

Cite this: *Chem. Sci.*, 2020, **11**, 12512

All publication charges for this article have been paid for by the Royal Society of Chemistry

# Mechanically tightening, untying and retying a protein trefoil knot by single-molecule force spectroscopy†

Han Wang  and Hongbin Li \*

Knotted conformation is one of the most surprising topological features found in proteins, and understanding the folding mechanism of such knotted proteins remains a challenge. Here, we used optical tweezers (OT) to investigate the mechanical unfolding and folding behavior of a knotted protein *Escherichia coli* tRNA (guanosine-1) methyltransferase (TrmD). We found that when stretched from its N- and C-termini, TrmD can be mechanically unfolded and stretched into a tightened trefoil knot, which is composed of *ca.* 17 residues. Stretching of the unfolded TrmD involved a compaction process of the trefoil knot at low forces. The unfolding pathways of the TrmD were bifurcated, involving two-state and three-state pathways. Upon relaxation, the tightened trefoil knot loosened up first, leading to the expansion of the knot, and the unfolded TrmD can then fold back to its native state efficiently. By using an engineered truncation TrmD variant, we stretched TrmD along a pulling direction to allow us to mechanically unfold TrmD and untie the trefoil knot. We found that the folding of TrmD from its unfolded polypeptide without the knot is significantly slower. The knotting is the rate-limiting step of the folding of TrmD. Our results highlighted the critical importance of the knot conformation for the folding and stability of TrmD, offering a new perspective to understand the role of the trefoil knot in the biological function of TrmD.

Received 16th May 2020  
Accepted 16th October 2020

DOI: 10.1039/d0sc02796k

rsc.li/chemical-science

## Introduction

A protein knot is one of the most remarkable features found in proteins over the last two decades, and it adds a layer of topological complexity to the protein folding problem.<sup>1–7</sup> More than 1300 proteins with knotted or slipknotted conformations have been identified from proteins whose three-dimensional structures are deposited in Protein Data Bank (PDB), including trefoil (3<sub>1</sub>), figure-of-eight (4<sub>1</sub>), Gordian (5<sub>2</sub>) and stevedore (6<sub>1</sub>) knots.<sup>5,8,9</sup> These knots in proteins are believed to be functionally relevant, as well as provide extra structural stability to proteins.<sup>9–12</sup> Understanding the molecular mechanism *via* which knotted proteins overcome the topological barriers to fold represents a significant challenge.<sup>13</sup>

Both experimental,<sup>8,14–18</sup> including ensemble and single-molecule measurements, and computational studies<sup>19–23</sup> have started to provide invaluable mechanistic insights into the folding of knotted/slipknotted proteins. It has been shown that such knotted proteins are able to overcome the high topological barrier to knot themselves and fold, although generally very slowly.<sup>8,15,18,24,25</sup> Molecular dynamics (MD) simulations are of

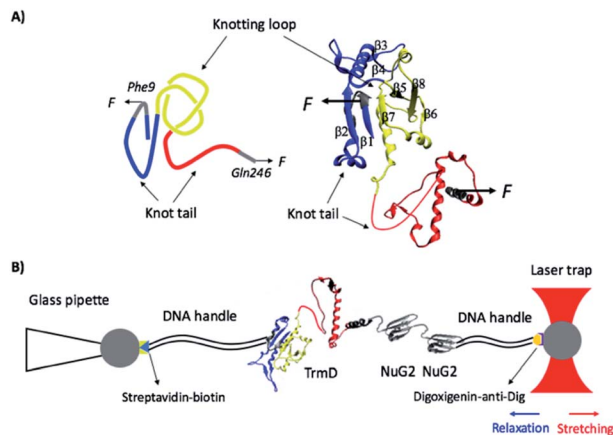
great importance to understanding the molecular mechanism of the folding of these proteins. These studies have revealed the complexity of the folding of knotted proteins and proposed several possible knotting mechanisms, including threading, slipknotting, mouse trapping and folding on ribosomes, to offer important insights into the folding mechanism of such knotted proteins at the molecular level.<sup>20–23</sup>

*Escherichia coli* tRNA (guanosine-1) methyltransferase D (TrmD) (PDB: 1P9P) possesses a deep trefoil knot within its protein structure.<sup>26–31</sup> TrmD is responsible for the methylation of G37 in tRNA containing the sequence of G36pG37, which is critical to preventing frameshifting in the protein translation process.<sup>26</sup> TrmD catalyzes the modification process by using *S*-adenosyl-L-methionine (AdoMet) as the methyl donor, and it is suggested that the conserved knot structure is key to accommodating the adenosine moiety of AdoMet by the loosening and retightening of the trefoil knot during the methylation process.<sup>27</sup> Therefore, understanding the folding mechanism of TrmD as well as the structural dynamics of its knotted structure is not only important for elucidating the general folding mechanism underlying knotted proteins, but also key to understanding the functional role of the trefoil knot in the biological function of TrmD.

TrmD is an  $\alpha/\beta$  protein and its three dimensional structure can be readily divided into two subdomains (Fig. 1):<sup>26</sup> the N-terminal  $\alpha/\beta$  assembly domain (residues 1–161) and the C-

Department of Chemistry, University of British Columbia, Vancouver, BC, V6T 1Z1, Canada. E-mail: hongbin@chem.ubc.ca

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc02796k



**Fig. 1** (A) The three-dimensional structure of the knotted protein TrmD (PDB code: 1P9P). TrmD is composed of two subdomains: the N-terminal Rossman fold with the motif of alternating  $\beta$  strand– $\alpha$  helix– $\beta$  strand (residues 1–161, colored in blue and yellow) and the C-terminal  $\alpha$  helical domain (residues 162–246, colored in red). The knotting loop is colored in yellow (residues 81–161), and the knot tails are colored in blue and red. Residues 9 and 246, along which TrmD is stretched in the OT experiment, are colored in grey. (B) Schematics of the experimental setup of optical tweezers. Protein chimera TrmD<sub>9,246</sub>–(NuG2)<sub>2</sub> is coupled with two DNA handles via thio-maleimide chemistry. The NuG2 domain serves as a fingerprint for identification of single-molecule stretching events. This DNA–protein chimera is then held by two different polystyrene beads, with streptavidin and anti-digoxigenin modified on their surface, respectively. One bead is held by a glass pipette and the other is trapped by the OT. Stretching or relaxation of the protein chimera can be achieved by moving the laser trap.

