nature structural & molecular biology

A structural ensemble of a ribosome–nascent chain complex during cotranslational protein folding

Lisa D Cabrita[1,2,5], Anaïs M E Cassaignau[1,2,5], Hélène M M Launay[1,2,5], Christopher A Waudby[1,2], Tomasz Wlodarski[1,2], Carlo Camilloni[3], Maria-Evangelia Karyadi[1,2], Amy L Robertson[1,2], Xiaolin Wang[1,2], Anne S Wentink[1,2], Luke S Goodsell[1,2], Cheryl A Woolhead[4], Michele Vendruscolo[3], Christopher M Dobson[3] & John Christodoulou[1,2]

Although detailed pictures of ribosome structures are emerging, little is known about the structural and cotranslational folding properties of nascent polypeptide chains at the atomic level. Here we used solution-state NMR spectroscopy to define a structural ensemble of a ribosome–nascent chain complex (RNC) formed during protein biosynthesis in *Escherichia coli*, in which a pair of immunoglobulin-like domains adopts a folded N-terminal domain (FLN5) and a disordered but compact C-terminal domain (FLN6). To study how FLN5 acquires its native structure cotranslationally, we progressively shortened the RNC constructs. We found that the ribosome modulates the folding process, because the complete sequence of FLN5 emerged well beyond the tunnel before acquiring native structure, whereas FLN5 in isolation folded spontaneously, even when truncated. This finding suggests that regulating structure acquisition during biosynthesis can reduce the probability of misfolding, particularly of homologous domains.

The manner by which a protein acquires its correct tertiary structure, while avoiding alternative pathways leading to aberrant folding, is a fundamental processes that underpins the biological activity of all living systems[1]. A mechanistic understanding of the inherent nature of protein folding has come predominantly from extensive studies of isolated polypeptides renatured in dilute aqueous solutions, in which the folding process has been elegantly described by using energy landscapes[2,3]. However, the extent to which these characteristics describe folding within the cell is a prominent question in contemporary biology[4]. For the majority of proteins, folding processes begin in a cotranslational manner during biosynthesis on the ribosome[5–7], which leads to constant remodeling of the energy landscape as translation proceeds[8]. Cotranslational folding is thought to be a crucial means by which the cell carries out successful folding, particularly of polypeptide chains that would otherwise readily misfold[7,9,10].

A mechanistic understanding of protein biosynthesis is emerging through detailed structures of the functional ribosome[11], but there is little structural understanding of the emerging nascent chain because its inherent dynamics has eluded most high-resolution techniques. During biosynthesis, the nascent chain is synthesized at a rate of ~10–20 amino acids per second in prokaryotes[12], and its folding is under at least some form of translational control; thus, for example, the presence of synonymous codons within mRNA sequences has been observed to adversely affect the folding efficiency[6,10] of nascent chains. As the nascent chain elongates, it emerges in a vectorial manner from the restricted environment of the ribosomal exit tunnel, enters the crowded cellular milieu and begins to explore conformational space. A range of ancillary proteins such as molecular chaperones[13] and those mediating processing and translocation[14] are present, and the ribosome is a central hub for many of these proteins, which compete for binding to the nascent chain[15]. Most notable among these proteins is the ribosome-associated molecular chaperone trigger factor[16,17], which binds to emerging polypeptide chains at the ribosomal exit tunnel[18]. In addition, the ribosomal surface itself has been suggested to influence this process through transient electrostatic interactions between the emerging nascent chain and the ribosomal surface[9,19]; in some cases, these interactions appear to alter the rate and efficiency of folding[9].

The manner by which nascent chains sample structural conformations has been primarily investigated with translationally arrested RNCs, in which local compaction observed with fluorescence resonance energy transfer (FRET) probes on the nascent chain has been used to propose structure formation[10]. Putative cotranslational protein-folding intermediates[20] have also been identified through fluorescence measurements[20] and through biochemical studies[21], which have proposed that structural conformations formed on the ribosome may differ from those populated *in vitro*. Cryo-EM analysis of RNCs[22] has shown that nascent chains remain largely extended as they are extruded through the ribosomal exit tunnel, and additional structural[23] and biochemical evidence[24,25] has indicated that some amino acid sequences promote the formation of incipient structures, such as α-helices, in distinct regions of the tunnel. Although it has also more recently been shown that a simple tertiary motif can form within the exit tunnel[26], higher-order structure appears to be formed only when a nascent chain has emerged. A detailed understanding of the progressive acquisition of the tertiary structure of the nascent chain

[1]Institute of Structural and Molecular Biology, University College London, London, UK. [2]Institute of Structural and Molecular Biology, Birkbeck College, University of London, London, UK. [3]Department of Chemistry, University of Cambridge, Cambridge, UK. [4]Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK. [5]These authors contributed equally to this work. Correspondence should be addressed to J.C. (j.christodoulou@ucl.ac.uk).

Received 10 July 2015; accepted 2 February 2016; published online 29 February 2016; doi:10.1038/nsmb.3182

outside the ribosome is absent. In this study, therefore, we set out to use NMR spectroscopy to report on both structure and dynamics during folding at a residue-specific level[27,28]. We produced a structural ensemble of a highly dynamic nascent chain of a pair of immunoglobulin-like domains emerging during biosynthesis. In addition, we characterized, in solution, a set of RNCs generated *in vivo* in *E. coli* and produced a series of high-resolution snapshots that reveal structural details of cotranslational protein folding.

RESULTS

Isotopically labeled RNCs produced *in vivo* in *E. coli*

To explore the structure and dynamics of nascent chains as they emerge from the ribosome, we studied a polypeptide chain whose sequence was based on a pair of immunoglobulin-like proteins, $FLN5_{646–750}$ and $FLN6_{751–857}$, the fifth and sixth filamin domains of the *Dictyostelium discoideum* gelation factor[27]. We initially designed an FLN5-6 RNC, denoted FLN5+110 (ref. 27), in which the C terminus of the 105-residue FLN5 domain was separated from the peptidyl transferase center (PTC) by 110 residues comprising a folding-incompetent FLN6 domain (with an 18-residue truncation at its C terminus (residues 840–857)[28], referred to here as FLN6Δ18) and the 17-residue SecM translation-arrest motif[29]. All RNC and isolated protein designations for the FLN5 and FLN6 variants used in this study are shown in **Supplementary Figure 1**. We generated all the RNCs in *E. coli*[27], in which folding takes place in the cellular milieu, and purified the intact RNCs in high yield as previously described[27] (**Fig. 1a**).

FLN5 acquiring native-like structure on the ribosome

To use NMR spectroscopy to probe folded and unfolded conformations of FLN5+110 RNC, we developed a dual isotopic labeling scheme, using both selective protonation and uniform labeling approaches. Because methyl-group resonances are highly sensitive reporters of changes in protein tertiary structure, we generated selectively labeled RNCs on a perdeuterated (^{2}H) background in which only the isoleucine δ1 side chain of the nascent chain was labeled with $^{13}CH_3$. The replacement of all surrounding ^{1}H by ^{2}H nuclei results in longer relaxation times and more intense signals[30]. We then examined samples of the uniform (U) [U-^{2}H] Ileδ1-[$^{13}CH_3$] labeled RNC via ^{1}H-^{13}C correlation spectra (by using methyl TROSY NMR methods[30]). We identified resonances from all five FLN5 isoleucine residues in ^{1}H-^{13}C correlation spectra of FLN5+110 RNC, and these overlaid closely (^{1}H and ^{13}C chemical-shift changes <0.01 and 0.1 p.p.m., respectively) with those of isolated FLN5 (**Fig. 1b**), thus indicating that in this nascent chain the FLN5 domain had folded into a native conformation. In parallel, we produced U-^{15}N–labeled FLN5+110 RNCs, in which the peptide backbone was isotopically labeled, and we recorded ^{1}H-^{15}N correlation spectra via band-selective optimized flip-angle short transient heteronuclear multiple quantum coherence (SOFAST-HMQC) experiments[31]. Examination of the ^{1}H-^{15}N correlation spectra of the U-^{15}N–labeled FLN5+110 RNC showed nascent-chain resonances within a narrow window of ^{1}H chemical shifts, a result indicative of disordered structure. The chemical shifts of the nascent-chain resonances corresponded closely to those observed for unfolded FLN6 (in spectra of isolated FLN5-6Δ18 (ref. 28)) rather than unfolded FLN5 (in spectra of a 12-residue C-terminal truncation, FLN5Δ12) (**Fig. 1c**).

