{6-85} at $\Delta PTC_i = 22$ AA; the crystal structure is shown in green. (d) From the restrained FTE simulations we find ubiquitin's β-strands 1 and 5 (residues 2−6 and 66−71, respectively, and colored red) can on steric grounds fit deeply in the tunnel ($\Delta PTC_i = 19$ AA) in their folded conformation. The corresponding FTE from the crystal structure is superimposed (green). We note that due to the issue of perspective it appears in these figures the L22 protein is in contact with the NC close to the exit of the tunnel when in fact it is not.

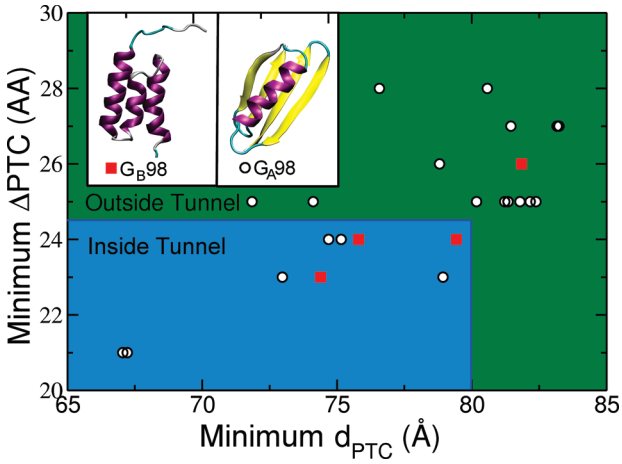

Figure 6. Location of non-native FTEs as a function of minimum ΔPTC_i and minimum d_{PTC} for the protein sequences G_B98 (red squares) and G_A98 (open circles). Sequences G_B98 and G_B98 differ by one residue.[31] The non-native structure for G_B98 is a three-helix bundle (shown in inset), while the non-native structure of G_A98 consists of a β-sheet platform with an α-helix on top. The boundary between the inside and outside of the exit tunnel, as defined in this study, is indicated by the solid blue line. For G_A98, 6 out of 21 non-native tertiary structures can fit a portion of their structure inside the tunnel, while for G_B98, 3 out of 4 can.

G_A98 three-fourths of its non-native tertiary structure can sterically fit (Figure 6). These findings indicate that *both* native and non-native structure can fit in the exit tunnel.

Simple Shape, Size, and Volume Arguments Lack Sufficient Resolution to Explain where FTEs can Fit along the Tunnel. In the light of our findings, we have explored the question of whether the ability of a given FTE to fit sterically in a region of the exit tunnel is related to its geometric properties of shape, size, or volume. It is important to address this question because simple geometric considerations of the nascent chain and tunnel can lead to the conclusion that tertiary structure formation in the tunnel cannot occur. We tested this issue by plotting, as a function of minimum $d_{PTC,i}$, the shape parameters of the FTEs in terms of their spherical and cylindrical character using the metrics of asphericity and acylindricity, respectively,[32] the characteristic size of each FTE in terms of its radius of gyration (R_g), and the eigenvalues (λ_0, λ_1, λ_2) of its R_g tensor, as well as the volume of each FTE.

The results of this analysis (Figure 7) show that the Pearson correlation coefficient is lower than 0.41 for any of these plots, implying that simple arguments based on the overall shape, size, and volume properties of FTEs do not explain adequately the ability of the latter to fit in various locations along the tunnel. Examining the simulated structures in Figure 5, it is clear that the weakness of this correlation arises from the irregular geometries and shapes of the tunnel and of the FTEs, factors that allow larger objects to fit in the tunnel than one might expect from consideration of the interaction of regular geometric objects, such as a sphere (the NC) packing in a cylinder (the ribosome exit tunnel). Therefore, higher resolution details of the ribosome tunnel and FTEs, as probed in the simulations in this paper, are necessary to understand the ability of folded portions of NCs to be accommodated in the exit tunnel.

FTEs are Thermodynamically Unstable inside the Exit Port and are at their Midpoints of Stability at the Exit Port Interface. The structural insights described in the previous two sections are based on restrained simulations in which equilibrium NC folding and unfolding were not permitted. While this approach is useful in determining precisely which FTEs can fit in the tunnel (Figure 2), it does not provide information on the

(32) Theodorou, D. N.; Suter, U. W. *Macromolecules* **1985**, *18*, 1206–1214.