terminal  $\alpha$  helical domain (residues 162–246, colored in red). The N-terminal  $\alpha/\beta$  assembly domain is a fold related to the Rossman fold, with a rather compact globular structure, and the C-terminal  $\alpha$  helix domain consists of a long disordered polypeptide coil and four relatively flexible  $\alpha$  helices. TrmD contains a deep trefoil knot in its structure, with residues 81–161 (colored in yellow) serving as the knot core (also called the knotting loop), and residues 1–80 (colored in blue) and 162–246 (colored in red) as knot tails. Thus, TrmD serves as an ideal model system for investigating the folding–unfolding mechanism of trefoil-knotted proteins.

Amongst the experimental techniques used to study knotted and slipknotted proteins, single-molecule force spectroscopy (SMFS) is of particular advantages due to its ability to manipulate the knot/slipknot structure by directly stretching the structure, offering a unique perspective to study the knotting/unknottting process of these proteins.<sup>8,16,32–35</sup> Atomic force microscopy (AFM) and optical tweezers (OT)-based SMFS techniques have been exploited for such studies. Due to the relatively low force resolutions ( $\sim 10$  pN), AFM was mainly used to probe the unfolding/untying mechanism of knotted and slipknotted proteins, and direct observation of their folding was largely not possible in these studies.<sup>16,32,35</sup> Due to its superb resolution in force (0.1 pN) and long-term stability,<sup>8,36–40</sup> OT has just been utilized to investigate the folding/unfolding mechanism of the 5<sub>2</sub>-knotted protein ubiquitin C-terminal hydrolase

isoenzyme L1 (UCH-L1) and slipknotted proteins AFB3-109 and pyruvoyl-dependent arginine decarboxylase (PADC).<sup>8,33,34</sup> Given the diversity of knots/slipknots in proteins, much more can be learned about their folding mechanisms using the OT-based SMFS technique. Due to the simplicity of the trefoil knot, TrmD serves as an ideal model system for the OT study.

Here, we combined OT with protein engineering techniques to investigate the unfolding and folding mechanisms of TrmD. Our results showed that upon stretching from its N- and C-termini, TrmD can be mechanically unfolded and stretched into an unfolded polypeptide chain with a tightened trefoil knot. Both two-state and three-state mechanical unfolding pathways were observed. Upon relaxation, the tightened knot can be loosened by thermal energy and the unfolded polypeptide can fold back to its native state efficiently. The folding of TrmD from such an unfolded state containing a knot is fast and robust. Distinct from a typical two-state folding process, the folding of TrmD from an unstructured polypeptide containing a knot involves a loosening process of the tightened trefoil knot prior to the folding into its native state. In addition, by using a truncation mutant of TrmD, we unfolded TrmD and untied the trefoil knot. Our results showed that the folding of TrmD from such an unknotted and unfolded state was much slower, with the knotting process as the rate-limiting step during the folding of the whole knotted protein.

## Materials and methods

### Protein engineering

The wild-type TrmD contains two cysteine residues in its sequence (C112 and C178). To avoid undesired coupling to the designed DNA handles during the coupling reactions, both C112 and C178 were mutated to serine. A codon-optimized gene encoding the cysteine-free TrmD with the desired restriction sites (5' *Bam*HI and 3' *Bgl*II and *Kpn*I) was custom synthesized (GeneScript). To avoid the inefficient coupling between the DNA handles and the first amino acid residue of wild-type TrmD, which was partially buried in the N-terminal subdomain, the 9<sup>th</sup> amino acid residue phenylalanine was mutated to cysteine (F9C) using the megaprimer approach of site-directed mutagenesis.<sup>41</sup> Following our well-established iterative molecular biology strategy,<sup>42</sup> we constructed the gene TrmD<sub>9,246</sub>–(NuG2)<sub>2</sub>. The gene TrmD<sub>9,246</sub>–(NuG2)<sub>2</sub> was then subcloned into a modified expression vector pQE80L–Cys, which carried a cassette allowing us to add one cysteine residue only at the C-terminus of the target protein, to build pQE80L/TrmD<sub>9,246</sub>–(NuG2)<sub>2</sub>–Cys. The full sequence of the engineered TrmD<sub>9,246</sub>–(NuG2)<sub>2</sub>–Cys is shown in the ESI.†

The gene encoding TrmD<sub>45,128</sub> was constructed using the megaprimer approach. TrmD  $\Delta$ helix<sub>9,161</sub> and TrmD  $\Delta$ helix<sub>45,128</sub> were obtained by a further regular polymerase chain reaction (PCR) to remove the C-terminal  $\alpha$  helical domain (residues 162–246). TrmD  $\Delta$ helix<sub>45,128</sub> and TrmD<sub>45,128</sub> were subcloned into the pQE80L vector for protein expression. All of the constructed genes were confirmed by DNA sequencing.

The engineered protein was overexpressed in *Escherichia coli* strain DH5 $\alpha$  at 37 °C in 250 mL 2.5% LB media with 100 mg L<sup>–1</sup>



antibiotic ampicillin. 1 mM isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG, Thermo Fisher Scientific, Waltham, MA) was added to induce protein overexpression at an optical density of around 0.8. The protein overexpression was allowed to continue for 4 hours at 37 °C. The bacterial cell pellets were harvested by centrifugation at 5000 rpm at 4 °C for 10 minutes and resuspended in 10 mL phosphate buffered saline (PBS, 10 mM, pH 7.4) buffer. 10  $\mu$ L protease inhibitor cocktail (SIGMA-ALDRICH, St. Louis, MO), 50  $\mu$ L 100 mg mL<sup>-1</sup> lysozyme from egg white (SIGMA-ALDRICH, St. Louis, MO), 1 mL 10% (w/v) Triton X-100 (VWR, Tualatin, OR), 50  $\mu$ L 1 mg mL<sup>-1</sup> DNase I (SIGMA-ALDRICH, St. Louis, MO) and 50  $\mu$ L 1 mg mL<sup>-1</sup> RNase A (Bio Basic Canada Inc, Markham, ON) were added for cell lysis. The lysis reaction was kept at 4 °C for 40 minutes. The supernatant containing the target protein was isolated by centrifugation at 10 000 rpm at 4 °C for 1 hour, and the protein was purified using a Co<sup>2+</sup> affinity column with a TALON His-tag purification kit (TaKaRa Bio USA Inc, Mountain View, CA). The protein was eluted and stored in elution buffer (10 mM PBS, 300 mM NaCl, 250 mM imidazole). The purified protein was at a concentration of  $\sim$ 0.5 mg mL<sup>-1</sup> and stored at -20 °C.

