



# Investigating the Effect of Chain Connectivity on the Folding of a Beta-Sheet Protein On and Off the Ribosome

Andrew P. Marsden<sup>†</sup>, Jeffrey J. Hollins<sup>†</sup>, Charles O'Neill, Pavel Ryzhov, Sally Higson, Carolina A.T.F. Mendonça, Tristan O. Kwan, Lee Gyan Kwa, Annette Steward and Jane Clarke

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK

Correspondence to Annette Steward and Jane Clarke: [as376@cam.ac.uk](mailto:as376@cam.ac.uk); [jc162@cam.ac.uk](mailto:jc162@cam.ac.uk)

<https://doi.org/10.1016/j.jmb.2018.10.011>

Edited by Sheena Radford

## Abstract

Determining the relationship between protein folding pathways on and off the ribosome remains an important area of investigation in biology. Studies on isolated domains have shown that alteration of the separation of residues in a polypeptide chain, while maintaining their spatial contacts, may affect protein stability and folding pathway. Due to the vectorial emergence of the polypeptide chain from the ribosome, chain connectivity may have an important influence upon cotranslational folding. Using MATH, an all  $\beta$ -sandwich domain, we investigate whether the connectivity of residues and secondary structure elements is a key determinant of when cotranslational folding can occur on the ribosome. From  $\Phi$ -value analysis, we show that the most structured region of the transition state for folding in MATH includes the N and C terminal strands, which are located adjacent to each other in the structure. However, arrest peptide force-profile assays show that wild-type MATH is able to fold cotranslationally, while some C-terminal residues remain sequestered in the ribosome, even when destabilized by 2–3 kcal mol<sup>-1</sup>. We show that, while this pattern of  $\Phi$ -values is retained in two circular permutants in our studies of the isolated domains, one of these permutants can fold only when fully emerged from the ribosome. We propose that in the case of MATH, onset of cotranslational folding is determined by the ability to form a sufficiently stable folding nucleus involving both  $\beta$ -sheets, rather than by the location of the terminal strands in the ribosome tunnel.

© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

The relationship between the folding of domains in isolation and cotranslational folding on the ribosome has been investigated in recent years using a range of methods [1,2] including studies that have compared *in vitro* folding with folding *in vivo* [3–12]. Such studies have shown that elements of secondary and tertiary structure can form, and small peptides can fold, in the ribosome tunnel [13–16] and that entire protein domains can fold in the exit vestibule of the ribosome [17,18]; the ribosome itself has also been shown to affect the folding process [12,19–21]. While some proteins appear to fold by the same pathway on and off the ribosome [18,22], other studies suggest that the folding pathways may differ [17,23–25]. Recent studies also suggest that the process of cotranslational folding

can provide a further level of protein regulation [26–29].

The vectorial emergence of the peptide chain from the ribosome tunnel prompts the question of how circular permutation of a protein affects cotranslational folding. Circular permutants (CPs) of isolated domains have been extensively studied to investigate the relationship between protein topology, chain connectivity, and folding pathways [30–35]: the permuted protein is covalently linked at the N and C termini, and new termini are generated elsewhere in the sequence, usually in a loop region [36]. This allows retention of the same amino acid composition and chain length as wild-type, but alters the connectivity of secondary structure elements; this in turn may lead to alterations in protein stability [37–39], enzyme activity [40,41], and folding pathway [42–44].

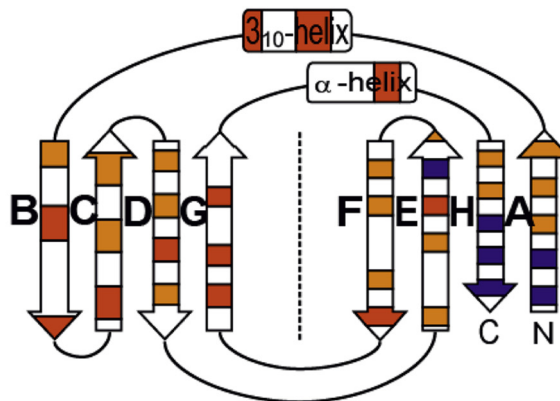
In order to study the effect of circular permutation on folding, both on and off the ribosome, we employed a MATH domain, a member of the TRAF domain-like superfamily, a  $\beta$ -sandwich domain containing a circularly permuted immunoglobulin (Ig)-fold topology with an extra strand [45] (Fig. 1). The N and C termini are adjacent to one another and the terminal H-strand is central to the structure. Using  $\Phi$ -value analysis, we determined the folding pathway for MATH wild-type and two CPs. We also use an *in vitro* arrest-peptide-based force-measurement assay (AP assay) [46–48], to determine when MATH can begin to fold, in a stalled ribosome system, and how this is affected by changes in chain connectivity upon circular permutation.

Translational arrest peptides [49,50] such as SecM have been employed previously to study protein insertion into membranes and also to study cotranslational protein folding on the ribosome [14,17,46–48,51]. In such assays, an amino acid linker of increasing length ( $L$ ) is placed between the protein and the arrest site (Fig. 2); when translation is stalled by the arrest peptide, the stall can be released by the application of a moderate force, such as that generated by a protein as it folds in close proximity to the ribosome [51,52]. The probability of release is proportional to the force applied: to release the stall, the protein must be able to fold at that force and have a sufficiently long lifetime to release the stall [51]. The fraction of full-length protein ( $f_{FL}$ ) (Fig. S1) released in a given time interval, at each value of  $L$ , gives insights into when folding occurs during translation.

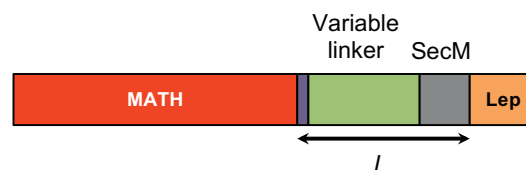
## Results

### WT folding and analysis of the kinetic data

Kinetic data for MATH wild-type and variants show rollover in both unfolding and refolding (Fig. S2).



**Fig. 1.** MATH wild-type structure and  $\Phi$ -value analysis. (left) MATH domain with  $\Phi$ -values displayed. (right) Schematic view of MATH showing the topology of the protein with  $\Phi$ -values displayed. High  $\Phi$ -values in blue (0.6–1), medium  $\Phi$ -values in orange (0.3–0.59), and low  $\Phi$ -values in red (0–0.29).