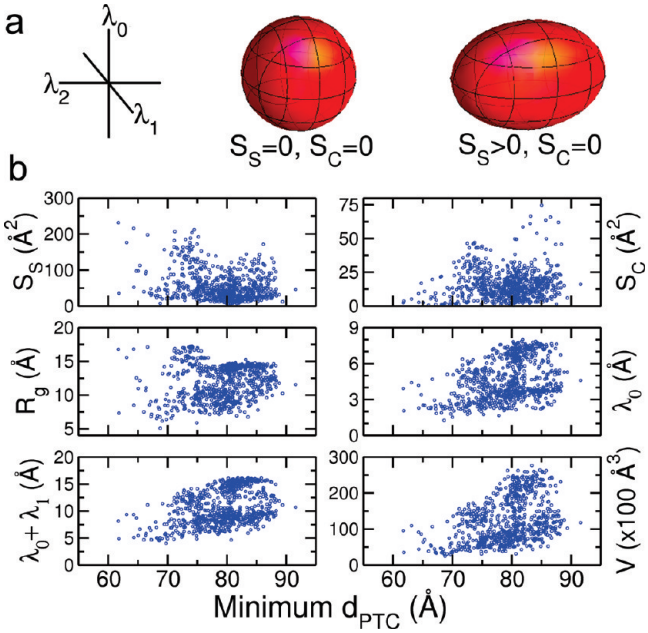

Figure 7. (a) Example of the use of the radius of gyration tensor's eigenvalues, denoted λ, to determine the shape parameters asphericity ($S_S = \lambda_2^2 - 0.5(\lambda_0^2 + \lambda_1^2)$) and acylindricity ($S_C = \lambda_0^2 - \lambda_1^2$). The eigenvalues have the property $\lambda_0 \leq \lambda_1 \leq \lambda_2$ (see Experimental Procedures). $S_S = 0$ and $S_C = 0$ when the FTE is spherical ($\lambda_0 = \lambda_1 = \lambda_2$) or cylindrical ($\lambda_0 = \lambda_1$) in shape. (b) Shape, size, and volume characteristics of the 795 FTEs from the nine proteins listed in main text Table 1 as a function of their minimum d_{PTC}. The radius of gyration, $R_g = (\lambda_0^2 + \lambda_1^2 + \lambda_2^2)^{0.5}$, is plotted along with the two smallest eigenvalues, λ_0 and λ_1. The volumes (V) of the folded FTEs are also plotted. The volume corresponds to the volume of the FTE interaction sites only; excluded volume effects from the ribosome or other NC interaction sites are not included in the calculation. A probe radius of 1.4 Å, corresponding to the radius of a water molecule, was used in the volume calculation.

thermodynamic stability of the FTEs as a function of ΔPTC_i and hence on the probability of their being present under equilibrium conditions. To examine this issue, we carried out unrestrained (where the NC is free to fluctuate structurally) replica exchange simulations of protein G-RNC and λ_{6-85}-RNC, in each of which hundreds of folding and unfolding transitions are observed for $N_L \geq 27$ (Figure S4, Supporting Information).

The nascent chain model used (see Experimental Procedures) has, notably, been shown to produce realistic folding/unfolding thermodynamics in terms of calculated heat capacities, osmolyte m-values, and cooperativity metrics such as the van't Hoff divided by the calorimetric enthalpy.[33] For example, for protein G in bulk, we calculate this ratio of enthalpies to equal 1.05, a value in excellent agreement with the experimentally measured value of 1.04.[34] In addition, the calculated bulk native state stabilities of protein G and λ_{6-85} are, respectively, −4.8 and −2.7 kcal/mol at 310 K, in good agreement with the experimentally measured values of −4.8 and −4.3 kcal/mol, respectively.[35,36] Thus, this model produces realistic values for the thermodynamics and free energies of stability of these proteins where experimental data exist for comparison.

Figure S5a,b (Supporting Information) shows the probability $P_i(\Delta PTC_i)$ that FTE i of protein G and λ_{6-85} are folded as a function of ΔPTC_i. Converting these probabilities into free energies of stability $\Delta G_i(\Delta PTC_i)$ $(= -k_B T \ln(P_i(\Delta PTC_i)/(1 - P_i(\Delta PTC_i)))$, Figure S5c shows that none of the FTEs of protein G is thermodynamically stable (i.e., $\Delta G_i < 0$) until $\Delta PTC_i \geq 28$ AA, which is at the interface between the inside and outside of the tunnel (Figure S6, Supporting Information). This means that no FTE that folds inside the tunnel is thermodynamically stable relative to its unfolded state and reveals that these folded structures will therefore form only transiently in the exit port. This result for the small proteins studied here is consistent with the established conclusion from *in vitro* studies that the native structures of globular proteins exhibit highly cooperative folding transitions and are generally unstable when their sequences are truncated.[37] Therefore, when a significant fraction of native contacts is not present, the folded structure becomes unstable, resembling the situation in which a portion of an NC is sequestered in the exit tunnel.