### Preparation of DNA handles and DNA-protein chimera

DNA handles were prepared *via* the method described previously.<sup>34,40,43</sup> Two DNA handles were synthesized *via* regular PCR amplification by using pGEMEX-1 plasmid as the template and the modified primers purchased from Integrated DNA Technologies (IDT Inc, San Jose, CA), as reported previously.<sup>34</sup> A QIAquick PCR purification kit (QIAGEN, Germantown, MD) was used to purify the PCR products followed by PCR amplification. The length of the DNA handles is 802 bp (biotin-DNA-maleimide) and 558 bp (digoxigenin-DNA-maleimide), which corresponds to the contour length of 273 nm and 190 nm, respectively. Then the DNA handles were allowed to react with 4-(*N*-maleimidomethyl) cyclohexanecarboxylic acid *N*-hydroxysuccinimide ester (SMCC, SIGMA-ALDRICH, St. Louis, MO) overnight. The freshly expressed proteins were reduced with 1 mM tris(2-carboxyethyl) phosphine (TCEP, SIGMA-ALDRICH, St. Louis, MO) for 1 hour and purified using Zeba desalting columns (7k MW, Thermo Fisher Scientific, Waltham, MA) to remove the extra TCEP. The reduced proteins were diluted to  $\sim$ 3  $\mu$ M using Tris buffer (20 mM Tris, 150 mM NaCl, pH 7.4) and 1  $\mu$ L of the diluted protein was allowed to react with 1  $\mu$ L of 3  $\mu$ M prepared DNA handles at room temperature overnight. The DNA-protein chimera was diluted to  $\sim$ 10 nM and stored at -80 °C.

### Optical tweezers experiments

The optical tweezers experiments were carried out on a Mini-Tweezers setup.<sup>44,45</sup> 1  $\mu$ L of streptavidin-coated polystyrene beads (1% w/v 1  $\mu$ m, Spherotech Inc, Lake Forest, IL) was diluted using 3 mL Tris buffer and injected into the fluid chamber *via* a syringe. Subsequently, a single streptavidin-coated bead was captured by a laser beam and moved onto the tip of a glass pipette. The bead was fixed tightly by applying a vacuum through the pipette. 1  $\mu$ L of the 10 nM DNA-protein

chimera was added to 4  $\mu$ L of anti-digoxigenin-coated polystyrene beads (0.5% w/v, 2  $\mu$ m, Spherotech Inc, Lake Forest, IL) for coupling. The coupling reaction was kept at room temperature for 30 minutes. The beads coupled with the DNA-protein chimera on its surface were also diluted with 3 mL Tris buffer and injected into the fluid chamber. A single anti-digoxigenin-coated bead was then trapped by a laser beam and brought into contact with the streptavidin-coated bead to establish the dumbbell for stretching and relaxation (Fig. 1B).

Stretching the protein-DNA chimera resulted in force-distance curves, in which the distance contains the contribution of the extension of the protein-DNA construct as well as the compliance of the optical trap. The MiniTweezers setup is a nonlinear optical trap, and its stiffness varies with the stretching force.<sup>44,45</sup> Therefore, force-distance curves measured using the MiniTweezers setup cannot be directly converted into force-extension curves of the protein-DNA construct itself. As such, the force-distance curves cannot be directly fitted to the worm-like-chain (WLC) model of polymer elasticity to measure the contour length increment ( $\Delta L_c$ ) of protein unfolding. For a given unfolding event with a contour length increment of  $\Delta L_c$  in the force-distance curve, the length increase at a given force  $F$  gives the extension  $E$  of the unfolded polypeptide chain of a contour length of  $\Delta L_c$ . By fitting the measured force-extension relationship to the WLC model, we can measure the persistence length and  $\Delta L_c$  of the unfolded polypeptide chain being released from the unfolding.<sup>33,40,46</sup>

### Extracting the protein unfolding/folding kinetics

A direct and model-free calculation method proposed by Oesterhelt *et al.*<sup>47</sup> was used to measure the force-dependent two-state unfolding/folding rate constants of proteins. The force-distance curves, generated by stretching (or relaxing) the target protein at a constant velocity, were divided into small time windows ( $\Delta t$ ) and the force can be regarded as constant within the time window. The unfolding/folding rate constants at force  $F$  can be calculated as  $k(F) = N(F)/M(F) \times \Delta t$ , where  $N(F)$  is the total number of unfolding/folding events observed at force  $F$  and  $M(F)$  is the total number of time windows at force  $F$ . In our experiments, we used a bin of 1 pN in force, which will give rise to a  $\Delta t$  of 0.2–0.3 s. This time window was small enough to allow the use of the Oesterhelt method to extract rate constants with good accuracy.

Force-dependency of the unfolding rate constants of short-lived intermediate states was measured directly by a single exponential fitting to the dwell-time distribution of the intermediate states.<sup>48</sup> The relationship of probability density *vs.* dwell-time can be described as  $Pd(t) = \alpha(F)\exp(-\alpha(F) \times t)$ , where  $\alpha(F)$  is the unfolding rate constant under force  $F$ ,  $t$  is the dwell-time of short-lived intermediate states and  $Pd(t)$  is the probability density for a specific dwell-time.