**Fig. 2.** Construct used in the arrest peptide assay. MATH domain (red), SGSG linker (purple), variable linker (green), SecM stall site (gray), terminal Lep domain (orange). (Full amino acid sequences for MATH wild-type, CP-CD, and CP-DE are shown in Supplementary Information).

Rollover in the refolding arm is protein concentration independent, suggesting that MATH folds *via* an early, off-pathway intermediate (see discussion in Supplementary Information). Moreover, rollover observed in the unfolding arm is consistent with a high-energy unfolding intermediate [53] (Fig. S3). Analysis of kinetic data is described in Supplementary Information. Since the refolding kinetics are complicated by the presence of the roll-over, all  $\Phi$ -values were determined (Eq. (1)) using unfolding rate constants of wild-type and mutant ( $k_u^{wt}$  and  $k_u^{mut}$ ) at 3 M guanidinium chloride (GdmCl), where we had accurate data for wild-type and all mutants:

$$\Phi = 1 - (\Delta\Delta G_{\ddagger-N} / \Delta\Delta G_{D-N}) \quad (1)$$

where  $\Delta\Delta G_{\ddagger-N} = -RT \ln(k_u^{wt}/k_u^{mut})$ . The change in free energy on unfolding,  $\Delta\Delta G_{D-N}$ , was determined using equilibrium denaturation curves.

### Terminal strands A and H are most structured region in MATH transition state

A  $\Phi$ -value analysis was performed to elucidate the folding pathway of wild-type MATH: 39 mutations at 37 separate locations were made across the protein (Table S1). Only mutations that were sufficiently

destabilizing ( $\Delta\Delta G_{D-N} > 0.75$  kcal mol<sup>-1</sup>) [54] were considered for  $\Phi$ -value analysis.

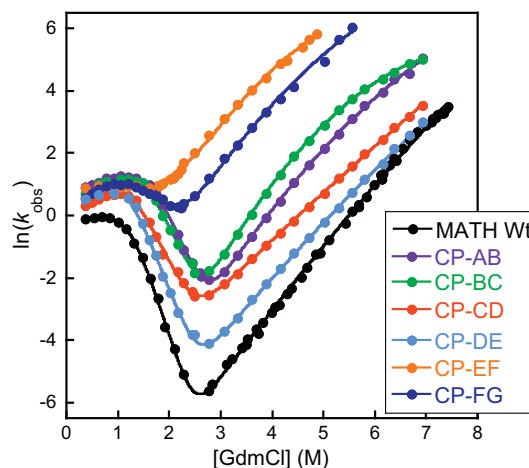
Except for the cross-sheet helical regions and strands B and G,  $\Phi$ -values are all  $>0$ , suggesting that most regions of MATH are at least partially structured in the transition state for folding (Fig. 1 and Table S1). We find the highest  $\Phi$ -values are in the N and C terminal strands, which includes the peripheral A-strand (N-terminus) and structurally central H-strand (C-terminus). Overall, residues in the sheet containing strands F, E, H, and A have higher  $\Phi$ -values than those in the sheet containing strands B, C, D, and G (Fig. 1 and Table S1). The most structured region of the transition state for folding in wild-type MATH is located in the area around the N- and C-terminal strands (strands A and H, respectively).

### Circular permutation increases both folding and unfolding rates

To determine the effects of circular permutation upon the stability and folding and unfolding rates of MATH, a series of CPs were made. The N and C termini were joined (with no additional residues added) and new N and C termini were located in each of the seven loops and turns (see [Materials and Methods](#) and Fig. S4 for details). All CPs were designed to disrupt the native connectivity of the wild-type domain while minimizing the potential impact on native structure and topology; we use the nomenclature CP-XX, where XX denotes the position of the new N and C termini (thus, CP-AB has new termini in the A–B loop). We characterized the stability and kinetics of six of these CPs (Table S2). The exception is CP-GH, which could not be produced solubly. All CPs are stable and, like wild-type, fold in a reversible two-state manner at equilibrium; all CPs showed rollover in the refolding arm, as observed in wild-type MATH (Fig. 3). Most significantly, our chevron plots reveal that both the folding and unfolding rate constants of MATH are increased in all permutants.

### CP-CD and CP-DE display a similar pattern of $\Phi$ -values to wild-type

In order to determine how changes in connectivity affect the folding pathway of MATH, partial  $\Phi$ -value analyses were performed on two of the permutants, CP-CD and CP-DE (Fig. 4, Tables S3 and S4). These permutants were selected as new termini are generated in the center of the BCDG sheet in CP-CD, whereas new termini are generated between the two  $\beta$ -sheets in CP-DE (Fig. S4); also, both permutants are stable and easy to produce and have linear unfolding arms (Fig. 3). A number of conservative mutations were made throughout the core of the proteins to probe tertiary structure