Because these combined NMR data were exquisite probes for both the folded and unfolded structural preferences of FLN5 and FLN6 tethered to the ribosome, they enabled us to use chemical shifts measured for FLN5+110 RNC as replica-averaged structural restraints in molecular dynamics simulations[32] to determine a structural ensemble of the RNC (**Fig. 1d**,**e** and **Supplementary Video 1**). This ensemble showed folded FLN5 tethered to the ribosome by a disordered FLN6. Despite lacking persistent structure, FLN6 was compact and exhibited transient populations (of ~20% on average) of native-like secondary-structure elements and inter-residue contacts (**Fig. 1d**–**h**). The ensemble also illustrated that FLN5 had substantial access to a broad region of the ribosomal surface and formed transient contacts with both 23S rRNA and ribosomal protein L29, as a result of its tethering to the disordered FLN6 (**Fig. 1d**,**e**). We analyzed the regions of the ribosome in proximity to FLN6 and observed that FLN6 formed transient interactions with 23S RNA and several ribosomal proteins associated with the exit port (**Fig. 1f**). The most frequent degrees of contact were with L24 (55%), specifically with a prominent loop in proximity to the exterior of the exit port; such contact is further supported by nascent-chain cross-linking studies[33] and a cryo-EM structure of a ribosome–SecYE complex[34], which has shown that this loop undergoes marked conformational changes in the presence of a nascent chain derived from the periplasmic protein FtsQ. In addition, FLN6 made transient yet substantial contacts with L23 (30%), whose position on the surface near the exit tunnel has been shown in structural studies[15] to be an adaptor site for ancillary proteins including trigger factor.

These RNC ensemble structures suggested that the conformational freedom of the nascent chain was likely to be tempered by its interactions with the ribosomal surface (**Fig. 1d**) and that these interactions had both structural and dynamical implications for the processes by which a vectorially emerging nascent-chain sequence forms its complex tertiary structure beyond the ribosomal tunnel. Previous studies[35] have shown that isolated FLN5 folds highly cooperatively via a weakly populated folding intermediate, thus raising the question of a possible role for the ribosome itself in modulating the folding of FLN5 nascent chains as they emerge during biosynthesis.

Structural evidence for cotranslational folding of FLN5

To probe how FLN5 acquires its structure during biosynthesis, we extended our NMR approach to analyze a series of 12 RNCs, in which the FLN6 linker was progressively shortened (**Fig. 2a**,**b**); each NMR spectrum represented a unique snapshot during biosynthesis and reported on cotranslational protein folding at equilibrium. The series of SecM-arrested nascent chains consisted of FLN5 with decreasing numbers of residues of the FLN6 sequence, ranging from 110 to 21 residues (**Fig. 2b**). The RNCs, denoted FLN5+L (with linker length L = 21 to 110), purified from *E. coli* cells in similar yields to that of the FLN5+110 RNC, and a series of biochemical and biophysical analyses showed that all RNCs were completely intact (**Fig. 2c**) and free of any extraneous proteins, including trigger factor and DnaK (**Supplementary Fig. 1**). The continuous cycling of these ubiquitous cytosolic chaperones[17,36] with the ribosome and nascent chains alike indicated that the RNCs had considerable access to such chaperones during cotranslational folding within the cell; however, their absence after purification (<1.5% occupancy; **Supplementary Fig. 1**) indicated that FLN5 RNCs are relatively poor substrates[16] for these particular species. We isotopically labeled each of the FLN5 RNC samples as either Ileδ1-[$^{13}CH_3$] or U-^{15}N, and we acquired ^{1}H-^{13}C and ^{1}H-^{15}N correlation spectra, respectively. For all samples, the acquisition of these spectra was accompanied by rigorous control experiments including interleaved NMR diffusion and cross-peak intensity measurements, in conjunction with western blotting (**Fig. 2c** and **Supplementary Fig. 2**), to ensure that the data used for structural analysis were derived exclusively from intact RNCs.

As shown in the ^{1}H-^{13}C correlation spectra of FLN5+110 RNC (**Fig. 1b**), resonances from all five FLN5 isoleucine residues could

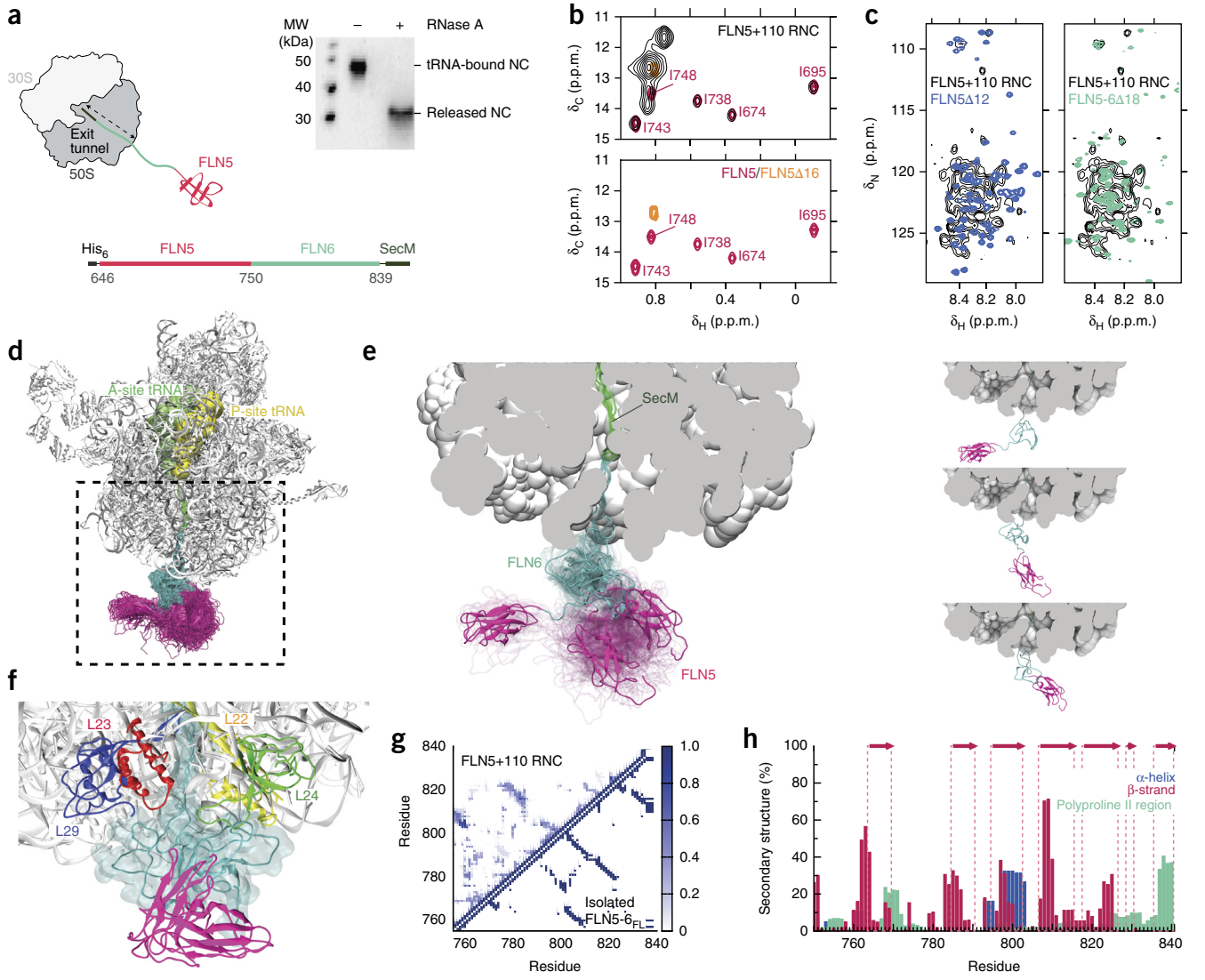

Figure 1 Structural ensemble of a ribosome-bound nascent chain. (**a**) Schematic of the FLN5+110 RNC used for the ensemble calculations. The FLN5 sequence is tethered to the ribosome by a C-terminally truncated $FLN6_{751–839}$ sequence and stalled with the SecM translation-arrest motif[27,29]. Anti-hexahistidine (His_6) western blot of purified FLN5+110 RNC, in both the ribosome-attached form with bound prolyl P-site tRNA and the released form. MW, molecular weight. (**b**) Overlay of ^{1}H-^{13}C correlation spectra of Ileδ1-[$^{13}CH_3$]-labeled FLN5+110 RNC (black), isolated, natively folded FLN5 (pink) and isolated unfolded FLN5Δ16 (orange). (**c**) Overlay of ^{1}H-^{15}N correlation spectra of U-^{15}N–labeled FLN5+110 RNC with isolated FLN5Δ12 (blue) and unfolded FLN5-6Δ18 (green). (**d**) NMR chemical shift–restrained structural ensemble of FLN5+110 RNC, showing the disordered FLN6 linker (cyan) and the native fold acquired by FLN5 (pink) (accession codes, PDB 2N62 and BMRB 25748). (**e**) Close-up view of the ribosomal exit tunnel, highlighting three representative conformations of the nascent-chain ensemble (left); the three representative conformations are also shown separately (right). (**f**) Transient interactions made between the disordered FLN6 linker (cyan) and the proximal ribosomal proteins at the surface. (**g**) Probability of the formation of inter-residue contacts in the FLN5+110 RNC (shown above diagonal) and in the native state of isolated full-length (FL) FLN5-6 (shown below diagonal) (**h**) Secondary-structure populations of the RNC, depicting β-strands (magenta), α-helices (blue) and polyproline II regions (green); native β-strands are indicated (arrows).

be similarly identified in ^{1}H-^{13}C correlation spectra of FLN5+67 and FLN5+47 samples, thus indicating that in these nascent chains the FLN5 domain had also folded into a native or near-native conformation (**Fig. 3a**). The intensities of the dispersed resonances in FLN5+47 were, however, only 30% (±12%) of those within the corresponding spectra of FLN5+67 and FLN5+110, a feature further discussed below. Moreover, in the spectra of FLN5+45, only three of the five native-like isoleucine resonances were visible, all with very low intensity (**Fig. 3a**), no such resonances at all could be detected in spectra of the RNC with the shortest linker length, FLN5+21.