The force-dependency of unfolding/folding rate constants was subsequently fitted using the Bell-Evans model, to extract the intrinsic protein unfolding/folding kinetics at zero force.<sup>49,50</sup> The Bell-Evans model can be described as  $\alpha(F) = \alpha_0 \exp(F\Delta x_u/k_B T)$  and  $\beta(F) = \beta_0 \exp(-F\Delta x_f/k_B T)$ , where  $k_B$  is the Boltzmann





constant,  $T$  is the temperature in Kelvin,  $\alpha(F)$  (or  $\beta(F)$ ) is the unfolding (or folding) rate constant under force  $F$ , and  $\alpha_0$  (or  $\beta_0$ ) is the intrinsic unfolding (or folding) rate constant at 0 pN.  $\Delta x_u$  (or  $\Delta x_f$ ) is the distance from the native protein to transition state (or from the unfolded to transition state).

A standard double-pulse protocol<sup>42</sup> was applied to measure the knotting kinetics of mutant TrmD  $\Delta\text{helix}_{45,128}$ . The protein was first untied under stretching, and then quickly relaxed to its initial length. After a time delay  $\Delta t$ , a second stretching was applied to the molecule to check if the knotting was achieved again within the time  $\Delta t$ . Protein knotting probability was dependent on the time delay for protein refolding and can be expressed as  $P(t) = 1 - \exp(-\beta_0 \times t)$ , where  $\beta_0$  is the intrinsic protein knotting rate constant.

## Results

### Stretching TrmD leads to the mechanical unfolding of TrmD and the tightening of the trefoil knot

To investigate the mechanical unfolding of the trefoil knotted protein TrmD using OT, we stretched TrmD from its C-terminus (residue 246) and residue 9 using the TrmD variant Phe9Cys (termed TrmD<sub>9,246</sub>) of which the more solvent exposed residue 9 (than the N-terminus) (Fig. 1) helped improve the coupling efficiency of DNA handles to the protein. We then engineered the dsDNA-protein chimera, DNA-TrmD<sub>9,246</sub>-(NuG2)<sub>2</sub>-DNA, by coupling doubled stranded DNA handles to the construct TrmD<sub>9,246</sub>-(NuG2)<sub>2</sub> via thiol-maleimide coupling chemistry.<sup>33,34,40,43</sup> In this DNA-protein construct, the well-characterized NuG2 domains serve as a fingerprint for identifying single-molecule stretching events.<sup>34,43</sup>

Stretching TrmD<sub>9,246</sub>-(NuG2)<sub>2</sub> resulted in force-distance curves with the characteristic sawtooth-like appearance (Fig. 2A, inset). Three unfolding events were observed prior to the B-S overstretching transition of the dsDNA handles at  $\sim 65$  pN. Two unfolding events occurred at higher forces (typically higher than 30 pN) with a  $\Delta L_c$  of 18 nm (colored in black), corresponding to the mechanical unfolding of the fingerprint NuG2 domains. Thus, the third unfolding event at  $\sim 10$  pN can be readily attributed to the unfolding of TrmD (colored in red). Since TrmD unfolds at forces that are significantly lower than those for NuG2, subsequent stretching of TrmD was limited to  $\sim 15$  pN so that only TrmD was unfolded and NuG2 domains remained folded in the experiment (Fig. 2A and B).

The mechanical unfolding of TrmD<sub>9,246</sub> was observed to follow bifurcated pathways, involving both two-state and three-state pathways. Fig. 2A shows typical force-distance curves of TrmD following the two-state unfolding pathway. Fitting the force-extension relationships to the WLC model of polymer elasticity gave a contour length increment  $\Delta L_c$  of  $36.8 \text{ nm} \pm 1.8 \text{ nm}$  upon the unfolding of TrmD (Fig. 2C). TrmD is 246 amino acid residues long, but only 238 residues were involved in the pulling experiments of TrmD<sub>9,246</sub>. The complete unfolding of TrmD<sub>9,246</sub> would lead to a  $\Delta L_c$  of  $\sim 80 \text{ nm}$  ( $238 \text{ aa} \times 0.36 \text{ nm per aa} - 5.4 \text{ nm} = 80.3 \text{ nm}$ , where 5.4 nm is the distance between residues 9 and 246) if there was no knotted conformation. It is evident that the calculated  $\Delta L_c$  is