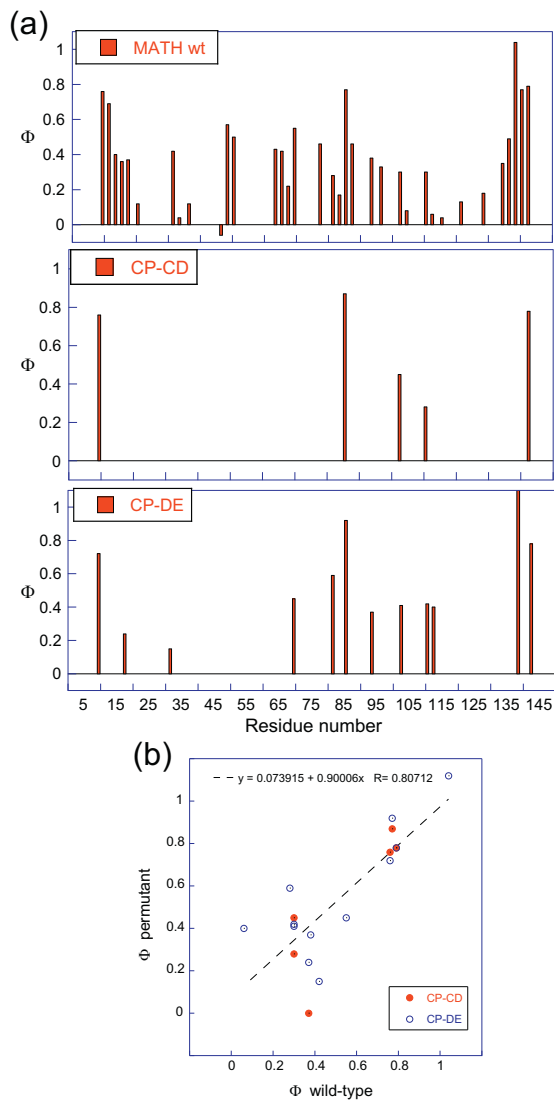


**Fig. 3.** Kinetic data for MATH wild-type and CPs. All MATH permutants fold and unfold more rapidly than wild-type.

formation and, where possible, matched to mutations made previously in wild-type. As for wild-type MATH,  $\Phi$ -values were determined using unfolding rate constants at 3 M GdmCl and equilibrium values of  $\Delta\Delta G$ . For CP-CD, six variant proteins could be produced, which were sufficiently destabilized for  $\Phi$ -value analysis; a further nine variants were found to be insoluble and could not be isolated, seven of which contained a mutation in the BCDG sheet. We observed that mutations in CP-CD are less destabilizing on average than equivalent mutations in wild-type, except V10A (A-strand) and V143A (H-strand), which are destabilized by a similar amount (Table S3). A comparison with wild-type shows that high  $\Phi$ -values in the A- and H-strands are maintained in CP-CD (Fig. 4). For CP-DE, 12 variant proteins were sufficiently destabilized for  $\Phi$ -value analysis; only two mutations, located in strands D and H, resulted in insoluble protein, which could not be isolated. Three variants with mutations located in the D, F, and G strands (L70A, Y103A, and W111L, respectively) are destabilized to a greater extent than in wild-type; however, mutations in the A, B, and H strands have a similar effect on stability (Table S4). CP-DE also displays a pattern of  $\Phi$ -values similar to that of wild-type, with high values in the A- and H-strands (Fig. 4). The most structured region of the transition state in both CPs is maintained in the FEHA sheet and particularly around the terminal strands A and H, as in wild-type MATH.

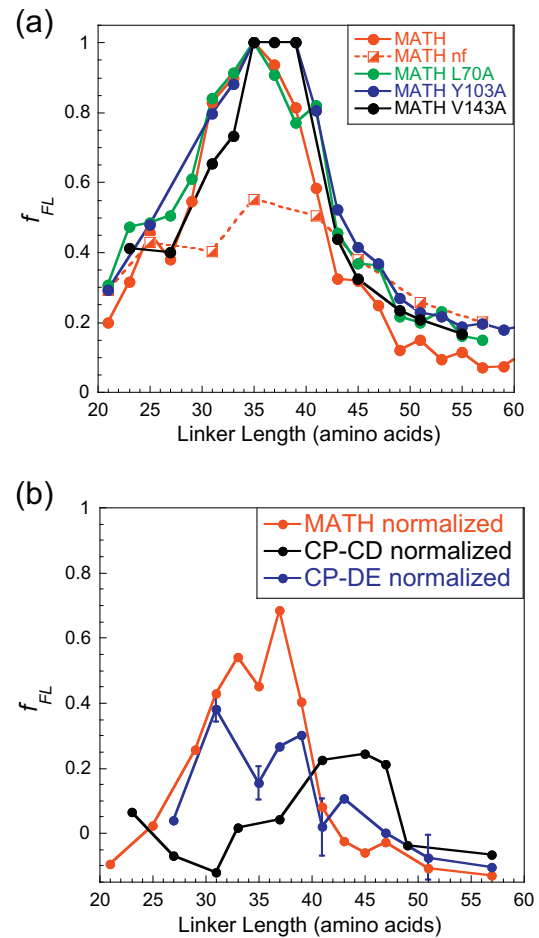
### MATH folds close to the ribosome

Can cotranslational folding begin before the entire MATH domain is fully emerged from the ribosome? To investigate this, we used an *in vitro* AP assay, as described in Refs. [14,17] to determine when wild-type MATH can commence folding on the



**Fig. 4.** Comparison of  $\Phi$ -values for MATH wild-type, CP-CD, and CP-DE. (a) Histograms of  $\Phi$ -values for MATH wild-type (top), CP-CD (center), and CP-DE (bottom). (b) A plot of  $\Phi$ -values for MATH wild-type *versus*  $\Phi$ -values for CPs showing a linear fit of the data, which gives an  $R$ -value of 0.81.

ribosome (Fig. 5a). In the study we report here, we use this assay to determine when the MATH domain can fold productively by assessing from the force profile at what linker length a force is generated sufficient to release the stall. In addition to obtaining force-profile results for wild-type MATH, we also performed arrest peptide experiments on a non-folding (nf) variant of MATH (W16G/W47G/F84E) at 14 different linker lengths; this is to ensure that fraction full-length values ( $f_{FL}$ ) obtained are due to a folding event and not due to non-specific interactions with the ribosome tunnel. Previous studies have estimated that the ribosome tunnel can sequester in the range of 30–40 amino



**Fig. 5.** Following folding of MATH on the ribosome using the arrest peptide assay. (a) Force profiles of wild-type MATH with non-folding controls MATH (W16G/W47G/F84E) (red) and three variants MATH L70A (green), Y103A (blue), and V143A (black). The small differences between the four force profiles reflect the reproducibility expected from such assays as reported previously [17]. (b) Force profiles of CP-CD and CP-DE compared with wild-type MATH normalized against non-folding controls. Data points shown at 14 linker lengths for wild-type MATH (red) are normalized against the nf control MATH (W16G/W47G/F84E); data shown at 10 linker lengths for CP-CD (black) are normalized against the nf control CP-CD (W16G/W47G/F84E); data shown at 10 linker lengths for CP-DE (blue) are normalized against the nf control CP-DE (W16G/W47G/F84E). Error bars are shown at linker lengths for CP-DE where experiments (sample plus nf control) were performed in triplicate.

acids [12,55,56]. Our results show that MATH can release the stall at linker length ( $L$ ) = 31 with a force peak extending to ( $L$ ) = 39 (Fig. 5a), suggesting that MATH can fold cotranslationally when a proportion of C-terminal (H-strand) residues are still located in the ribosome tunnel. Three variants of wild-type MATH were also employed in the AP assay, each destabilized by 2–3 kcal mol<sup>-1</sup> relative to wild-type (Table S1): L70A (D-strand) and

Y103A (F-strand) are centrally located, and V143A is located in the C-terminal H-strand. All three variants exhibit a force-measurement profile similar to wild-type (Fig. 5a) showing that a reduction in intrinsic stability of this magnitude does not influence the cotranslational folding of MATH close to the ribosome. (We define “intrinsic” as relating to the properties of the native domain in isolation).