Using each of the FLN5 RNCs, we also recorded ^{1}H-^{15}N correlation spectra to monitor the presence or absence of resonances of the unfolded form of the FLN5 domain in each sample. Examination of the spectra (**Fig. 3b**) revealed that when *L* was between 21 and 44 residues, all the resonances of the nascent chain appeared within a narrow window of ^{1}H chemical shifts, a result indicative of disordered structure. The chemical shifts of the nascent-chain resonances corresponded closely to those observed in spectra of unfolded forms of isolated FLN5 generated by a C-terminal truncation, FLN5Δ12 (**Fig. 3b** and **Supplementary Fig. 3**) or by a destabilizing mutation in the FLN5 variant Y719E (**Supplementary Fig. 3**). The average intensities of these RNC cross-peaks were, however, reduced substantially in spectra of FLN5+43 and FLN5+44 RNCs, and no comparable unfolded FLN5 resonances were visible in the spectra of FLN5+45 to FLN5+110

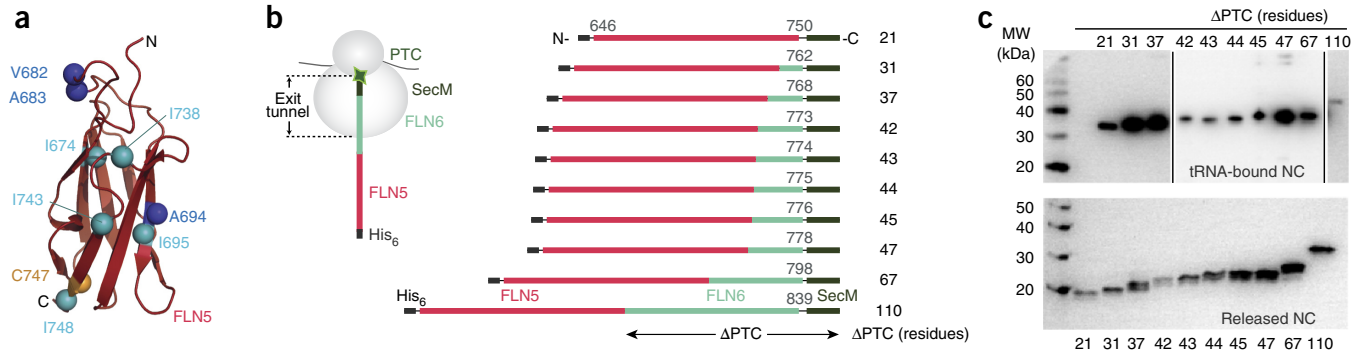

Figure 2 Design and *in vivo* production of FLN5 ribosome–nascent chain complexes in *E. coli*. (**a**) Structure of isolated natively folded FLN5 (PDB 1QFH). Mapped onto the FLN5 structure are the five isoleucines (δ1 methyl groups) of FLN5 (Ile674, Ile695, Ile738, Ile743 and Ile748, cyan) used as probes of native structure acquisition and the amide groups of three residues (Val682, Ala683 and Ala694, blue) selected for analysis of unfolded conformations (described in main text). (**b**) Design of the translationally arrested RNCs[27] used to monitor nascent-chain emergence and folding, in which the FLN5 sequence is tethered to the PTC with increasing lengths of the FLN6 sequence and the SecM translation-arrest motif. (**c**) Anti-histidine western blots of the library of purified FLN5 RNCs in ribosome-bound (top; additional data in **Supplementary Fig. 1**) and released (bottom) forms. MW, molecular weight.

RNCs. In addition, we identified cross-peaks attributable to the emerging FLN6 sequence in an unfolded state (**Supplementary Fig. 3**) in spectra of FLN5+67 (**Fig. 3b**), as in the FLN5+110 RNC (**Fig. 1c**). These NMR data clearly showed an increasing population of the folded state of FLN5 relative to its unfolded state as the length of the sequence joining it to the PTC increased, and they also showed a concomitant appearance of peaks from disordered residues from FLN6.

To further evaluate the transition from the unfolded to the folded state as FLN5 emerged from the tunnel, we selected three amide resonances of FLN5 that were particularly well resolved and did not overlap with other resonances in the spectra of U-^{15}N–labeled RNCs (**Fig. 3b** and **Supplementary Fig. 3**). These resonances had comparable ^{1}H linewidths (20 ± 3 Hz) in all RNCs from FLN5+21 to FLN5+42 (**Supplementary Fig. 3**), thus indicating that, at least for these residues, any differences in intensity associated with nascent-chain length were attributable to changes in the population of the unfolded form of the nascent chain rather than to changes in relaxation behavior. Indeed, analysis of the signal intensities indicated that the population of the unfolded state of FLN5 decreased substantially in samples for which L = 42 to 45, and the length-dependent changes in the amide resonance intensities of the disordered FLN5 nascent chain were consistent with an unfolded-to-folded transition with a midpoint between L = 42 and 45 (**Fig. 4a**). Also consistent with this conclusion was our observation of native-like resonances of the isoleucine methyl groups of FLN5 in ^{1}H-^{13}C correlation spectra starting from FLN5+45 and continuing through to FLN5+110 RNCs. We attributed the weak intensity of the methyl resonances in nascent chains with L = 45 and 47 to the low mobility of the folded FLN5 domain, resulting from its proximity to the slowly tumbling ribosome, rather than attributing it to a reduction in the population of the folded state. The increases in the intensity of these resonances, as evident for nascent chains with L = 67 and 110, then reflected the gain in mobility of the folded FLN5 domain as the length of the chain linking it to the PTC increased (**Fig. 3a**).

Folding of FLN5 RNCs is offset relative to isolated FLN5

To rationalize these spectral observations of folding as the nascent chain elongates, we examined the accessibility to solvent of the emerging FLN5 domain by probing the native cysteine residue (Cys747; **Fig. 2a**), which is located close to the FLN5-FLN6 boundary, for its susceptibility to modification by methoxypolyethylene glycol maleimide (PEG-Mal) (**Fig. 4a,b**). Under the experimental conditions used, the 5-kDa moiety, as shown previously[24], can substantially (>80%) modify a cysteine in a nascent chain only if it is beyond the exit tunnel, i.e., more than ~100 Å from the PTC[37]. Therefore, we used a series of RNCs of the folding-incompetent variant FLN5 Y719E (**Supplementary Fig. 4**) to monitor the emergence of Cys747 from the ribosomal exit tunnel, without the complication of the cysteine residue being shielded from solvent as a result of structure acquisition

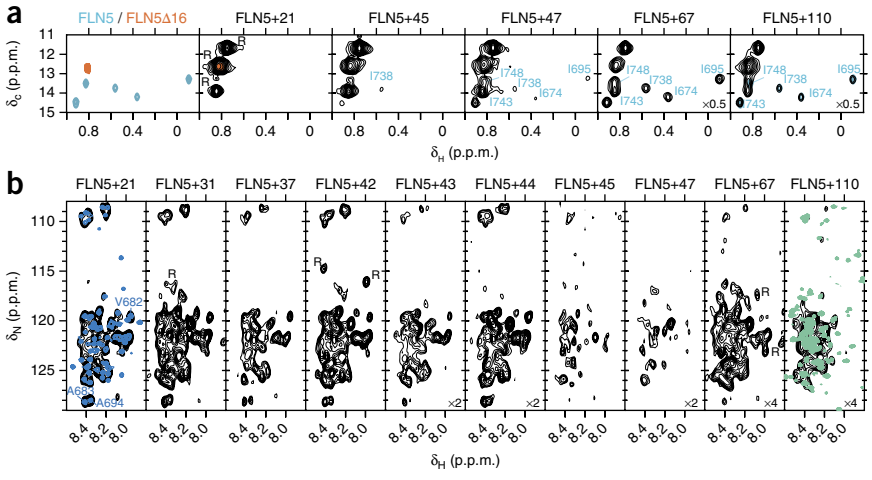

Figure 3 Nascent chains of FLN5 emerging from the ribosome, as monitored by NMR spectroscopy. (**a**) ^{1}H-^{13}C correlation spectra of Ileδ1-[^{13}CH$_3$]-labeled FLN5 RNCs (black), isolated, natively folded FLN5 (cyan) and isolated unfolded FLN5Δ16 (orange). Resonances marked R arise from background labeling of 70S ribosomal proteins[27]. (**b**) ^{1}H-^{15}N correlation spectra of U-^{15}N–labeled FLN5 RNCs, compared with disordered controls, unfolded FLN5 in FLN5Δ12 (blue) and unfolded FLN6 in FLN5-6Δ18 (green). Resonances used for the analysis of unfolded FLN5 conformations are labeled in the FLN5+21 RNC spectrum.