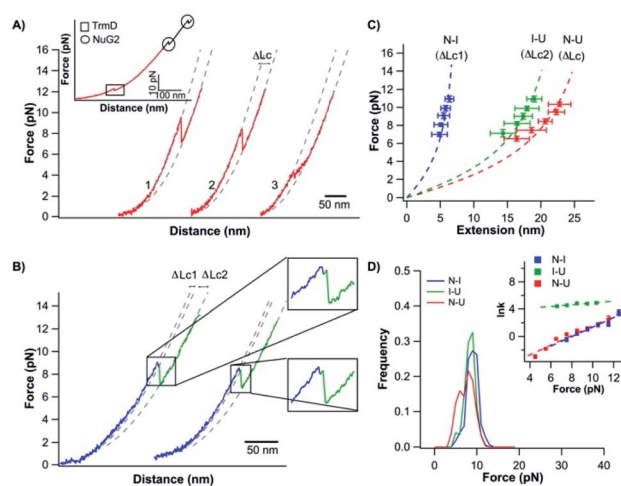


Fig. 2 Stretching TrmD<sub>9,246</sub> to a tightened trefoil knot involves two parallel pathways. (A) Representative force–distance curves showing two-state unfolding of TrmD<sub>9,246</sub>. The pulling speed was  $50 \text{ nm s}^{-1}$ . Inset: representative force–distance curves showing the unfolding of full-length TrmD<sub>9,246</sub>-(NuG2)<sub>2</sub>. The mechanical unfolding events of the fingerprint domains NuG2 are colored in black and indicated with circles, and the unfolding events of TrmD<sub>9,246</sub> are indicated with squares. (B) Representative force–distance curves of TrmD<sub>9,246</sub> via the three-state unfolding pathway. The pulling speed was  $50 \text{ nm s}^{-1}$ . The zoom-in view showed that the protein unfolds via an intermediate state. The protein unfolding from N  $\rightarrow$  I and I  $\rightarrow$  U is colored in blue and green, respectively. The dashed lines in (A) and (B) are pseudo WLC fits to the force–distance curves. The fits were primarily used for identification of various states (folded, intermediate and unfolded state) of protein in the force–distance curves, and the fitting parameters were not physically meaningful. (C) Force–extension relationships of TrmD<sub>9,246</sub>. Extension is the length increase at a given force upon unfolding. Error bars correspond to the standard deviation of the data measured from different molecules. Dashed lines are WLC fits to the force–extension data measured for N  $\rightarrow$  I (blue), I  $\rightarrow$  U (green) and N  $\rightarrow$  U (red), respectively. The WLC fits yielded a persistence length of  $0.8 \text{ nm}$  and a  $\Delta L_{c1}$  of  $9.8 \pm 0.4 \text{ nm}$ ,  $\Delta L_{c2}$  of  $28.9 \pm 1.0 \text{ nm}$  and  $\Delta L_c$  of  $36.8 \pm 1.8 \text{ nm}$ , respectively. (D) Unfolding force histograms of TrmD<sub>9,246</sub> at a pulling speed of  $50 \text{ nm s}^{-1}$ . Inset: force-dependency of unfolding rate constants measured by the model-free method proposed by Oesterhelt *et al.* Fitting to the Bell–Evans model yielded the intrinsic unfolding rate constant  $\alpha_0$  and the unfolding distance  $\Delta x_u$ , as summarized in Table S1.†

significantly larger than the experimentally measured one, suggesting that part of TrmD<sub>9,246</sub> may have unfolded at low forces, which are below the detection limit of the MiniTweezers.

It is of note that TrmD contains a large C-terminal  $\alpha$ -helical domain (residues 133 to 246). The  $\alpha$ -helix structure is mechanically labile and often unfolds at very low forces, resulting in the missing of a clear mechanical unfolding signature in single-molecule force spectroscopy experiments.<sup>16,51</sup> We speculated that the unfolding of the labile  $\alpha$ -helix (including the C-terminal  $\alpha$ -helical domain and the  $\alpha$ -helix in the knot core (residues 133–161)) would account for the missing  $\Delta L_c$  in TrmD. To verify this hypothesis, we engineered an  $\alpha$ -helix truncation variant TrmD  $\Delta\text{helix}$  (residues 9–161), in which residues 162–246 were deleted. Stretching TrmD  $\Delta\text{helix}_{9,161}$  resulted in a  $\Delta L_c$  of  $\sim 37 \text{ nm}$ , identical to that of wt TrmD

(Fig. S1†), suggesting that indeed the unfolding of the C-terminal  $\alpha$ -helix occurred at low forces and in a gradual fashion and did not result in a clear unfolding signature in the force–distance curves.

If the  $\alpha$ -helix in the knot core (residues 133–161) also unfolds at low forces similar to the C-terminal  $\alpha$ -helical domain, the unfolding of the rest of the folded protein structure (residues 9 to 132) would yield a  $\Delta L_c$  of 42.7 nm ( $124 \text{ aa} \times 0.36 \text{ nm per aa} - 1.9 \text{ nm} = 42.7 \text{ nm}$ , where 1.9 nm is the distance between Cys9 and Ser132). This  $\Delta L_c$  of 42.7 nm is about 6 nm longer than the experimentally measured  $\Delta L_c$  of TrmD<sub>9,246</sub>. Distinct from the unfolding of NuG2, the unfolding of TrmD<sub>9,246</sub> would result in a tightened trefoil knot present within the unstructured polypeptide chain. Thus, the “missing  $\Delta L_c$ ” of  $\sim 6 \text{ nm}$ , corresponding to  $\sim 17$  amino acid residues, can be readily attributed to the formation of the tightened trefoil knot. This result is in good agreement with the size of the protein trefoil knot measured previously in single-molecule force spectroscopy experiments,<sup>8,32</sup> and suggested that TrmD can be mechanically stretched into an unfolded conformation with a tightened trefoil knot.

TrmD<sub>9,246</sub> unfolds *via* a two-state pathway, with an average unfolding force of  $\sim 8 \text{ pN}$  at a pulling speed of  $50 \text{ nm s}^{-1}$  (Fig. 2D and Table S1†). Using the method proposed by Oesterhelt *et al.*,<sup>47</sup> we measured the unfolding rate constant as a function of the stretching force (inset, Fig. 2D). Fitting the experimental data to the Bell–Evans model<sup>49,50</sup> gave an intrinsic unfolding rate constant  $\alpha_0$  of  $(3.8 \pm 0.5) \times 10^{-3} \text{ s}^{-1}$  at zero force and an unfolding distance between the native state and transition state  $\Delta x_u$  of  $2.9 \pm 0.3 \text{ nm}$ .

### The mechanical unfolding of TrmD involved a short-lived intermediate state

In our OT experiments, about 68% of TrmD<sub>9,246</sub> followed an apparent two-state unfolding pathway. A small portion ( $\sim 32\%$ ) of TrmD<sub>9,246</sub> was observed to unfold *via* a three-state pathway, involving a short-lived unfolding intermediate state (Fig. 2B). WLC fits to the extension at different forces yielded a  $\Delta L_{c1}$  of  $9.8 \pm 0.4 \text{ nm}$  and a  $\Delta L_{c2}$  of  $28.9 \pm 1.0 \text{ nm}$ , respectively ( $N \rightarrow I$  and  $I \rightarrow U$ , Fig. 2C). The sum of  $\Delta L_{c1}$  and  $\Delta L_{c2}$  equals  $38.7 \text{ nm}$ , in good agreement with the  $\Delta L_c$  yielded from two-state unfolding pathway. These results suggested that the mechanical unfolding pathways of TrmD are bifurcated,<sup>52,53</sup> and the three-state pathway involves a well-defined intermediate state during the mechanical unfolding pathway.