### Onset of folding does not correlate with location of strands A and H in the ribosome tunnel

As residues with the highest  $\Phi$ -values are found in the terminal strands A and H in wild-type MATH, CP-CD, and CP-DE, we ask: does the location of these strands in the ribosome tunnel influence the onset of cotranslational folding? To investigate this, we employed CP-CD and CP-DE in the AP assay. Very importantly, each of these CPs, although precisely the same length as wild-type MATH (see Fig. S5a), with precisely the same leader, variable linker and SecM sequence, will have different residues interacting with the ribosome tunnel at any given linker length; to control for this, we made non-folding mutants of each CP as well; thus, in the results, the data for each CP is shown normalized against *its own* non-folding variant (Fig. 5b).

CP-CD has new termini located in the middle of the BCDG sheet, central to the structure, and is destabilized by approximately 2 kcal mol<sup>-1</sup> compared with the isolated wild-type domain (Table S2); the H- and A-strands will emerge from the ribosome after strands D, E, F, and G, but before strands B and C (Fig. S5b). Our data show that the force profile of CP-CD results in a later force peak compared with wild-type (Fig. 5b). We infer that for this permutant, folding does not occur until most or all of the domain has emerged from the ribosome tunnel.

The circular permutant CP-DE has new termini located between the two  $\beta$ -sheets and has a similar stability to wild-type (Table S2). In this permutant, strands H and A are located in the center of the protein (Fig. S5). Force-profile data for CP-DE show that folding occurs at a similar linker length to wild-type,  $L = 31$ , and also like wild-type, the force peak extends to  $L = 39$  (Fig. 5b).

Taken together, our data show that the earlier emergence of the H–A-strand pair during translation does not result in earlier onset of productive cotranslational folding in the CPs.

## Discussion

We use a MATH domain and two CPs to investigate how alterations in chain connectivity affect the folding pathway and the onset of folding on the ribosome. The generation of CPs allows the

investigation of the folding of MATH variants with identical amino acid sequences, but different chain connectivity and order of secondary structural elements.

Our  $\Phi$ -value analysis of MATH wild-type shows that folding proceeds by a nucleation-condensation mechanism, in which secondary and tertiary structure form concomitantly. Previous experimental and simulation studies suggest that a folding nucleus can be subdivided into “obligate” and “critical” nuclei [57,58]: Residues that constitute the obligate nucleus interact early to establish the correct topology of the protein; these in turn are supported by a shell of additional interactions, which constitute the critical nucleus, turning the free energy profile downhill.  $\Phi$ -values report upon the extent of structure formation in the transition state, but not the order in which that structure is formed in that nucleus [59]. Our data are consistent with the A-strand being almost fully structured in the transition state, and thus, we infer, fully packed onto the terminal H strand (Fig. 1). However, it seems unlikely that interaction of the N and C terminal strands alone would be sufficient to set up the complex topology of the MATH domain fold. Instead we propose that residues with relatively high  $\Phi$ -values in strands C, D, E, and F interact to form an obligate folding nucleus, setting up the complex topology of the domain and establishing the correct register of the strands and the packing of the two sheets. This nucleus would then be further stabilized by the packing of the peripheral strands A and H (which in this case constitute the critical nucleus). In this model, formation of interactions between residues in strands A and H is energetically critical—the N and C terminal regions of the protein must come together and pack to provide sufficient energy to cross the transition state barrier, resulting in high  $\Phi$ -values in this region of the protein.

There have been a number of studies of Greek-key proteins that have given support to an obligate-critical nucleus model such as that we propose here: The Ig domain titin I27 (I27) and the evolutionarily unrelated fnIII domains, TNfn3, FNfn10, and CAfn2 share a structurally equivalent obligate nucleus, comprising key residues located in the four central strands, which establishes the topology of these complex Greek-key domains [60–64]. Similar results were seen for the all-beta Greek key protein S6, and for an all-alpha Greek key death domain [35,43].

Upon circular permutation of MATH, the terminal A- and H-strands are covalently linked and become adjacent in the primary sequence. We thus predicted that connection of the terminal strands and generation of new N and C termini in loop regions would increase the folding rate of the CPs due to a reduction in the entropic cost of bringing the terminal strands together during folding [65]. Indeed we find that all six CPs exhibited an increased rate of both folding and unfolding, indicating that the energetic

cost of forming the transition state is lower in the permutants. In most previous studies, circular permutation has been shown to alter the folding pathway of a protein [42–44] such as in CPs of S6 [30] and the SH3 domain of  $\alpha$ -spectrin [66], which have a shifted folding nucleus upon changes in connectivity. However, the pattern of  $\Phi$ -values seen in wild-type MATH is essentially maintained in CP-CD and CP-DE (Fig. 4): our data are consistent with the hypothesis that these permutants follow a similar folding pathway to wild-type.

The question of when cotranslational folding can begin on the ribosome is of particular interest in the case of MATH: Due to the vectorial synthesis of the peptide chain, the A-strand will emerge first in wild-type and be outside the ribosome, while the H-strand is still sequestered in the tunnel. Our  $\Phi$ -value analysis shows that interactions between strands H and A are energetically critical for folding. As shown in Fig. 1, the A-strand cannot pack onto the protein until the H-strand is formed, so we expected that the entire domain of MATH would need to be fully emerged before folding can occur. However, wild-type MATH and three destabilized variants (Fig. 5a) can fold cotranslationally, while a proportion of C-terminal residues are still sequestered in the ribosome.

We also predicted that CP-CD and CP-DE may fold earlier on the ribosome compared with wild-type, as the H-strand will emerge earlier from the tunnel (Fig. S5). The onset of the folding is revealed in the arrest peptide assay, and we see that neither permutant folds earlier than wild-type in our cotranslational folding assay (Fig. 5b). For CP-DE, folding occurs at the same linker length as wild-type, whereas CP-CD can fold only when fully emerged from the ribosome. We can rule out loss of intrinsic stability as being key to the failure of CP-CD to fold close to the ribosome, as it has a similar stability to the wild-type variants Y103A and V143A.

It appears that in the case of MATH, cotranslational folding is not influenced by formation of interactions between strands A and H. Instead we propose that formation of both  $\beta$ -sheets, critical to formation of the obligate nucleus, is required for formation of sufficiently stable structures to be able to fold close to the ribosome. In CP-DE, connectivity within each sheet is maintained as the new termini are created within a cross-sheet loop. However, in CP-CD, the new termini created in strands C and D disrupt the connectivity of the BCDG sheet. We infer that the BCDG sheet in CP-CD is now too destabilized to contribute to formation of the obligate nucleus, until the protein is fully emerged from the ribosome.