While tertiary folding in the exit port may be thermodynamically disfavored, it will still occasionally occur and may be significant. Thus, for example, in the case of λ_{6-85} at $\Delta PTC_i =$ 23 AA, 12 different FTEs of this protein are observed to fold in the exit port during the simulation (Figure S5, Supporting Information). Helices 1 (residues 4−21) and 5 (residues 73−79) of λ_{6-85} form an FTE for 0.2% of the simulation time ($\Delta G_i(\Delta PTC_i = 23$ AA$) = 3.7$ kcal/mol); indeed, the complete native structure is found to be present in the exit tunnel for 0.1% of the simulation time ($\Delta G_i(\Delta PTC_i = 23$ AA$) = 4.2$ kcal/mol). The midpoint of λ_{6-85} folding, that is, when $\Delta G_i = 0$ kcal/mol, occurs at $\Delta PTC_i = 26$ AA (Figure S5b). This means that the native state of λ_{6-85} is significantly populated ($P_i(\Delta PTC_i = 26$AA$)) \approx 0.5$) at the interface between the inside and outside of the exit port and is occasionally populated at smaller values of ΔPTC_i (Figure 5c); indeed, the simulated structures at $\Delta PTC_i = 26$ AA show that the natively folded λ_{6-85} is partially buried inside the exit port (Figure S6c,d, Supporting Information).

For protein G at $\Delta PTC_i = 23$ AA, five different FTEs fold in the exit port during the simulations (Table S1, Supporting Information). For example, β-strand 2 (residues 14−19) forms an FTE with the α-helix that is folded for 0.02% of the simulation time ($\Delta G_i = 5.2$ kcal/mol). Complete native-state structure formation is occasionally observed at $\Delta PTC_i = 24$ AA (Table S1), with a stability of 10.8 kcal/mol, and the midpoint of folding for this FTE occurs at $\Delta PTC_i = 28$ AA (Figure S5a, Supporting Information). On the basis of an analysis of structures from the simulations (Figure S6a,b, Supporting Information), this midpoint can be interpreted as being at the interface of the exit opening with a portion of the folded native state inside the exit port.

Discussion and Conclusions

Tertiary Structure can Occur at least Transiently inside the Ribosome Exit Port. We have provided three distinct lines of evidence that demonstrate that tertiary native structure formation and hence protein folding can start in the exit port, the last 20 Å of the 80 Å long ribosome exit tunnel (Figure 1a). First, by examining an RNC containing the disordered globular polypeptide chain A_{100} as a NC, we find that the exit port can accommodate a large number of NC residues (Figure 3e) and that long-range tertiary contacts, in some cases separated

(33) O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *36*, 13403–13408.
(34) Thoms, S.; Max, K. E. A.; Wunderlich, W.; Jacso, T.; Lilie, H.; Reif, B.; Heinemann, U.; Schmid, F. X.; Franz, X. *J. Mol. Biol.* **2009**, *391*, 918–932.
(35) Sancho, D. D.; Doshi, U.; Munoz, V. *J. Am. Chem. Soc.* **2009**, *131*, 2074–2075.
(36) McCallister, E. L.; Alm, E.; Baker, D. *Nat. Struct. Biol.* **2000**, *7*, 669–673.
(37) Neira, J. L.; Fersht, A. R. *J. Mol. Biol.* **1999**, *285*, 1309–1333.