In the three-state pathway, the first unfolding event (from the native to the intermediate state) occurred at an unfolding force of  $9.5 \pm 1.3 \text{ pN}$ . And the second unfolding event (from the intermediate state to the unfolded state) occurred at  $8.9 \pm 1.3 \text{ pN}$  (Fig. 2D and Table S1†). It is of note that the unfolding intermediate state was short-lived in the range of force we measured ( $7\text{--}10 \text{ pN}$ ), and the lifetime, with an average of  $\sim 0.02 \text{ s}$ , was not sensitive to the stretching force. Fitting the force-dependency of the unfolding rate of the three-state unfolding to the Bell–Evans model yielded an intrinsic unfolding rate constant  $\alpha_0$  of  $(3.6 \pm 0.6) \times 10^{-3} \text{ s}^{-1}$  ( $N \rightarrow I$ ) and  $(3.6 \pm$

$0.3) \times 10^{-3} \text{ s}^{-1}$  ( $I \rightarrow U$ ) (inset, Fig. 2D and Table S1†), respectively. The unfolding rate constant for  $I \rightarrow U$  was four orders of magnitude larger than that for  $N \rightarrow I$ , suggesting that the intermediate state was rather mechanically unstable and can be unfolded rapidly once the intermediate state is formed.

It is of note that the unfolding kinetics parameters for  $N \rightarrow I$  are similar to those of  $N \rightarrow U$ . This similarity together with the short life time of the intermediate state gives rise to the possibility that the observed two-state pathway is just a special case of the three-state pathway, in which, due to the insufficient time resolution, the intermediate state was not resolved. Since our MiniTweezers setup has a temporal resolution of  $\sim 0.1 \text{ ms}$  at a force of  $\sim 10 \text{ pN}$ , we carried out experiments at a higher sampling frequency ( $30 \text{ kHz}$ ) to gain a better resolution of the unfolding events. Our data at higher sampling frequency only showed a small increase of the observed three-state unfolding events ( $\sim 30\%$  at  $1 \text{ kHz}$  *vs.*  $\sim 35\%$  at  $30 \text{ kHz}$ ). This result suggested that most of the two-state unfolding events were likely not due to the insufficient time resolution, and the two-state and three-state pathways are more likely to be two different pathways. However, the difference between the two unfolding pathways is small.

Nevertheless, it is important to point out that the observed unfolding intermediate state and the bifurcation of the unfolding pathways are not associated with the trefoil knot, but a property of the folded protein structure itself.

### The compaction of the trefoil knot

It is of interest to note that when the mechanical unfolding of TrmD<sub>9,246</sub> occurred at  $\sim 9 \text{ pN}$ , the force–distance curve of the unfolded TrmD largely followed the typical WLC behavior of an unfolded polypeptide chain (Fig. 2A, curves 1 and 2). However, if the unfolding occurred at lower forces ( $< 6 \text{ pN}$ ) (Fig. 2A, curve 3, and Fig. S2†), the force–distance curve of the unfolded TrmD showed a clear deviation from the WLC behavior of an unfolded polypeptide chain (grey dotted line). This behavior suggests that at lower forces, TrmD<sub>9,246</sub> first unfolded into a state that is of a shorter  $L_c$ , and further stretching caused the lengthening of the unfolded polypeptide chain until its length is the same as that of the unfolded TrmD<sub>9,246</sub> with a tightened knot (of the size of  $6 \text{ nm}$ ). This behavior can be explained by the compaction of the trefoil knot by the stretching force as schematically shown in Fig. S2.† At forces higher than  $7 \text{ pN}$ , the force–distance curve returned to the normal WLC behavior of an unfolded polypeptide chain, suggesting that the trefoil knot had been stretched to a tightened knot at a force of  $\sim 6\text{--}7 \text{ pN}$ . Further compaction/tightening would require higher forces. Similar compaction behavior was observed for a  $5_2$  knot in the pulling experiments of UCH-L1.<sup>8</sup>

### The unfolded and knotted TrmD can refold back to its native state upon relaxation

After its mechanical unfolding, TrmD<sub>9,246</sub> was converted to an unfolded polypeptide with a tightened trefoil knot. To investigate whether the unfolded TrmD with a tightened trefoil knot can refold back to its native state, we relaxed the unfolded



polypeptide to allow protein refolding. As shown in Fig. 3A, during the relaxation process (curves colored in black), the protein was observed to refold, characterized by a force jumping event at  $\sim 4$  pN and a length shortening to the original length of TrmD<sub>9,246</sub>. Subsequent stretching led to an identical signature of the unfolding of the knotted protein TrmD<sub>9,246</sub>, suggesting that the relaxation had resulted in a successful refolding of the native conformation of TrmD<sub>9,246</sub>.

However, the folding behavior of TrmD<sub>9,246</sub> is distinct from that of a typical two-state folding event. During the relaxation/refolding process of a typical two-state folder, such as NuG2 (Fig. S3†), the unfolded polypeptide chain followed the typical WLC behavior until the refolding event occurs. In contrast, during the relaxation of the unfolded TrmD<sub>9,246</sub>, the force–distance relationship of the unfolded TrmD<sub>9,246</sub> polypeptide chain showed a clear deviation from that of WLC when the force was lower than  $\sim 7$  pN. When the force reached  $\sim 4$  pN, the unfolded TrmD refolded, as evidenced by a sudden force jump event. This behavior seemed to be the reverse of the compaction process of the trefoil knot observed during unfolding (Fig. S2†), suggesting that the global folding of the knotted TrmD<sub>9,246</sub> may proceed with the loosening process of the tightened trefoil knot. Moreover, the reversibility of the compaction and loosening processes and their deviation from the WLC model suggested that elastic energy was stored in the polypeptide chain during

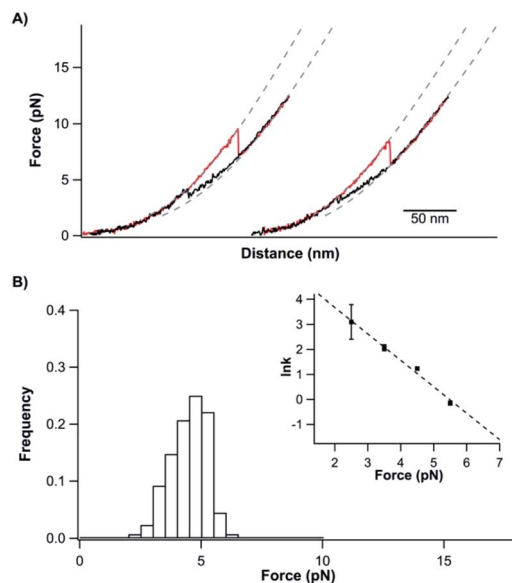
the knot compaction process, which was released during the loosening process.