We have shown that chain connectivity can influence the onset of productive cotranslational folding of a MATH domain on the ribosome, but not in a way we initially predicted. In the CPs, the earlier

emergence of residues which are the most highly structured in the rate-limiting transition state, in the isolated domains, does not result in folding occurring earlier on the ribosome. Taking all our data together, we propose that stable formation of an obligate nucleus (likely similar to that seen in our folding studies of isolated domains and involving strands C, D, E, and F) is required for the folding of MATH to be productive on the ribosome. Small all  $\beta$ -domains often form part of larger multi-domain proteins; we speculate that MATH is unable to fold on the ribosome until both  $\beta$ -sheets are formed and able to pack, to prevent the accumulation of partly folded structures which could lead to misfolding or aggregation.

## Materials and Methods

### DNA manipulation for expression of isolated domains

MATH wild-type (PBD:2CR2 residues 4–153) and circularly permuted constructs were generated as synthetic genes (Life Technologies) and cloned into pRSET A plasmid (Invitrogen) (previously modified to remove the sequence including the entire T7 gene 10 leader and EK recognition site up to, but not including, the BamH I site and replaced with a sequence encoding residues GLVPRGS).

CPs are given the nomenclature CP-XX, where XX denotes the position of the new termini. In each case, the residues DSGK (shown in the wild-type sequence below in bold) were used to form a linker connecting the original N and C-termini, and new termini are generated in a loop region. For the isolated constructs of all CPs, a glycine followed by residues SVN (shown underlined in the wild-type sequence below) were added to the newly generated C-terminus, as this was found to improve solubility. For CP-DE, an additional C73A mutation was made at the new C-terminus to avoid a free Cys residue. The amino acid sequence of MATH wild-type is as follows:

**GSSGK**VVKFSYMWTINNFSFCREEMGE  
VKSSTFSSGANDKLKWCLRVNPKGLDEESK  
DYLSLYLLLVSCKSEVRAKFKFSILNAKGEETKA  
MESQRAYRFVQGGKDWGFKKFIIRD<sup>FL</sup>L  
DEANGLLPDDKLT<sup>FL</sup>FC<sup>FL</sup>EVSVVQDS<sup>FL</sup>VN<sup>FL</sup>ISGQ.

### Protein expression and purification

Protein was expressed using *Escherichia coli* C41 in 2 × YT containing 100  $\mu\text{g mL}^{-1}$  IPTG at 37 °C with shaking. Cells were grown overnight at 22 °C and then harvested by centrifugation at 6000 rpm (Sorvall RC-5C). The cell pellets were resuspended in 1 × PBS (pH 7.4) and 2 mM  $\beta$ -mercaptoethanol

and lysed by sonication. Cell lysate was cleared by centrifugation at 18,000 rpm (Sorvall RC-5C), and the soluble protein supernatant was bound to Ni<sup>2+</sup> agarose resin (Agarose Bead Technologies). Resin was then washed four times in 1 × PBS (pH 7.4) and then resuspended in 50 mM Tris (pH 8.2), and protein was cleaved from the resin overnight at 25 °C using thrombin (Sigma). Protein was further purified by gel filtration chromatography using a HiLoad Superdex G75 column (GE Healthcare) equilibrated in 1 × PBS (pH 7.4).

### Equilibrium denaturation and kinetics studies

Stability of proteins was determined by GdmCl denaturation in 1 × PBS (pH 7.4) containing up to 10 mM DTT at 25 °C using the intrinsic fluorescence of tryptophan after excitation at 280 nm. Emission spectra between 300 and 400 nm were recorded using a PerkinElmer LS-55 or a Cary Eclipse fluorimeter. Final protein concentration was 1 μM. The kinetics of protein folding and unfolding was measured in GdmCl in 1 × PBS (pH 7.4) containing up to 10 mM DTT at 25 °C using intrinsic tryptophan fluorescence. Proteins were excited at 280 nm using an Applied Photophysics SX.18 or SX.20 stopped-flow spectrophotometer, and the fluorescence emission was collected above 320 nm. The final protein concentration was 1 μM, and six traces were averaged for each concentration of denaturant. Final protein concentrations of 0.1, 1.0, and 2 μM for MATH wild-type and 0.1, 1.0, and 6.5 μM for CP-DE were used to demonstrate protein concentration-independent rollover below 2 M GdmCl. Data were fitted to a single-exponential model using KaleidaGraph (Synergy Software). For analysis of kinetic data, see Supplementary Information.

### DNA manipulations for AP assays

A plasmid was generated containing the *E. coli* SecM arrest peptide (FSTPVWISQAQGIRAGP) [67,68] and a terminal Lep segment (23 amino acids) derived from leader peptidase [46] in a modified version of pRSET A (Invitrogen). Increasing linker lengths (from  $L = 21$  to  $L = 61$  in increments of 2 amino acids) were generated in the vector using site-directed mutagenesis. DNA templates for *in vitro* transcription–translation were prepared using the In-Fusion cloning system (Takara Bio). Genes of interest were amplified by PCR with overhanging homology to the vector and subsequently cloned into pRSET A containing the appropriate linker length, using the In-Fusion system. All constructs were verified by DNA sequencing. For MATH wild-type site-directed mutagenesis was performed to generate non-folding triple mutant controls (W16G/W47G/F84E) at 14 linker lengths ( $L = 21, 25, 29–47, 51,$  and 57). For MATH CPs, site-directed mutagenesis was performed to generate non-folding triple-mutant

controls (W16G/W47G/F84E) at 10 linker lengths for CP-CD ( $L = 23, 27, 31, 33, 37, 41, 45, 47, 49,$  and 57) and for CP-DE ( $L = 27, 31, 35–43, 47, 51,$  and 57). Site-directed mutagenesis was also performed on wild-type MATH to generate constructs with the non-functional FSTPVWISQAQGIRAGA arrest peptide (P to A mutation underlined) as full-length controls; also arrested controls were generated by insertion of a stop codon directly after the arrest peptide. Protein sequences of constructs used in the AP assays for MATH wild-type, CP-DE, and CP-CD ( $L = 61$ ) are shown in Supplementary Information (Fig. S5a).