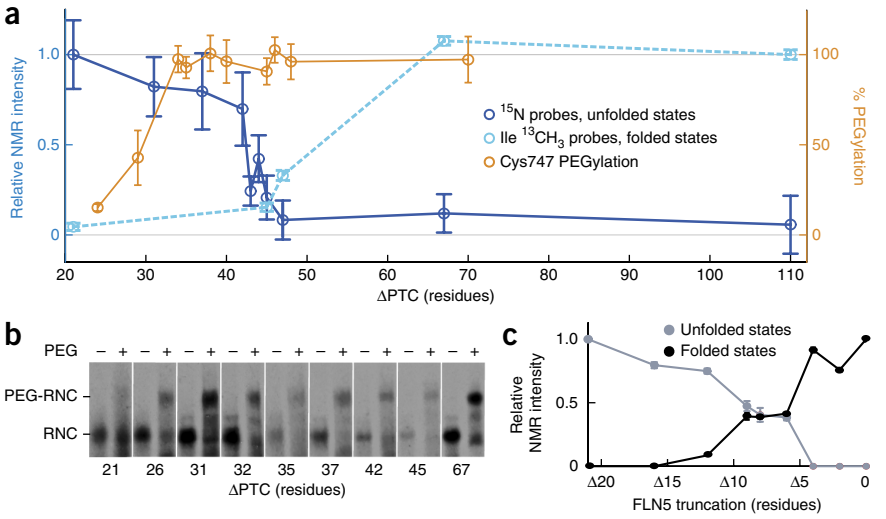

Figure 4 Folding of FLN5 on the ribosome, as monitored by NMR spectroscopy and PEGylation. (**a**) FLN5 nascent-chain folding, as measured by intensity changes of ^{15}N amide resonances (blue) arising from the unfolded FLN5 domain (mean ± s.d. for $n = 4$ ($n = 3$ for +45 and $n = 2$ for +47) nascent-chain concentrations from western blot replicates) and intensity changes in isoleucine $^{13}CH_3$ resonances (cyan) arising from native FLN5 structure (mean ± s.d. derived from spectral noise for one single measurement). Intensities are normalized and scaled relative to $L = 21$ (unfolded) or $L = 110$ (folded). The solvent accessibility of the FLN5 domain from the ribosomal exit tunnel was probed with PEGylation (orange) (mean ± s.d., $n = 5$ sample replicates) of folding-incompetent FLN5 Y719E RNCs, in which the native Cys747 is close to the boundary between FLN5 and FLN6. (**b**) Cys747 PEGylation of FLN5 Y719E RNCs results in a band shift (PEG-RNC). (**c**) C-terminal truncations of isolated FLN5, as measured by NMR. Averaged cross-peak intensities of folded (black) and unfolded (gray) states of FLN5 (additional data in **Supplementary Fig. 5**) are mapped against truncation length.

in the FLN5 domain. Under conditions analogous to those of the NMR experiments (and adapted from those well established[24] to achieve PEGylation of a nascent chain entirely emerged from the tunnel), we observed complete PEGylation for $L \geq 31$, i.e., when Cys747 was ≥34 residues from the PTC (**Fig. 4a**,**b** and **Supplementary Fig. 4**). This result showed that at these nascent-chain lengths, the entire FLN5 sequence had emerged from the tunnel to an extent that enabled it to be accessible by PEG-Mal well before the folding of the domain could be observed ($L > 44$) by NMR spectroscopy, as discussed above.

We next generated a series of C-terminal truncations of the isolated $FLN5_{646–750}$ domain to examine the length dependence of folding of this domain in the absence of the ribosome. We analyzed ^{1}H-^{15}N correlation spectra of nine C-terminal truncations ranging from FLN5Δ2 to FLN5Δ21. The spectra indicated that FLN5Δ12 and its shorter variants, which lacked the C-terminal β-strand G and its adjacent loop in the native structure, were fully unfolded under the conditions used in this study (**Fig. 4c** and **Supplementary Fig. 5**). In contrast, the longer variants FLN5Δ2 and FLN5Δ4 were fully natively folded, whereas sequences of lengths between FLN5Δ6 and FLN5Δ9 populated both folded and unfolded states at equilibrium. From these results, we concluded that the isolated FLN5 domain in bulk solution could tolerate truncation of up to nine residues and still populate a folded state to a substantial degree.

The observation that FLN5Δ4 (residues 646–746 of FLN5) was fully folded in its isolated state was notable because the next residue in the RNC FLN5+31, Cys747, was solvent accessible and hence was clear of the exit tunnel, as shown by the PEGylation experiments discussed above (**Fig. 4a**,**b**). Acquisition of native structure would therefore, in principle, be possible even in the case of FLN5+31, i.e., when $L = 31$ and Cys747 has emerged from the tunnel, as indicated by this residue's accessibility to PEGylation. The NMR data, however, showed that folding of the FLN5 domain took place only when a further 11 to 14 residues of FLN6 had been added to the sequence (**Fig. 4a**). There was thus a substantial difference between the length of the FLN5 polypeptide sequence required for the acquisition of native structure by the isolated

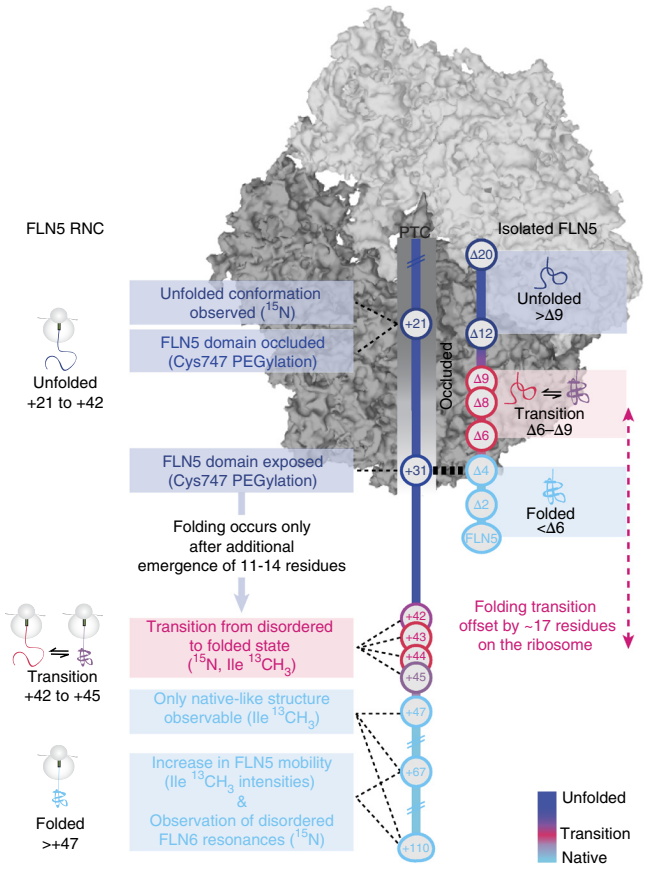

Figure 5 FLN5 folding is offset on the ribosome. Comparison of FLN5 folding on the ribosome and in isolation, as depicted by the conformational states observed for FLN5 RNCs with different linker lengths, L, from the PTC and those of C-terminal FLN5 truncations: blue, unfolded; cyan, folded; pink, folding transition. The sequence of the FLN5 nascent chains is solvent exposed, as monitored by PEGylation, at $L \geq 31$ residues, in which Cys747 is 34 residues from the PTC, yet the domain acquires native-like structure only upon addition of a further 11–14 residues, at $42 \leq L \leq 45$, as monitored by NMR spectroscopy. Isolated FLN5 truncations are shown alongside the RNC lengths with FLN5Δ4 as a reference for $L = 31$, at which point complete PEGylation is observed. A folding offset (pink dashed arrow) indicates the difference observed in the initiation of FLN5 structure acquisition on the ribosome relative to the protein in isolation.