The “quasi-two-state” folding forces of TrmD<sub>9,246</sub> showed a narrow distribution with an average force of  $\sim 4.5$  pN at a pulling speed of  $50 \text{ nm s}^{-1}$  (Fig. 3B and Table S1†). Moreover, fitting the force-dependent folding rate constants to the Bell–Evans model gave a folding rate constant at zero force of  $(3.3 \pm 0.3) \times 10^2 \text{ s}^{-1}$  (inset, Fig. 3B). These results indicated that the folding of TrmD from its unfolded state with the pre-formed knot is fast and robust.

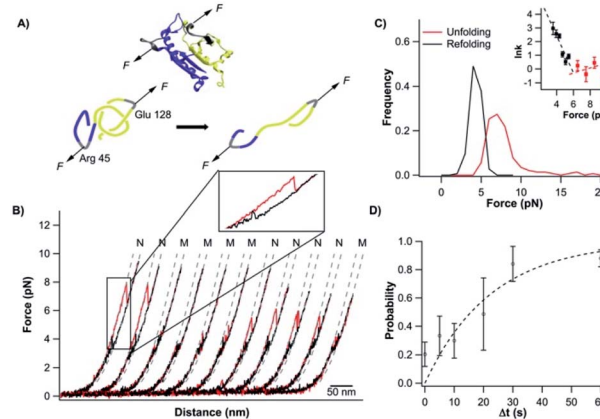
### Untying and tying the knotted structure of TrmD by using optical tweezers

Having characterized the mechanical unfolding of TrmD<sub>9,246</sub> and the folding of unfolded and knotted TrmD<sub>9,246</sub>, we then investigated the possibility of unfolding TrmD and untying its trefoil knot. The trefoil knot in TrmD is very deep, and even removing 79 residues from either N- or C-termini will not destroy the trefoil knot. This deep knot is difficult to untie in a force spectroscopy experiment. In order to untie the trefoil knot in TrmD, we used the helix truncation variant TrmD  $\Delta$ helix, in which the depth of the trefoil knot is significantly reduced. Residues Arg45 and Asp128 were mutated to cysteines to allow TrmD to be stretched from these two residues (this construct is termed TrmD  $\Delta$ helix<sub>45,128</sub>) (Fig. 4A).

Stretching TrmD  $\Delta$ helix<sub>45,128</sub> resulted in the typical force–distance curves shown in Fig. 4B. The unfolding of TrmD



**Fig. 3** Unfolded and tightened TrmD<sub>9,246</sub> can fold back to its native state. (A) Representative stretching–relaxation curves of TrmD<sub>9,246</sub> at a pulling speed of  $50 \text{ nm s}^{-1}$ . Consecutive stretching and relaxing curves revealed that the tightened protein was able to fold back to the native state. During relaxation, the force–distance relationship of TrmD deviated from the typical behaviour of a DNA–protein chimera starting from  $\sim 6$  pN (as indicated by the dotted pseudo WLC fits). The protein eventually folded back into its native state via a quasi-two-state mechanism at  $\sim 4$  pN. (B) The force histograms of the “quasi-two-state” folding of TrmD<sub>9,246</sub>. Inset: force-dependency of the folding rate constants measured using the Oesterhelt method. The dotted line is the fitting using the Bell–Evans model. Fitting parameters  $\beta_0$  and  $\Delta x_f$  are shown in Table S1.†



**Fig. 4** Untying and tying the knotting loop of TrmD  $\Delta$ helix<sub>45,128</sub>. (A) Schematics of stretching protein TrmD  $\Delta$ helix<sub>45,128</sub>. The stretching sites are colored in grey. (B) Representative stretching–relaxation curves of TrmD  $\Delta$ helix<sub>45,128</sub> showing distinct unfolding behavior of the same molecule. Relaxation of the unfolded protein led to a “refolding” event at  $\sim 4$  pN. However, it remained unknown whether the protein folded into its native state with the trefoil knot. In the subsequent stretching force–distance curves, either a typical two-state unfolding at a force of  $\sim 7$  pN or a “hump” feature was observed. (C) Histograms of unfolding and “refolding” forces of TrmD  $\Delta$ helix<sub>45,128</sub> at a pulling speed of  $50 \text{ nm s}^{-1}$ . Inset: force-dependency of the unfolding/folding rate constants. Dotted lines are fits of the experimental data to the Bell–Evans model. (D) Relationship of the knotting probability versus  $\Delta t$  at zero force. The dotted line is the fit of the experimental data to the first order rate law, yielding a knotting rate constant of  $(3.8 \pm 1.0) \times 10^{-2} \text{ s}^{-1}$ .



$\Delta\text{helix}_{45,128}$  largely follows a two-state unfolding pathway with an average unfolding force of  $\sim 8$  pN at a pulling speed of  $50 \text{ nm s}^{-1}$ . Fitting the WLC model of polymer elasticity to the extension of  $\Delta\text{helix}_{45,128}$  led to a  $\Delta L_c$  of  $25.6 \pm 0.7 \text{ nm}$  upon the unfolding of TrmD  $\Delta\text{helix}_{45,128}$  (Fig. S4†). When stretched from residues 45 and 128, 84 residues were involved. The complete unfolding of TrmD  $\Delta\text{helix}$  followed by the untying of the trefoil knot would result in a  $\Delta L_c$  of  $26.9 \text{ nm}$  ( $84 \text{ aa} \times 0.36 \text{ nm per aa} - 3.3 \text{ nm} = 26.9 \text{ nm}$ , where  $3.3 \text{ nm}$  is the distance between C45 and C128), in excellent agreement with the experimentally measured  $\Delta L_c$ . This result strongly indicated that stretching the TrmD  $\Delta\text{helix}$  from residues 45 and 128 resulted in the mechanical unfolding of the TrmD  $\Delta\text{helix}$  and untying of its trefoil knot, providing a unique opportunity to investigate the knotting and refolding process of TrmD.