### *In vitro* transcription and translation

Transcription and translation were performed using the PUREflex cell-free translation system according to the manufacturer's instructions (GeneFrontier Corporation) [69,70]. <sup>35</sup>S-Methionine (PerkinElmer) was added to a master mix containing all components of the PUREflex system and subsequently added to pre-prepared samples of template DNA (250 μg). Synthesis of labeled polypeptides was performed for 15 min at 37 °C with shaking at 500 rpm. Reactions were quenched using an equal volume of ice-cold 10% (v/v) trichloroacetic acid on ice for 30 min. Labeled polypeptides were pelleted by centrifugation at 14,440g for 5 min at 4 °C. The supernatant was discarded, and the pellet was resuspended in 1 × LDS buffer (Life Technologies) with shaking at 14,000 rpm, 15 min at 37 °C. RNA was digested with RNase A (ThermoFisher Scientific) with shaking at 700 rpm for 15 min at 37 °C. Samples were resolved by SDS-PAGE and fixed for 1 h using Instant Blue protein stain (Expedeon), and gels were dried for 2 h. Images were transferred to a phosphor screen for 2 days before imaging. Bands were quantified using ImageJ (<http://rsb.ingo.nih.gov/ij/>) and fit to a Gaussian distribution (Fig. S1c) using KaleidaGraph (Synergy Software). To obtain normalized fraction full-length values ( $f_{FL}$ ), the  $f_{FL}$  value for the nf control was subtracted from the  $f_{FL}$  value for the equivalent sample. For MATH wild-type, normalized data points in the force-measurement profiles are single experiments; for CP-CD, normalized data points are from a single experiment except for  $L = 41$  and 45, which are an average of two experiments. For CP-DE, normalized data points are from a single experiment except  $L = 27$ , which is an average of two experiments, and  $L = 31, 35, 41,$  and 51, which are an average of at least three experiments. For MATH wild-type variants L70A, Y103A, and V143A, data points are not normalized and are from a single experiment. The reproducibility of data obtained from arrest peptide force measurement assays has been discussed previously [17].

## Acknowledgments

All authors are supported by a grant from the Wellcome Trust (WT095195) except A.P.M., P.R., and C.A.T.F.M. J.C. is a Wellcome Trust Senior Research Fellow. A.P.M. is supported by the EPSRC. P.R. is supported by an Erasmus Plus Programme through Hochschule Bonn-Rhein-Sieg University, Germany. C.A.T.F.M. is supported by the Cambridge Trust and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) Science Without Borders Cambridge Scholarship. Plasmids used in this study for the arrest-peptide-based force-measurement assays were based on those developed by Professor Gunnar von Heijne, and initial constructs were a kind gift from this laboratory.

Author contributions: A.P.M. and S.H. characterized and purified the wild-type protein; C.O. and P.R. characterized and purified the CPs; J.J.H., A.P.M., A.S., and C.M. performed the arrest-peptide-based force-measurement assays; T.O.K. and L.G.K. provided assistance with molecular biology; A.S. and J.J.H. wrote the manuscript; J.C. conceived and planned the investigation and wrote the manuscript.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jmb.2018.10.011>.

Received 6 June 2018;

Received in revised form 11 September 2018;

Accepted 16 October 2018

Available online 23 October 2018

### Keywords:

protein folding;  
folding on the ribosome;  
SecM;  
immunoglobulin-like;  
Phi value

†A.P.M. and J.J.H. contributed equally to this work.

### Abbreviations used:

GdmCl, guanidinium chloride; CPs, circular permutants; AP assay, arrest-peptide-based force-measurement assay.

## References

[1] M. Thommen, W. Holtkamp, M.V. Rodnina, Co-translational protein folding: progress and methods, *Curr. Opin. Struct. Biol.* 42 (2017) 83–89.

- [2] A.A. Komar, Unraveling co-translational protein folding: concepts and methods, *Methods* 137 (2018) 71–81.
- [3] A.V. Nicola, W. Chen, A. Helenius, Co-translational folding of an alphavirus capsid protein in the cytosol of living cells, *Nat. Cell Biol.* 1 (1999) 341–345.
- [4] M.S. Evans, I.M. Sander, P.L. Clark, Cotranslational folding promotes  $\beta$ -helix formation and avoids aggregation in vivo, *J. Mol. Biol.* 383 (2008) 683–692.
- [5] Y. Han, A. David, B. Liu, J.G. Magadán, J.R. Bennick, J.W. Yewdell, S. Qian, Monitoring cotranslational protein folding in mammalian cells at codon resolution, *Proc. Natl. Acad. Sci. U. S. A.* 109 (2012) 12467–12472.
- [6] I.M. Sander, J.L. Chaney, P.L. Clark, Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design, *J. Am. Chem. Soc.* 136 (2014) 858–861.
- [7] S.J. Kim, J.S. Yoon, H. Shishido, Z. Yang, L.A. Rooney, J.M. Barral, W.R. Skach, Translational tuning optimizes nascent protein folding in cells, *Science* 348 (2015) 444–448.
- [8] K.S. Hingorani, L.M. Gierasch, Comparing protein folding in vitro: foldability meets the fitness challenge, *Curr. Opin. Struct. Biol.* 24 (2014) 81–90.
- [9] F. Buhr, S. Jha, M. Thommen, J. Mittelstaet, F. Kutz, H. Schwalbe, M.V. Rodnina, A.A. Komar, Synonymous codons direct cotranslational folding toward different protein conformations, *Mol. Cell* 61 (2016) 341–351.
- [10] P. Ciryam, R.I. Morimoto, M. Vendruscolo, C.M. Dobson, E.P. O'Brien, In vivo translation rates can substantially delay the cotranslational folding of the *Escherichia coli* cytosolic proteome, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) E132–E140.
- [11] C. Eichmann, S. Preissler, R. Riek, E. Deuerling, Cotranslational structure acquisition of nascent polypeptides monitored by NMR spectroscopy, *Proc. Natl. Acad. Sci. U. S. A.* 107 (2010) 9111–9116.
- [12] L.D. Cabrera, A.M.E. Cassaignau, H.M.M. Launay, C.A. Waudby, T. Wlodarski, C. Camilloni, M. Karyadi, A.L. Robertson, X. Wang, A.S. Wentink, L.S. Goodsell, C.A. Woolhead, M. Vendruscolo, C.M. Dobson, J. Christodoulou, A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding, *Nat. Struct. Mol. Biol.* 23 (2016) 278–285.
- [13] L. Tu, C. Deutsch, Determinants of helix formation for a Kv1.3 transmembrane segment inside the ribosome exit tunnel, *J. Mol. Biol.* 429 (2017) 1722–1732.
- [14] O.B. Nilsson, R. Hedman, J. Marino, S. Wickles, L. Bischoff, M. Johansson, A. Müller-Lucks, F. Trovato, J.D. Puglisi, E.P. O'Brien, R. Beckmann, G. von Heijne, Cotranslational protein folding inside the ribosome exit tunnel, *Cell Rep.* 12 (2015) 1533–1540.
- [15] T. Su, J. Cheng, D. Sohmen, R. Hedman, O. Berninghausen, G. von Heijne, D.N. Wilson, R. Beckmann, The force-sensing peptide VemP employs extreme compaction and secondary structure formation to induce ribosomal stalling, *elife* 6 (2017) e25642.
- [16] E. Mercier, M.V. Rodnina, Co-translational folding trajectory of the HemK helical domain, *Biochemistry* 57 (2018) 3460–3464.
- [17] O.B. Nilsson, A.A. Nickson, J.J. Hollins, S. Wickles, A. Steward, R. Beckmann, G. von Heijne, J. Clarke, Cotranslational folding of spectrin domains via partially structured states, *Nat. Struct. Mol. Biol.* 24 (2017) 221–226.
- [18] P. Tian, A. Steward, R. Kudva, T. Su, P.J. Shilling, A.A. Nickson, J.J. Hollins, R. Beckmann, G. von Heijne, J. Clarke, R.B. Best, The folding pathway of an Ig domain is conserved on and off the ribosome, *Proc. Natl. Acad. Sci. U. S. A.* (2018) <https://doi.org/10.1073/pnas.1810523115> (in Press).