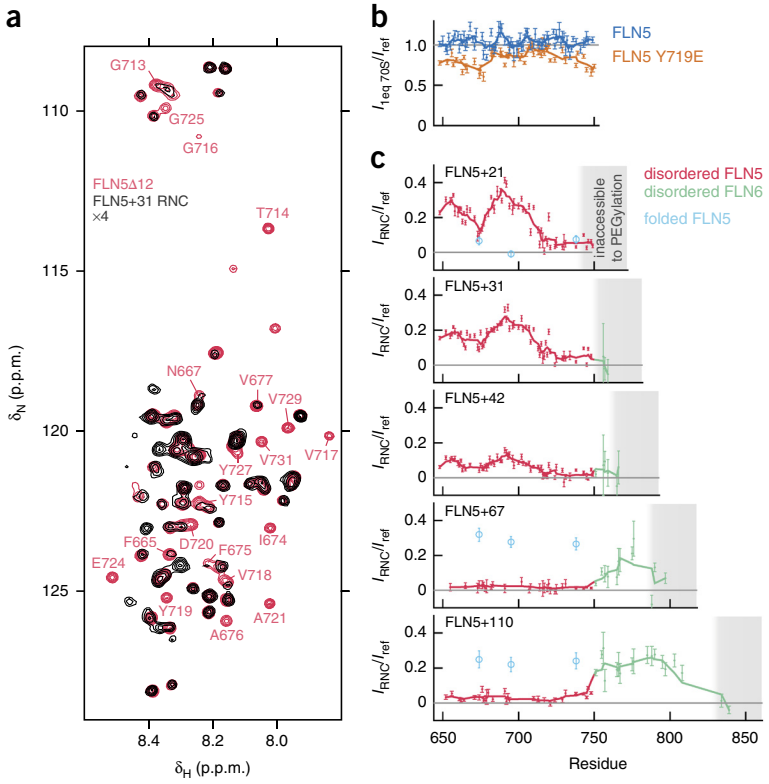

Figure 6 Residue-specific mapping of RNC interactions. (**a**) An overlay of ^{1}H-^{15}N correlation spectra (recorded at a ^{1}H frequency of 950 MHz) of FLN5+31 RNC (black) and unfolded, isolated FLN5Δ12 (red), highlighting resonances that are substantially broadened in the RNC. (**b**) Relative intensities of ^{1}H-^{15}N resonances of folded FLN5 (5 μM) and unfolded FLN5 Y719E (8 μM) in the presence of 1 molar equivalent of 70S ribosomes. (**c**) Relative intensities (I) of FLN5+21, FLN5+31, FLN5+42, FLN5+67 and FLN5+110 RNCs compared with an isolated reference comprising unfolded FLN5 (red), unfolded FLN6 (green) and folded FLN5 (cyan) are shown (additional information in Online Methods). A five-point moving average is plotted as a guide; errors are derived from spectral noise from a single measurement. The gray shaded area denotes occluded residues inaccessible to PEGylation.

domain in bulk solution and by the domain when attached to the ribosome; indeed, the folding transition on the ribosome, compared with that observed for the isolated protein, remarkably, required the availability of an additional 17 residues (**Fig. 5**).

We initially explored the origins of the offset between the solvent accessibility of the complete FLN5 domain and its folding during biosynthesis by examining the interdomain interactions between the emerged FLN5 and sections of the successive FLN6 linking sequence (**Fig. 1d**,**g**), by substituting the FLN6 residues with a poly(glycine-serine) linker (L_{GS}) to generate a series of FLN5+L_{GS} RNCs (**Supplementary Fig. 6**). Under identical conditions to those used for the FLN5 RNCs, complete PEGylation of the FLN5+L_{GS} RNCs occurred at L_{GS} ≥35 (Cys751, i.e., 34 residues from the PTC). NMR observations of L_{GS} = 31, 37 and 42 showed only disordered FLN5 resonances, thus suggesting that FLN5 folded independently, regardless of the linking sequence, and it was unlikely that interdomain interactions alone were the cause of the offset observed for folding (**Supplementary Fig. 6**).

The ribosome surface modulates the energy landscape of FLN5

Our structural ensemble of the FLN5+110 RNC revealed that the emerging nascent chain interacted transiently with ribosomal surface proteins (**Fig. 1d**,**f**), and we assessed this issue further with high-resolution 2D ^{1}H-^{15}N correlation spectra (**Fig. 6a**). We reasoned that such interactions might influence the capacity for a nascent chain to acquire structure. Addition of an equimolar concentration of 70S ribosomes to samples of the isolated unfolded variants FLN5 Y719E (**Fig. 6b**) and FLN5Δ12 (**Supplementary Fig. 7**) resulted in moderate (~30%) reductions in the intensities of the resonances of Lys663 to Val677 and Gly713 to Gly750. However, we did not observe analogous intensity changes after addition of 70S ribosomes to a sample of full-length, folded FLN5 (**Fig. 6b** and **Supplementary Fig. 7**), thus indicating that the intensity changes were the result of broadening attributable to the binding of unfolded FLN5 to the slowly tumbling ribosome particle. Analysis of the spectra of the various FLN5 RNCs showed that when L = 21 to 42, in which the domain was unfolded, resonances of Phe665 to Val677 and Gly713 to Gly750 were similarly reduced in intensity. The effects were, however, much more substantial than those observed for the isolated domain: the resonances of Phe665 to Val667 lost more than 70% of their

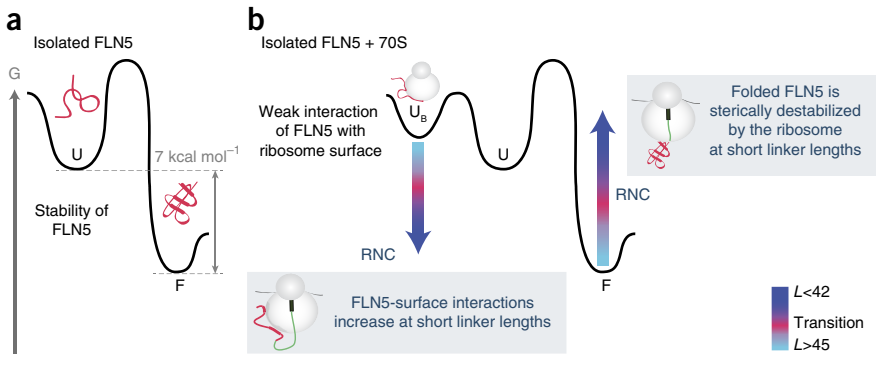

Figure 7 The ribosome modulates the folding landscape of FLN5 nascent chains. (**a**) Schematic of a free-energy diagram for isolated FLN5, showing the difference in free energy (G) of 7 kcal mol^{-1} between the folded state (F) and the unfolded state (U). (**b**) Schematic free-energy diagram for isolated FLN5 in the presence of ribosomes, showing a ribosome-bound state (U_B), which is accessible from the unfolded state. A model for how the ribosome may alter this landscape and inhibit nascent-chain folding is indicated (arrows). At short linker lengths, the tethered nascent chain is subject to high effective ribosome concentrations, thus favoring the ribosome-bound state. The native state is also likely to be energetically unfavorable, owing to steric interactions with the ribosome. As the nascent chain increases in length, the steric effects and the ribosome-associated interactions experienced by the tethered nascent chain are overcome by the stability of the folded FLN5 domain.

intensities, and those of Gly713 to Gly750 became completely undetectable (**Fig. 6a**,**c**).

The same FLN5 residues showed similar reductions in intensity in analogous spectra of L_{GS} = 31 and 42 in the FLN5+L_{GS} RNCs (**Supplementary Fig. 6**). These data therefore indicated that the specific stretches of sequence identified from the spectra of unfolded FLN5 interact with the ribosomal surface, a phenomenon previously observed for other RNC systems[9,19]. The greater extent of ribosome surface interactions of FLN5 in the RNCs, relative to that of the isolated FLN5 domain in the presence of the ribosome, may be attributed to a higher effective concentration of the ribosome as a result of its anchoring to the PTC. Such an effect would be most pronounced at short linker lengths (L = 21 to 42); indeed, we estimated the effective concentration to be 20 mM for a C-terminal residue located ten residues beyond the exit tunnel (Online Methods). This effect would increase the magnitude of the interaction between unfolded FLN5 and the ribosomal surface, particularly at the C terminus, which includes residues Gly713 to Gly750, a result consistent with our observations. As a consequence, the unfolded state would be stabilized relative to the native state[8] at short RNC linker lengths, $21 < L < 44$ (**Fig. 7b**) and could therefore inhibit folding of the domain when attached to the ribosome, as compared to its isolated state. As the nascent chain elongated, the interactions with the ribosome surface were progressively reduced, and only at $L > 42$ did they become insufficient to overcome the 7 kcal mol^{-1} free energy of folding measured for isolated FLN5 in bulk solution[35] (**Fig. 7a**,**b**).

The mechanism by which the ribosome–nascent chain interactions specifically acquire the capacity to modulate nascent-chain folding and achieve the observed folding offset is likely to be related to the effects of steric occlusion (particularly at short linker lengths) (**Fig. 7b**) and to be directed by the sequence determinants inherent to the nascent chain. In the latter case, the sequence similarity between FLN5 and FLN6 suggests that the two domains are likely to form similar transient interactions with the ribosome in their disordered states (**Fig. 1**). Because replacing FLN6 with a poly(GS) linker did not abrogate the folding offset (**Supplementary Fig. 6**), the transient interactions with the ribosome appear to act independently on each emerging domain, rather than requiring a preceding domain to interact with a specific ribosomal protein or RNA at the ribosomal exit, and to be responsible for transmitting a 'folding trigger'. Therefore, for a multidomain protein such as FLN, composed of homologous domains, ribosome-nascent chain interactions implicated in a folding offset may occur as each domain emerges sequentially, rather than in a coordinated intradomain manner.