by more than 70 residues, can in principle form within the exit port (Figure 3f,g); the ability to make such tertiary contacts is a prerequisite for the initiation of protein folding.[9,26,27] Second, by using restrained simulations in which a given FTE remains folded while the rest of the NC is unfolded on the RNC, we show that, for nine different proteins, some 35% of their folded tertiary structural elements can, on average, fit a portion of their structure inside the tunnel, at least on steric grounds. Therefore, these tertiary structures have a chance of folding in the tunnel starting at $\Delta PTC_i \geq 19$ AA and $d_{PTC} \geq 61$ Å (Figure 4b−e). Applying this same methodology to possible non-native structures of the proteins[31] G_A98 and G_B98, we demonstrated that non-native structure can also fit in the exit tunnel (Figure 6). Third, by using unrestrained simulations of protein G-RNC and λ_{6-85}-RNC that allow reversible folding and unfolding to occur, we find that, while the FTEs are not thermodynamically stable inside the exit port (Figure S5, Supporting Information), many elements of natively folded tertiary structure can be sampled during the simulations (Figure 5). At the interface of the exit port, the protein G-RNC and λ_{6-85}-RNC are folded in their native structure half the time in the simulations (Figure S5). Folding at this interface corresponds structurally to a portion of the folded NC being inside and a portion of it being outside the exit port (Figure 5c). This last finding is consistent with studies using NMR[38] and cross-linking[24] of stalled RNCs that observed the presence of elements of native structure near the exit port interface in two different nascent chains. These three lines of evidence from simulations strongly suggest that a significant level of transient tertiary protein folding can occur in the last quarter of the exit tunnel. Moreover, the general conclusions are consistent with those of a simulation study of a coarse-grained chymotrypsin inhibitor, (CI2)-RNC, where it was also found that tertiary structure can form in the tunnel and that it is unstable during synthesis.[13] This consistency is particularly relevant, as the assumptions and procedures differ and vary significantly, providing considerable confidence in the overall conclusions.

The finding that partial folding of a NC can occur inside the exit port has added significance because it suggests that protein targeting and chaperone recruitment could be initiated through recognition of these tertiary structures by chaperone molecules. In addition to the potential response of the tunnel to folding,[17] it is possible that partially folded NCs might contain clusters of exposed hydrophobic residues, or patterns of residues, that are recognized and bound by chaperones;[39] thus, for example, trigger factor (TF) has been shown to form a cradle that covers the ribosome exit tunnel opening and interacts with nascent chains.[40] In addition, the binding of TF may enhance the stability of folded NC tertiary structures in the exit port through an entropic mechanism.[41] Finally, and in agreement with previous conclusions,[24] this work suggests that, when chaperones such as TF[42] bind to the ribosome, the effective volume accessible to the NC inside the ribosome−TF complex is larger than that expected if folding could only occur outside the tunnel (Figure 1a). This conclusion is relevant to models of the mode of action of TF, where long NCs have been speculated to reptate through a cleft in the TF structure before folding begins, as a consequence of the limited volume that has been assumed to exist in the TF−ribosome cradle.[40]

Simulation Results Provide a Framework for Interpreting Experimental Studies. Recently, maleimide cross-linking experiments that can separate and detect the approximate populations of folded and unfolded species[43] have been reported for the Kv1.3 potassium protein stalled on the ribosome at different lengths of elongation.[24] The authors conclude that non-native tertiary contacts within the NC can form inside the exit port, that an α-hairpin can form native tertiary contacts starting at $\Delta PTC_i = 19$ AA, which is inside the exit port, and that this α-hairpin has a midpoint of stability located at $\Delta PTC_i = 24$ AA. Our finding of significant, albeit transient, tertiary contact formation for a structurally disordered NC inside the exit port (Figure 3d,f) is broadly consistent with this experimental finding of contact formation within the unfolded ensemble of a protein bound to the ribosome.

To gain additional insight into the structural details giving rise to these findings, we carried out further simulations of a protein of direct relevance to this system. As no crystal structure has been reported for the Kv1.3 protein, we carried out the simulations described in this paper using the Kv1.2 protein, which shares 94% sequence identity with Kv1.3 and for which a high-resolution structure is available. The very high sequence similarity is such that we can be confident in comparing the simulation data of Kv1.2 with the experimental data of Kv1.3. In our restrained simulations of the Kv1.2 protein, we find that the α-hairpin described above can sterically fit into the tunnel of the ribosome crystal structure with a minimum $\Delta PTC_i = 21$ AA. Furthermore, in our simulations, the midpoints of native-state stability for protein G and λ_{6-85} are between 26 and 28 AA, which is in agreement with the cross-linking transition midpoint of 24 AA found experimentally for the α-hairpin.