- [19] C.M. Kaiser, D.H. Goldman, J.D. Chodera, I. Tinoco Jr., C. Bustamante, The ribosome modulates nascent protein folding, *Science* 334 (2011) 1723–1727.
- [20] A.J. Samelson, M.K. Jensen, R.A. Soto, J.H.D. Cate, S. Marqusee, Quantitative determination of ribosome nascent chain stability, *Proc. Natl. Acad. Sci. U. S. A.* 113 (2016) 13402–13407.
- [21] A.M. Knight, P.H. Culviner, N. Kurt-Yilmaz, T. Zou, S.B. Ozkan, S. Cavagnero, Electrostatic effect of the ribosomal surface on nascent polypeptide dynamics, *ACS Chem. Biol.* 8 (2013) 1195–1204.
- [22] E.J. Guinn, P. Tian, M. Shin, R.B. Best, S.M. Marqusee, A small single-domain protein folds through the same pathway on- and off- the ribosome, *Biorxiv* (2018) <https://doi.org/10.1101/347864>.
- [23] W. Holtkamp, G. Kokic, M. Jager, J. Mittelstaet, A.A. Komar, M.V. Rodnina, Cotranslational protein folding on the ribosome monitored in real time, *Science* 350 (2015) 1104–1107.
- [24] K.G. Ugrinov, P.L. Clark, Cotranslational folding increases GFP folding yield, *Biophys. J.* 98 (2010) 1312–1320.
- [25] M.S. Evans, T.F. Clarke, P.L. Clark, Conformations of co-translational folding intermediates, *Protein Pept. Lett.* 12 (2005) 189–195.
- [26] F. Glöge, A.H. Becker, G. Kramer, B. Bukau, Co-translational mechanisms of protein maturation, *Curr. Opin. Struct. Biol.* 24 (2014) 24–33.
- [27] S. Kirchner, Z. Cai, R. Rauscher, N. Kastelic, M. Anding, A. Czech, B. Kleizen, L.S. Ostedgaard, I. Braakman, D.N. Sheppard, Z. Ignatova, Alteration of protein function by a silent polymorphism linked to tRNA abundance, *PLoS Biol.* 15 (2017), e2000779.
- [28] J.L. Chaney, A. Steele, R. Carmichael, A. Rodriguez, A.T. Specht, K. Ngo, J. Li, S. Emrich, P.L. Clark, Widespread position-specific conservation of synonymous rare codons within coding sequences, *PLoS Comput. Biol.* 13 (2017), e1005531.
- [29] F. Wruck, A. Katranidis, K.H. Nierhaus, G. Büldt, M. Hegner, Translation and folding of single proteins in real time, *Proc. Natl. Acad. Sci. U. S. A.* 114 (2017) E4399–E4407.
- [30] E. Haglund, M. Lindberg, M. Oliveberg, Changes of protein folding pathways by circular permutation: overlapping nuclei promote global cooperativity, *J. Biol. Chem.* 283 (2008) 27904–27915.
- [31] D.T. Capraro, S. Gosavi, M. Roy, J.N. Onuchic, P.A. Jennings, Folding circular permutants of IL-1 $\beta$ : route selection driven by functional frustration, *PLoS One* 7 (2012), e38512.
- [32] G. Hultqvist, S.R. Haq, A.S. Punekar, C.N. Chi, Å. Engström, A. Bach, K. Strømgaard, M. Selmer, S. Gianni, P. Jemth, Energetic pathway sampling in a protein interaction domain, *Structure* 21 (2013) 1193–1202.
- [33] B.L. Kier, J.M. Anderson, N.H. Anderson, Circular permutation of a WW domain: folding still occurs after excising the turn of the folding-nucleating hairpin, *J. Am. Chem. Soc.* 136 (2015) 741–749.
- [34] R.P. Nobrega, K. Arora, S.V. Kathuria, R. Graceffa, R.A. Barrea, L. Guo, S. Chakravarthy, O. Bilsel, T.C. Irving, C.L. Brooks III, C.R. Matthews, Modulation of frustration in folding by sequence permutation, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 10562–10567.
- [35] K.R. Kemplen, D. De Sancho, J. Clarke, The response of Greek key proteins to changes in connectivity depends on the nature of their secondary structure, *J. Mol. Biol.* 427 (2015) 2159–2165.
- [36] Y. Lindqvist, G. Schneider, Circular permutations of natural protein sequences: structural evidence, *Curr. Opin. Struct. Biol.* 7 (1997) 422–427.
- [37] J. Cellitti, M. Llinas, N. Echols, E.A. Shank, B. Gillespie, E. Kwon, S.M. Crowder, F.W. Dahlquist, T. Alber, S. Marqusee, Exploring subdomain cooperativity in T4 lysosyme I: structural and energetic studies of a circular permutant and protein fragment, *Protein Sci.* 16 (2007) 842–851.
- [38] J.S. Butler, D.M. Mitrea, G. Mitrousis, G. Cingolani, S.N. Loh, Structural and thermodynamic analysis of a conformationally strained circular permutant of barnase, *Biochemistry* 48 (2009) 3497–3507.
- [39] C. Chen, J. Yun, J. Kim, C. Park, Effect of circular permutations on transient partial unfolding in proteins, *Protein Sci.* 25 (2016) 1483–1491.
- [40] F. Li, Z. Luan, Q. Chen, J. Xu, H. Yu, Rational selection of circular permutation sites in characteristic regions of the  $\alpha/\beta$ -hydrolase fold enzyme *RhEst1*, *J. Mol. Catal. B Enzym.* 125 (2016) 75–80.
- [41] Y. Yu, I.D. Petrik, K.N. Chacón, P. Hosseinzadeh, H. Chen, N.J. Blackburn, Y. Lu, Effect of circular permutation on the structure and function of type 1 blue copper center in azurin, *Protein Sci.* 26 (2017) 218–226.
- [42] E.A. Shank, C. Cecconi, J.W. Dill, S. Marqusee, C. Bustamante, The folding cooperativity of a protein is controlled by its chain topology, *Nature* 465 (2010) 637–640.
- [43] I.A. Hubner, M. Lindberg, E. Haglund, M. Oliveberg, E.I. Shakhnovich, Common motifs and topological effects in the protein folding transition state, *J. Mol. Biol.* 359 (2006) 1075–1085.
- [44] M. Lindberg, J. Tangrot, M. Oliveberg, Complete change of the protein folding transition state upon circular permutation, *Nat. Struct. Biol.* 9 (2002) 818–822.
- [45] A.G. Uren, D.L. Vaux, TRAF proteins and mepriins share a conserved domain, *Trends Biochem. Sci.* 21 (1996) 244–245.
- [46] N. Ismail, R. Hedman, N. Schiller, G. von Heijne, A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration, *Nat. Struct. Mol. Biol.* 19 (2012) 1018–1022.
- [47] F. Cymer, G. von Heijne, Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 14640–14645.
- [48] J.A. Farias-Rico, S.K. Goetz, J. Marino, G. von Heijne, Mutational analysis of protein folding inside the ribosome exit tunnel, *FEBS Lett.* 591 (2017) 155–163.
- [49] K. Ito, S. Chiba, Arrest peptides: cis-acting modulators of translation, *Annu. Rev. Biochem.* 82 (2013) 171–202.
- [50] D.N. Wilson, S. Arenz, R. Beckmann, Translation regulation via nascent polypeptide-mediated ribosome stalling, *Curr. Opin. Struct. Biol.* 37 (2016) 123–133.
- [51] D.H. Goldman, C.M. Kaiser, A. Milin, M. Righini, I. Tinoco Jr., C. Bustamante, Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo, *Science* 348 (2015) 457–460.
- [52] M.E. Butkus, L.B. Prundeanu, D.B. Oliver, Translocon “pulling” of nascent SecM controls the duration of its translational pause and secretion-responsive *secA* regulation, *J. Bacteriol.* 185 (2003) 6719–6722.
- [53] A. Bachmann, T. Kiefhaber, Apparent two-state tendamistat folding is a sequential process along a defined route, *J. Mol. Biol.* 306 (2001) 375–386.
- [54] I.E. Sánchez, T. Kiefhaber, Origin of unusual  $\Phi$ -values in protein folding: evidence against specific nucleation sites, *J. Mol. Biol.* 334 (2003) 1077–1085.