DISCUSSION

In summary, we used NMR spectroscopy to determine the structural ensemble of a folded multidomain nascent chain on its parent ribosome. The ensemble provided clear insights into the dynamic process of cotranslational folding: the globular FLN5 domain possessed a high degree of conformational freedom resulting from the presence of a compact, disordered FLN6 domain; the latter showed transient yet substantial interactions with both ribosomal RNA and the ribosomal proteins surrounding the exit site, particularly L24. Our studies of the changes to the structural ensembles formed by shortened RNCs afforded a residue-specific understanding of how a nascent chain acquires native-like structure during its progressive emergence from the ribosomal exit tunnel. Indeed, in the case of the protein domain studied here, the folding of the tethered nascent chain did not take place as soon as a sequence of polypeptide chain capable of folding in bulk solution emerged from the ribosomal tunnel.

A certain degree of compaction of the nascent chain along with contributions from specific interactions of the disordered state of FLN5 (that are likely to be analogous to those observed for FLN6 in the structural ensemble), and the ribosomal surface are observed. These appear to permit persistent folding to occur only after emergence of an additional segment, here consisting of 11–14 residues of the subsequent FLN6 sequence. In living cells, however, cotranslational folding is not an equilibrium process but occurs in parallel with the process of translation (10–20 residues s^{-1})[12], thus potentially resulting in an offset between the points at which folding occurs on actively translating ribosomes compared with stalled ribosomes[38]. Thus, the continuous translation process would indicate that the folding of FLN5 may be completed at linker lengths longer than that at which we observed FLN5 folding to occur in stalled RNCs. Folding *in vitro* on a timescale of ~1 s^{-1}, a value typical of immunoglobulin domains[39], may produce an offset of 10–20 residues between the polypeptide chain length at which FLN5 folding becomes thermodynamically favorable and the point at which folded populations can form kinetically.

Therefore, at least in this system, interactions with the ribosome during emergence from the tunnel inhibit the acquisition of stable structure by a nascent chain rather than promote native-like contacts in a progressive manner, as suggested for other systems[5,7,9,10]. This phenomenon has apparent similarities to the behavior of some molecular chaperones, described as holdases[40], that inhibit the formation of misfolded and potentially toxic aggregates by stabilizing more highly unfolded states[41]. We suggest that regulating the acquisition of partially folded structures within a nascent chain during cotranslational folding of a protein may act in a similar manner to ensure efficient generation of functional proteins within living systems by reducing the probability of misfolding, particularly of multidomain proteins with high sequence identities among domains[42]. Indeed, such a mechanism suggests that cotranslational folding of neighboring individual domains may be remarkably similar to the cooperative folding *in vitro*[3] rather than to acquisition of native-like structure occurring gradually during the process of biosynthesis.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Coordinates for the structural ensemble have been deposited in the Protein Data Bank under accession code PDB 2N62, and NMR chemical-shift restraints have been deposited in the Biological Magnetic Resonance Data Bank under accession code 25748.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank J. Kirkpatrick for NMR technical assistance and useful discussions and B. Bukau (Ruprecht-Karls-Universität Heidelberg) for the kind gift of the anti-SecM antibody. J.C. acknowledges the use of the Biomolecular NMR Facility, University College London, and thanks T. Frenkiel and G. Kelly of the Medical Research Council Biomedical NMR Centre at the Crick Institute, London for the use of the facility. J.C. and T.W. acknowledge the use of the Advanced Research Computing High End Resource (ARCHER) UK National supercomputing service (http://www.archer.ac.uk/). L.D.C. is supported by the Wellcome Trust and by an Alpha-1 Foundation grant. A.L.R. is supported as a National Health and Medical Research Council (Australia) C.J. Martin Fellow. T.W. is supported as an European Molecular Biology Organization Long-Term Fellow and is also supported by the Wellcome Trust. The work of C.M.D. and M.V. is supported by a Wellcome Trust Programme Grant (094425/Z/10/Z to C.M.D. and M.V.). This work was supported by a Biotechnology and Biochemical Sciences Research

Council New Investigators Award (BBG0156511 to J.C.) and a Wellcome Trust Investigator Award (097806/Z/11/Z to J.C.).

AUTHOR CONTRIBUTIONS

L.D.C., C.M.D. and J.C. designed the research. J.C. supervised the overall project. L.D.C., A.M.E.C., H.M.M.L., C.A. Waudby, M.-E.K., A.S.W. and X.W. performed the research. A.L.R. and L.D.C. performed the PEGylation experiments, and C.A. Woodhead contributed valuable technical experience in *in vitro*–translation experiments. C.C., T.W., L.S.G. and M.V. performed the NMR-restrained molecular dynamics simulations. L.D.C., A.M.E.C., A.L.R., C.A. Waudby and J.C. analyzed the data. L.D.C., A.M.E.C., C.A. Waudby, M.V., C.M.D. and J.C. wrote the paper. All authors discussed the results and contributed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Dobson, C.M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
2. Onuchic, J.N., Luthey-Schulten, Z. & Wolynes, P.G. Theory of protein folding: the energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
3. Dobson, C.M., Sali, A. & Karplus, M. Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng.* **37**, 868–893 (1998).
4. Cabrita, L.D., Dobson, C.M. & Christodoulou, J. Protein folding on the ribosome. *Curr. Opin. Struct. Biol.* **20**, 33–45 (2010).
5. Nicola, A.V., Chen, W. & Helenius, A. Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat. Cell Biol.* **1**, 341–345 (1999).
6. Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274–280 (2009).
7. Frydman, J., Erdjument-Bromage, H., Tempst, P. & Hartl, F.U. Co-translational domain folding as the structural basis for the rapid *de novo* folding of firefly luciferase. *Nat. Struct. Biol.* **6**, 697–705 (1999).
8. O'Brien, E.P., Christodoulou, J., Vendruscolo, M. & Dobson, C.M. New scenarios of protein folding can occur on the ribosome. *J. Am. Chem. Soc.* **133**, 513–526 (2011).
9. Kaiser, C.M., Goldman, D.H., Chodera, J.D., Tinoco, I. Jr. & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).
10. Kim, S.J. *et al.* Protein folding. Translational tuning optimizes nascent protein folding in cells. *Science* **348**, 444–448 (2015).
11. Schmeing, T.M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234–1242 (2009).
12. Siller, E., DeZwaan, D.C., Anderson, J.F., Freeman, B.C. & Barral, J.M. Slowing bacterial translation speed enhances eukaryotic protein folding efficiency. *J. Mol. Biol.* **396**, 1310–1318 (2010).
13. Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A. & Bukau, B. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **400**, 693–696 (1999).
14. Schaffitzel, C. *et al.* Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature* **444**, 503–506 (2006).
15. Kramer, G., Boehringer, D., Ban, N. & Bukau, B. The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat. Struct. Mol. Biol.* **16**, 589–597 (2009).
16. Rutkowska, A. *et al.* Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.* **283**, 4124–4132 (2008).
17. Kaiser, C.M. *et al.* Real-time observation of trigger factor function on translating ribosomes. *Nature* **444**, 455–460 (2006).
18. Ferbitz, L. *et al.* Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* **431**, 590–596 (2004).
19. Knight, A.M. *et al.* Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics. *ACS Chem. Biol.* **8**, 1195–1204 (2013).
20. Kelkar, D.A., Khushoo, A., Yang, Z. & Skach, W.R. Kinetic analysis of ribosome-bound fluorescent proteins reveals an early, stable, cotranslational folding intermediate. *J. Biol. Chem.* **287**, 2568–2578 (2012).
21. Clark, P.L. & King, J. A newly synthesized, ribosome-bound polypeptide chain adopts conformations dissimilar from early *in vitro* refolding intermediates. *J. Biol. Chem.* **276**, 25411–25420 (2001).
22. Bhushan, S. *et al.* SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol.* **9**, e1000581 (2011).
23. Bhushan, S. *et al.* α-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **17**, 313–317 (2010).
24. Lu, J. & Deutsch, C. Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **12**, 1123–1129 (2005).
25. Woolhead, C.A., Johnson, A.E. & Bernstein, H.D. Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol. Cell* **22**, 587–598 (2006).
26. Nilsson, O.B. *et al.* Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep.* **12**, 1533–1540 (2015).
27. Cabrita, L.D., Hsu, S.T., Launay, H., Dobson, C.M. & Christodoulou, J. Probing ribosome-nascent chain complexes produced *in vivo* by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* **106**, 22239–22244 (2009).
28. Hsu, S.T. *et al.* Structure and dynamics of a ribosome-bound nascent chain by NMR spectroscopy. *Proc. Natl. Acad. Sci. USA* **104**, 16516–16521 (2007).
29. Nakatogawa, H. & Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629–636 (2002).
30. Rosenzweig, R. & Kay, L.E. Bringing dynamic molecular machines into focus by methyl-TROSY NMR. *Annu. Rev. Biochem.* **83**, 291–315 (2014).
31. Schanda, P., Kupce, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR* **33**, 199–211 (2005).
32. Camilloni, C., Cavalli, A. & Vendruscolo, M. Replica-averaged metadynamics. *J. Chem. Theory Comput.* **9**, 5610–5617 (2013).
33. Peterson, J.H., Woolhead, C.A. & Bernstein, H.D. The conformation of a nascent polypeptide inside the ribosome tunnel affects protein targeting and protein folding. *Mol. Microbiol.* **78**, 203–217 (2010).
34. Frauenfeld, J. *et al.* Cryo-EM structure of the ribosome–SecYE complex in the membrane environment. *Nat. Struct. Mol. Biol.* **18**, 614–621 (2011).
35. Hsu, S.T., Cabrita, L.D., Fucini, P., Dobson, C.M. & Christodoulou, J. Structure, dynamics and folding of an immunoglobulin domain of the gelation factor (ABP-120) from *Dictyostelium discoideum*. *J. Mol. Biol.* **388**, 865–879 (2009).
36. Deuerling, E. *et al.* Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.* **47**, 1317–1328 (2003).
37. Nissen, P., Hansen, J., Ban, N., Moore, P.B. & Steitz, T.A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
38. O'Brien, E.P., Ciryam, P., Vendruscolo, M. & Dobson, C.M. Understanding the influence of codon translation rates on cotranslational protein folding. *Acc. Chem. Res.* **47**, 1536–1544 (2014).
39. Clarke, J., Cota, E., Fowler, S.B. & Hamill, S.J. Folding studies of immunoglobulin-like beta-sandwich proteins suggest that they share a common folding pathway. *Structure* **7**, 1145–1153 (1999).
40. Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M. & Hartl, F.U. Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).
41. Hartl, F.U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
42. Wright, C.F., Teichmann, S.A., Clarke, J. & Dobson, C.M. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature* **438**, 878–881 (2005).