Our simulations provide a framework for understanding the available experimental data that indicate that native tertiary structure can form at least transiently inside the exit port at approximately the same position as that observed experimentally (21 vs 19 AA). Moreover, the simulations offer two main insights into these cross-linking experiments.[24] First, at $\Delta PTC_i = 24$ AA, which is at the inflection point in the probability of cross-linking, the α-hairpin studied by experiment may be at the interface between the exit port and the external environment rather than completely inside the tunnel. Second, our simulations suggest that cross-linking events in the exit port are likely to be observable between cysteine residues that are separated by more than 70 residues along the NC, separation lengths that have not yet been tested experimentally[24] but would be a valuable means of exploring further the accuracy of the simulations.

The numerous crystal structures of the ribosome that have been determined thus far indicate that the exit tunnel does not undergo large-scale structural rearrangements upon binding elongation factor, trigger factor, or tRNA molecules in various states of translation.[44−46] The tunnel appears, from these data at least, to exhibit only local structural fluctuations with amplitudes on the order of 1 Å².[4] Our simulation results clearly

(38) Cabrita, L. D.; Hsu, S. T. D.; Launay, H.; Dobson, C. M.; Christodoulou, J. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 22239–22244.
(39) Stan, G.; Brooks, B. R.; Lorimer, G. H.; Thirumalai, D. *Protein Sci.* **2005**, *14*, 193–201.
(40) Merz, F.; Boehringer, D.; Schaffizel, C.; Preissler, S.; Hoffmann, A.; Maier, T.; Rutkowska, A.; Lozza, J.; Ban, N.; Bukau, B.; Deuerling, E. *EMBO J.* **2008**, *27*, 1622–1632.
(41) Klimov, D. K.; Newfield, D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *12*, 8019–8024.
(42) Frydman, J. *Annu. Rev. Biochem.* **2001**, *70*, 603–647.
(43) Kosolapov, A.; Deutsch, C. *J. Biol. Chem.* **2003**, *278*, 4305–4313.
(44) Ishida, H.; Hawardy, S. *Biophys. J.* **2008**, *95*, 5962–5973.
(45) Voorhees, R. M.; Weixlbaumer, A.; Loakes, D.; Kelley, A. C.; Ramakrishnan, V. *Nat. Struct. Mol. Biol.* **2009**, *16*, 528–533.
(46) Blaha, G.; Stanley, R. E.; Steitz, T. A. *Science* **2009**, *325*, 966–970.

show that neither short nor long length scale motions are necessary to allow tertiary contacts, or even elements of native tertiary structure, to form within the exit port. The general conclusion drawn in this paper, that transient tertiary structure can often form within the exit port, is based on the fact that simulations in which the ribosome is held rigid or is allowed to undergo structural fluctuations both display this phenomena (Figure 2b). If larger fluctuations occur, the ribosome tunnel is likely to be able to accommodate even greater amounts of tertiary folding than that observed in the present work, which means that the structures described here are a lower bound on the extent to which tertiary structure can develop within the tunnel. More specifically, it has been suggested that the ribosome exit port would have to be highly dynamic to accommodate the Kv1.3 α-hairpin observed experimentally to form within the exit port since the radius of the α-hairpin, when it is assumed to be a sphere, is larger than the exit port radius, when it is assumed to be a cylinder.[24] Our results, however, indicate that such fluctuations are not required when the detailed geometries of both the tunnel and FTEs are accounted for in an explicit manner, such as has been done in these simulations.

Assuming that the ribosomal tunnel can be treated as a cylinder has been a useful approximation for estimating, experimentally or by simulations,[47] the characteristic length scales involved in co-translational folding. This assumption can, however, give rise to the conclusion that tertiary structure cannot form in the tunnel because of a mismatch in dimensions. Our results clearly demonstrate, however, that simple quantities that characterize the shape, size, and volume of FTEs do not explain where along the tunnel those FTEs can fit and possibly fold (Figure 7). This is because such simple geometric arguments lack sufficient spatial resolution to account for the irregular geometry of the tunnel and FTEs that determines the ability of an FTE to pack in the exit port.

Besides these general conclusions, we have been able to make specific predictions for a number of RNCs containing globular proteins concerning the nature of possible structure formation in the tunnel, which can, in principle, be tested using current technologies including NMR, cryo-EM, single-molecule FRET, and cross-linking experiments. The realistic modeling of FTEs in these simulations, as opposed to simplistic shape, size, and volume descriptors, explains the ability of elements of tertiary structure to fit into different portions of the tunnel, because such detail is required to capture accurately the packing effects that determine the minimum $d_{PTC,i}$ and minimum ΔPTC_i values. These conclusions suggest that simulations of the type described here represent a significant advance in assessing the behavior of the nascent chain during its biosynthesis by the ribosome and that it is possible, and likely to be valuable, to apply them to other macromolecular systems that are of interest or importance.