- [55] N.R. Voss, M. Gerstein, T.A. Steitz, P.B. Moore, The geometry of the ribosomal polypeptide exit tunnel, *J. Mol. Biol.* 360 (2006) 893–906.
- [56] S. Bhushan, T. Hoffmann, B. Seidelt, J. Frauenfeld, T. Mielke, O. Berninghausen, D.N. Wilson, R. Beckmann, SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center, *PLoS Biol.* 9 (2011), e1000581.
- [57] T. Ternström, U. Mayor, M. Akke, M. Oliveberg, From snapshot to movie:  $\Phi$  analysis of protein folding transition states taken one step further, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 14854–14859.
- [58] E.I. Shakhnovich, Folding nucleus: specific or multiple? Insights from lattice models and experiments, *Fold. Des.* 3 (1998) R108–R111.
- [59] A.R. Fersht, S. Sato,  $\Phi$ -value analysis and the nature of protein-folding transition states, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004) 7976–7981.
- [60] S.B. Fowler, J. Clarke, Mapping the folding pathway of an immunoglobulin domain: structural detail from phi value analysis and movement of the transition state, *Structure* 9 (2001) 355–366.
- [61] C.D. Geierhaas, E. Paci, M. Vendruscolo, J. Clarke, Comparison of the transition states for folding of two Ig-like proteins from different superfamilies, *J. Mol. Biol.* 343 (2004) 1111–1123.
- [62] E. Cota, A. Steward, S.B. Fowler, J. Clarke, The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold, *J. Mol. Biol.* 305 (2001) 1185–1194.
- [63] A.A. Nickson, B.G. Wensley, J. Clarke, Take home lessons from studies of related proteins, *Curr. Opin. Struct. Biol.* 23 (2013) 66–74.
- [64] I. Lappalainen, M.G. Hurley, J. Clarke, Plasticity within the obligatory nucleus of an immunoglobulin-like domain, *J. Mol. Biol.* 375 (2008) 547–559.
- [65] A.A. Nickson, K.E. Stoll, J. Clarke, Folding of a LysM domain: entropy-enthalpy compensation in the transition state of an ideal two-state folder, *J. Mol. Biol.* 380 (2008) 557–569.
- [66] J.C. Martinez, A.R. Viguera, R. Berisio, M. Wilmanns, P.L. Mateo, V.V. Filimonov, L. Serrano, Thermodynamic analysis of alpha-spectrin SH3 and two of its circular permutants with different loop lengths: discerning the reasons for rapid folding in proteins, *Biochemistry* 38 (1999) 549–559.
- [67] H. Nakatogawa, A. Murakami, I. Koreaki, Control of SecA and SecM translation by protein secretion, *Curr. Opin. Struct. Biol.* 7 (2004) 145–150.
- [68] M.N. Yap, H.D. Bernstein, The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel, *Mol. Cell* 34 (2009) 201–211.
- [69] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, Cell-free translation reconstituted with purified components, *Nat. Biotechnol.* 19 (2001) 751–755.
- [70] Y. Shimizu, T. Kanamori, T. Ueda, Protein synthesis by pure translation systems, *Methods* 36 (2005) 299–304.