ONLINE METHODS

Generation of ribosome–nascent chain complexes (RNCs) and isolated C-terminal truncations. DNA constructs of RNCs of tandem domains FLN5 and FLN6 were derived from a FLN5+110 RNC construct described previously[27]. Site-directed mutagenesis was used to manipulate the length of the 110–amino acid FLN6 linker to generate a set of SecM-stalled FLN5 RNCs with linker lengths, *L*, ranging from 21 to 110 residues (*L* = 21, 26, 31, 35, 37, 42, 43, 44, 45, 47, 67 and 110). Selectively isotopically labeled, histidine-tagged RNCs were generated in BL21(DE3) *E. coli* with an *in vivo* procedure described previously[27] with modifications. Cells were grown in an unlabeled MDG medium at 37 °C and then were washed and resuspended in an M9-based expression medium (EM9, adapted from ref. 43) enriched with the relevant isotopes. RNC expression was induced with 1 mM IPTG, and after 10 min, 150 mg/mL rifampicin[44] was added, and the cells were harvested 35 min later. Uniform ^{15}N labeling was performed as described previously[27]. The production of [U-^{2}H]-Ileδ1-[$^{13}CH_3$]–labeled RNCs, in which the δ1 methyl group of the isoleucine side chain was selectively protonated, was achieved under perdeuterated (^{2}H) conditions with the isoleucine precursor 2-ketobutyric-4-^{13}C,3,3-d_2 acid in a procedure adapted from that described previously for U-^{15}N–labeled RNCs[27], in which the cells were progressively adapted into the deuterated isotopes and precursors. Rifampicin was omitted during the induction period, and cells were harvested after 1.5 h. The purification of RNCs from *E. coli* was performed as described previously[27], except that the ribosomal material was recovered from the lysate with a 35% (w/v) high-salt sucrose cushion before purification with an Ni-IDA column and then with a 10–35% (w/v) sucrose gradient. Site-directed mutagenesis was used to introduce the Y719E point mutation into FLN5 RNCs and into isolated FLN5, and to substitute the FLN6 linker for a glycine-serine repeat sequence (poly(GS)). Isolated C-terminal truncations of FLN5 (residues 646–750) were generated by removing between 2 and 21 amino acids (FLN5Δ2 Δ4, Δ6, Δ8, Δ9, Δ12, Δ16 and Δ21) with mutagenesis. Each of the FLN5 variants was expressed and purified as previously described for full-length FLN5 (ref. 28).

Determination of RNC integrity and nascent-chain occupancy. For evaluating RNC integrity, samples were run on denaturing 12% (w/v) polyacrylamide bis-Tris gels at neutral pH and with a sample dye at pH 5.7 to maintain the ester bond between the tRNA and the nascent chain. Released forms of the nascent chain were obtained by treatment of the RNC samples with RNase A. For determination of nascent-chain occupancy, RNase A–treated RNCs were run alongside isolated protein concentration standards, and anti-histidine western blots were analyzed with ImageJ (http://imagej.nih.gov/ij/) software.

RNC integrity over the time course of NMR experiments, as monitored by western blotting. Purified RNCs were incubated at 25 °C, and 5-pmol aliquots were collected periodically to examine the integrity of the tRNA-bound form of the nascent chain over time, in the same samples used in NMR experiments. All samples were analyzed by western blotting, and the intensity of the band corresponding to tRNA-bound nascent chain was assessed by densitometry.

Detection and quantification of trigger factor and DnaK in RNC samples. Purified RNC samples were treated with RNase A and assessed by western blotting for the presence of trigger factor with a rabbit polyclonal anti–trigger factor antibody (cat. no. A01329, GenScript). A similar procedure was used for the detection of DnaK with an anti-DnaK antibody (cat. no. LS-C63274-50, Source Bioscience). The residual amount of both trigger factor and DnaK present within the RNCs was determined with densitometry analysis with purified trigger factor and DnaK proteins as standards, as described above.

Coupled transcription and translation of RNCs *in vitro*. An *E. coli* S30 cell extract was prepared as previously described[25]. A pair of primers (5′ primer upstream of the T7 promoter, 5′-CTCGATCCCGCGAAATTAATACG-3′; 3′ primer partially overlapping the SecM-stalling sequence, 5′-AGGTCCATGGTTAAGGGCCAG-3′) was used to produce linear templates encoding SecM-stalled RNCs from the relevant plasmids. Reactions were performed in 25-μL volumes with an S30 extract and a translation premix[25], containing 1.5 μg linear DNA, 0.04 mM L-amino acids, 20 μL (10 units) T7 RNA polymerase, 5 μCi [^{35}S]methionine and 200 ng/μL antisense ssrA oligonucleotide. Transcription-translation reactions were incubated at 37 °C for 30 min, and the RNCs were isolated from 30% (w/v) sucrose cushions and centrifuged at 100,000 r.p.m. for 1 h.

PEGylation gel-shift assay of RNCs. Pelleted *in vitro*–derived RNCs, corresponding to approximately 6 pmol of 70S ribosomes were resuspended in buffer A (20 mM HEPES, pH 7.2, 100 mM NaCl, and 5 mM $MgCl_2$). Samples were divided, and the aliquots subjected to PEGylation reactions were incubated in buffer A containing 1 mM methoxypolyethylene glycol maleimide (5 kDa). Samples were then incubated at 25 °C for 1 h. After PEGylation reactions, the samples were run with PAGE conditions as described above. The gel was exposed to film, and the extent of PEGylation in each RNC was evaluated by densitometry with ImageJ software, in which the intensity of the PEGylated, tRNA-bound form was evaluated relative to that of the unPEGylated tRNA-bound form within the same sample. The PEGylation data reported are the average of at least five independent experiments.

NMR spectroscopy of RNCs. Before NMR spectroscopy, each sample was buffer-exchanged into Tico buffer[27], pH 7.5 (containing d_8-HEPES for $^{13}CH_3$-labeled RNCs), supplemented with 1 mM EDTA and protease inhibitors. The samples also contained 7% (v/v) D_2O (U-^{15}N–labeled samples) or 100% D_2O (Ileδ1-[$^{13}CH_3$]–labeled samples) as a lock signal and 0.001% (w/v) DSS as an internal reference. Sample concentrations were based on the nascent-chain content and ranged from 2 to 12 μM. NMR data were acquired on a 700 MHz Bruker Avance III spectrometer (University College London) equipped with a TXI cryoprobe, and in specific cases with 800- and 950-MHz Bruker Avance III HD spectrometers (NMR Centre, Crick Institute). All spectra were recorded at 298 K unless otherwise stated, in an interleaved manner[27]. For samples of U-^{15}N–labeled RNCs, ^{1}H-^{15}N SOFAST-HMQC spectra at 700 MHz were recorded with 1,024 points in the direct (^{1}H) dimension (T_{aq} = 46 ms) and 64 points (128 points for poly(GS)-linker RNCs) in the indirect (^{15}N) dimension (T_{aq} = 14.1 ms), with a recycling delay of 50 ms. ^{1}H-^{13}C HMQC spectra of Ileδ1-[$^{13}CH_3$]–labeled RNCs at 700 MHz were recorded with 3,072 points in the direct (^{1}H) dimension (T_{aq} = 137.6 ms) and 128 points in the indirect (^{13}C) dimension (T_{aq} = 12.1 ms). For all RNCs recorded at 700 MHz, either ^{15}N XSTE[45] or ^{1}H STE–^{1}H-^{13}C HMQC[46] diffusion measurements were acquired with a diffusion delay of 100 ms and bipolar trapezoidal gradient pulses (total length 4 ms, shape factor 0.9) with strengths of 0.028 and 0.529 T m^{-1}. Spectra were recorded at 800 and 950 MHz with a nonuniform weighted sampling scheme, a 50-ms acquisition time in the direct (^{1}H) dimension (spectral width 16 p.p.m.), 160 points in the indirect (^{15}N) dimension (spectral width 22 p.p.m.), and a recycling time of 50 ms. The indirect dimension was acquired with a cosine nonuniform weighted scheme, providing an 11% increase in intensity[47]. These data were interleaved with SORDID diffusion measurements[48] with a diffusion delay of 190 ms and trapezoidal gradient pulses (total length 4 ms, shape factor 0.9) with strengths of 0.058 and 0.387 T m^{-1}. All data were processed and analyzed with NMRPipe[49] and Sparky (http://www.cgl.ucsf.edu/home/sparky/).