Experimental Procedures

All-atom simulations of the ribosome nascent chain complex were carried out using the Charmm 22 force field[48] in conjunction with the CHARMM program (version c33b2).[49] A distance-dependent dielectric constant was used to model implicitly the charge-shielding effect of water. Coarse-grained simulations were carried out using the C_α side-chain model force field[33] for the nascent chain and ribosomal proteins, and the 3,4-interaction site model for the ribosomal RNA bases as detailed in the Supporting Information. The coarse-grained linker was modeled as an extended strand whose force-field parameters were determined on the basis of PDB statistics. A number of simulation protocols were used in this study. In all simulations, low-friction Langevin dynamics was used with a 0.05 ps^{-1} damping coefficient and 2 and 10 fs integration time steps for the all-atom and coarse-grained simulations, respectively.

We determined the minimum ΔPTC_i and related quantities using two different protocols. In the first, we used constant temperature (310 K) simulations with a root-mean-squared-deviation (rmsd) restraint applied to the NC to keep the various FTEs folded during the simulation and used the interaction energy criterion to determine the minimum ΔPTC_i (Figure 1c). In the second protocol, we used replica exchange simulations but did not apply an rmsd restraint, which allowed the nascent chain to fold and unfold at equilibrium. From this protocol, the minimum ΔPTC_i was calculated as the smallest ΔPTC_i for which $P_i(\Delta PTC_i) > 0$ at 310 K. From those replica exchange simulations, the stability of the FTEs was calculated using the WHAM equations.[50] The coarse-grained A_{100}-RNC simulations were run at 310 K. The "cons fix" command in the Constraints module of CHARMM was used to restrain the ribosome interaction sites for the fixed ribosome simulations, and the "cons harm" command was used to allow the ribosome interaction sites to fluctuate with a variance of 3 $Å^2$ in the flexible ribosome simulations. Additional details of the methods used in this study are provided in the Supporting Information.

Acknowledgment. We thank Sophie Jackson and Stephanie O'Brien for a careful reading of the manuscript, Robert Best for providing the dihedral parameters for the Karanicolas and Brooks model, and Neil Voss and Peter Moore for sharing with us their volume-filling model of the ribosome tunnel. This work was supported in part by a NSF postdoctoral fellowship to E.P.O., a Human Frontier Science Program Long-term Fellowship (LT0798/2005) and the National Science Council of the Republic of China Taiwan (NSC97-2917-1-564-102) to S.-T.D.H., a Human Frontier Young Investigators Award (RGY67/2007) and BBSRC grant (9015651/1) to J.C., and a Wellcome Trust Programme grant to C.M.D.

Supporting Information Available: Detailed simulation and analysis procedures. This material is available free of charge via the Internet at http://pubs.acs.org

JA106530Y

(47) Ziv, G.; Haran, G.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 18956–18961.
(48) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(49) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(50) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.
(51) Improta, S.; Politou, A. S.; Pastore, A. *Structure* **1996**, *4*, 323–337.
(52) Gronenborn, A. M.; Filpula, D. R.; Essig, N. Z.; Achari, A.; Whitlow, M.; Wingfield, P. T.; Clore, G. M. *Science* **1991**, *253*, 657–661.
(53) Beamer, L. J.; Pabo, C. O. *J. Mol. Biol.* **1992**, *227*, 177–196.
(54) Musacchio, A.; Noble, M.; Pauptit, R.; Wierenga, R.; Saraste, M. *Nature* **1992**, *359*, 851–855.
(55) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531–544.
(56) McCoy, A. J.; Fucini, P.; Noegel, A. A.; Stewart, M. *Nat. Struct. Biol.* **1999**, *6*, 836–841.
(57) Zweifel, M. E.; Leahy, D. J.; Hughson, F. M.; Barrick, D. *Protein Sci.* **2003**, *12*, 2622–2632.
(58) Minor, D. L.; Lin, Y. F.; Mobley, B. C.; Avelar, A.; Jan, Y. N.; Jan, L. Y.; Berger, J. M. *Cell* **2000**, *102*, 657–670.
(59) van Thor, J. J.; Georgiev, G. Y.; Towrie, M.; Sage, J. T. *J. Biol. Chem.* **2005**, *280*, 33652–33659.