RNC labeling efficiency and selectivity, as assessed by ^{15}N filtered/edited difference spectroscopy. Isotopic labeling of the 70S ribosome particle was monitored in U-^{15}N–labeled RNCs. A ^{15}N-edited 1D experiment was recorded with modified ^{15}N SOFAST-HMQC sequences with 500-ms presaturation of water for suppression of the disordered nascent-chain resonances that exchange rapidly with the solvent. The observed signals therefore arise predominantly from nonlabile amides of the folded domain of ribosomal protein L7/L12 (ref. 50). A ^{15}N-filtered experiment was run identically, except with the phase cycle of the receiver inverted to reject ^{15}N-labeled magnetization (^{14}N-bound ^{1}H). The intensity of the ^{1}H envelope of 70S ribosomal resonances bound to ^{14}N (^{15}N-filtered 1D) was matched by scaling to that of ^{15}N-bound ^{1}H (^{15}N-edited 1D) to quantify the ratio of unlabeled to labeled ribosomal protons. From these measurements, the extent of background labeling arising from the ribosomal proteins was determined to range between 1% and 15% across all samples (~50 samples) of ^{15}N-labeled RNCs. An analogous approach was applied to purified, released nascent chains, and the extent of nascent-chain labeling was determined to be >90%.

Cotranslational folding, as monitored by ^{1}H-^{15}N correlation spectra. Three well-resolved resonances with signal-to-noise ratios of 12 ± 2 (corresponding

doi:10.1038/nsmb.3182

to residues Ala683, Ala694 and Val682) within ^{1}H-^{15}N correlation spectra of RNCs L = 21 to 42 were used for lineshape analysis. Spectra were processed with exponential window functions, and 1D cross-sections were fitted to Lorentzian lineshapes. The averaged calculated linewidths of the resonances of these residues were 12 ± 1 Hz for FLN5Δ16 and 20 ± 3 Hz for FLN5 RNCs (error, s.d.). The similar linewidths measured for Ala683, Ala694 and Val682 in disordered FLN5 RNCs (L = 21 to 42) indicated that these resonances had uniform relaxation properties and could therefore be used to monitor the populations of the disordered state. Peak intensities of FLN5+L RNCs (L = 21, 31, 37, 42, 43, 44, 45, 47, 67 and 110) were determined with Sparky (http://www.cgl.ucsf.edu/home/sparky/), scaled for the number of scans and relative nascent-chain concentrations, and averaged across the three peaks. Peak-height errors were calculated as the s.d. of 100 points randomly selected in the baseline of these RNC spectra. Nascent-chain concentrations were determined with anti-histidine western blot analysis of the RNC, as described above. At least two independent experiments were performed for each RNC sample, and the error was determined from the s.d. of these experiments. Relative intensities (I) of the residues from FLN5+21, FLN5+31, FLN5+42, FLN5+67 and FLN5+110 RNCs were measured relative to a reference of unfolded FLN5 (made up of a composite consisting of FLN5Δ12 and FLN5 Y719E) and unfolded FLN6 (from FLN5-6Δ18) from ^{1}H-^{15}N correlation spectra measured at high field (FLN5+21 and FLN5+110 RNCs at 800 MHz and FLN5+31, FLN5+42 and FLN5+67 at 950 MHz). Relative intensities (I) of the three well-dispersed Ileδ1-$^{13}CH_3$ resonances in FLN5+21, FLN5+67 and FLN5+110 RNCs were measured against those of folded FLN5 from ^{1}H-^{13}C correlation spectra.

Assignment of FLN5 Y719E and disordered FLN6. FLN5 Y719E amide chemical shifts were assigned on the basis of an assigned FLN5Δ12 spectrum and with ^{15}N NOESY-HSQC (200-ms mixing time) and ^{15}N TOCSY-HSQC (70-ms mixing time) experiments recorded at 277 K. Unfolded FLN6 $^{1}H_N$ and ^{15}N chemical shifts were assigned (except for residues 810–832) with a FLN5-6Δ18 construct[28], which gave rise to resonances that closely overlaid with those of natively folded FLN5 and showed additional resonances characteristic of disordered chemical shifts of unfolded FLN6. FLN5-6Δ18 amide chemical shifts were assigned at 283 K with uniformly ^{15}N-^{13}C–labeled samples via standard triple resonance experiments (HNCO, HN(CA)CO, HNCACB and HN(CO)CACB).

Estimation of the effective ribosome concentration experienced by a nascent chain. The effective local concentration of a binding site near the exit tunnel on the ribosome surface, as experienced by a residue in a nascent chain, can be estimated with previously described methods[51]. If the unfolded polypeptide outside the exit tunnel is analyzed with a random flight model, the mean distance from a residue at the end of the exit tunnel (taken here to be 31 residues from the PTC on the basis of PEGylation measurements, **Fig. 4a**,**b**) to a point N residues along the chain (i.e., N + 31 residues from the PTC) is approximately $\langle r^2 \rangle = CNl^2$, where the C_α-C_α distance l = 3.8 Å, and the characteristic ratio C = 9 accounts for the stiffness of a typical polypeptide chain[51]. By modeling the ribosome surface as an infinite plane, the effective local concentration of a binding site situated close to the exit tunnel can be determined to be $c_L = 2(3/2\pi\langle r^2 \rangle)^{3/2}/1{,}000N_A$ (in mol L^{-1}), where N_A is Avogadro's number[51]. For residues 10 to 20 amino acids beyond the exit tunnel (i.e., linker lengths L = 41–51), this corresponds to effective concentrations between 8 and 23 mM.

Structure calculations with chemical shift–restrained molecular dynamics simulations. Structural ensemble calculations of the FLN5+110 RNC were performed with the replica-averaged metadynamics (RAM) method previously described[32]. In these calculations, chemical shifts are used as replica-averaged structural restraints in molecular dynamics simulations with GROMACS[52] together with PLUMED2 (ref. 53). We used the CHARMM22* force field[54] with TIP3P water molecules[55]. A time step of 2 fs was used together with LINCS constraints[56]. The van der Waals and Coulomb interactions were cut off at 0.9 nm, whereas long-range electrostatic effects were treated with the particle mesh Ewald method. All simulations were carried out in the canonical ensemble by keeping the volume fixed and by thermosetting the system at 300 K with a Bussi-Donadio-Parrinello thermostat[57].

43. Sivashanmugam, A. *et al.* Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.* **18**, 936–948 (2009).
44. Rutkowska, A. *et al.* Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. *FEBS Lett.* **583**, 2407–2413 (2009).
45. Ferrage, F., Zoonens, M., Warschawski, D.E., Popot, J.L. & Bodenhausen, G. Slow diffusion of macromolecular assemblies by a new pulsed field gradient NMR method. *J. Am. Chem. Soc.* **125**, 2541–2545 (2003).
46. Didenko, T., Boelens, R. & Rüdiger, S.G. 3D DOSY-TROSY to determine the translational diffusion coefficient of large protein complexes. *Protein Eng. Des. Sel.* **24**, 99–103 (2011).
47. Waudby, C.A. & Christodoulou, J. An analysis of NMR sensitivity enhancements obtained using non-uniform weighted sampling, and the application to protein NMR. *J. Magn. Reson.* **219**, 46–52 (2012).
48. Augustyniak, R., Ferrage, F., Damblon, C., Bodenhausen, G. & Pelupessy, P. Efficient determination of diffusion coefficients by monitoring transport during recovery delays in NMR. *Chem. Commun. (Camb.)* **48**, 5307–5309 (2012).
49. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
50. Christodoulou, J. *et al.* Heteronuclear NMR investigations of dynamic regions of intact *Escherichia coli* ribosomes. *Proc. Natl. Acad. Sci. USA* **101**, 10949–10954 (2004).
51. Timpe, L.C. & Peller, L. A random flight chain model for the tether of the Shaker K^+ channel inactivation domain. *Biophys. J.* **69**, 2415–2418 (1995).
52. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
53. Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: new feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
54. Piana, S., Lindorff-Larsen, K. & Shaw, D.E. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* **100**, L47–L49 (2011).
55. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, M.L. & Klein, L. Comparison of simple potential functions for stimulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
56. Hess, B., Bekker, H., Berendsen, H.J.C. & Fraaije, J. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
57. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).