TrmD  $\Delta\text{helix}_{45,128}$  unfolds *via* a two-state pathway with an average unfolding force of  $\sim 8$  pN at a pulling speed of  $50 \text{ nm s}^{-1}$  (Fig. 4C and Table S1†). Fitting the force-dependent unfolding rate constants to the Bell-Evans model yields an intrinsic unfolding rate constant  $\alpha_0$  of  $(2.5 \pm 2.2) \times 10^{-1} \text{ s}^{-1}$  (Fig. 4C, inset). After the TrmD  $\Delta\text{helix}_{45,128}$  had been unfolded and untied, the unfolded and unknotted polypeptide chain was relaxed to zero force to allow the folding/knotting of TrmD. During the relaxation process, a “refolding” event was often observed at  $\sim 4$  pN, leading to a “refolded” state that is of similar length to the native  $\Delta\text{helix}_{45,128}$  (Fig. 4B). The average “refolding” force is  $\sim 4$  pN and the intrinsic “refolding” rate constant was estimated to be  $(2.5 \pm 0.4) \times 10^3 \text{ s}^{-1}$  (Fig. 4C, inset). However, it was unclear whether the trefoil knot had reformed in this “refolded” state. To check this, we stretched the “refolded” state in the next cycle. In about 20% of the traces, a clear unfolding event occurring at  $\sim 7$  pN was observed (Fig. 4B, curves marked by “N”). The unfolding signatures of these events were the same as those of the native TrmD  $\Delta\text{helix}$  with the trefoil knot, suggesting that in these “refolded” molecules, the trefoil knot had been retied. However, in 80% of the traces, the unfolding occurred at  $\sim 4$  pN, resulting in a hump-like feature (Fig. 4B, curves marked by “M”). This unfolding signature is different from that of unfolding of the native TrmD  $\Delta\text{helix}_{45,128}$ , suggesting that in these events, TrmD  $\Delta\text{helix}_{45,128}$  folded into a non-native or intermediate state that is mechanically labile, and the trefoil knot likely had not reformed. Therefore, the observed “refolding” at  $\sim 4$  pN for TrmD  $\Delta\text{helix}_{45,128}$  did not reflect the true refolding of the native trefoil TrmD. Instead, it is likely that such “refolding” events reflected the folding into an unknotted, non-native state or intermediate state.

The knotted conformation of TrmD is an important structural feature of the enzyme. Since the mechanical stretching of TrmD  $\Delta\text{helix}_{45,128}$  involved untying the trefoil knot by pulling the C-terminus of the protein threading out of the knot core, the correct folding of TrmD must involve tying the C-terminal knot tail into the knot core. It has been shown that knotting is the rate-limiting step in the folding of some knotted proteins.<sup>8,54</sup> To investigate if knotting is the rate limiting step for TrmD  $\Delta\text{helix}_{45,128}$ , we used OT to investigate the folding behavior of wt TrmD by stretching wt TrmD from its residues 45/128 (TrmD<sub>45,128</sub>).

To untie a protein knot, the length of the polypeptide chain between the two stretching points in the protein knot structure should be much longer than the length of the knot tail to be pulled out of the knot. For the pulling geometry of TrmD<sub>45,128</sub>, TrmD can be unfolded but its deep trefoil knot cannot be untied due to the long C-terminal knot tail (residues 162–246). Our OT experiments showed that the unfolding of TrmD<sub>45,128</sub> occurred at higher forces than those for wt TrmD, but displayed a  $\Delta L_c$  of  $27.6 \pm 0.7 \text{ nm}$  (Fig. 5B and C), which is the same as that of TrmD  $\Delta\text{helix}_{45,128}$ . But the folding of TrmD<sub>45,128</sub> was drastically different from that of TrmD  $\Delta\text{helix}_{45,128}$ . The folding of TrmD<sub>45,128</sub> occurred in a two-state fashion at  $\sim 6$  pN (Fig. 5D). Consecutive stretching–relaxation cycles showed that TrmD<sub>45,128</sub> could always fold correctly back to its native state, and no misfolding was observed. Kinetics analysis revealed a high folding rate constant at zero force (Fig. 5D, inset), indicating that TrmD<sub>45,128</sub> can fold rapidly in the presence of a pre-formed knotted structure.

The comparison of the folding behavior of TrmD<sub>45,128</sub> and TrmD  $\Delta\text{helix}_{45,128}$  clearly indicated the critical importance of the knotting step in the folding of TrmD. To determine the knotting kinetics of TrmD  $\Delta\text{helix}_{45,128}$ , we used a double-pulse protocol in the refolding experiment.<sup>42</sup> In the first pulse, the correctly folded and knotted TrmD  $\Delta\text{helix}_{45,128}$  was stretched to unfold the protein and untie the trefoil knot, and then the unfolded and untied polypeptide chain was allowed to relax to zero force and held at zero force for a period of time  $\Delta t$ . In the second pulse, the protein was stretched again. The occurrence

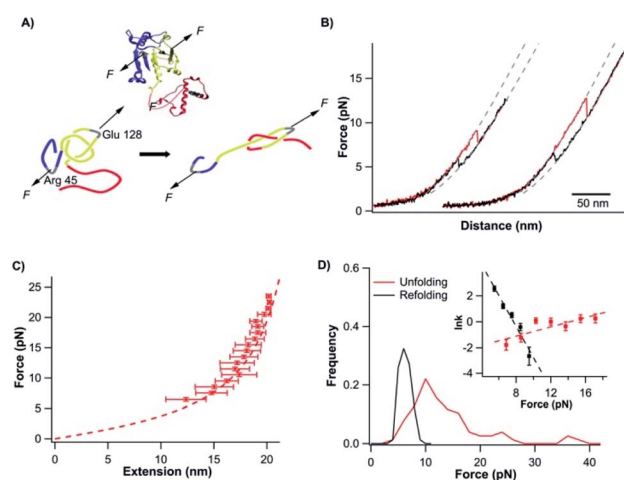


Fig. 5 Stretching the protein TrmD<sub>45,128</sub> does not untie the knotted conformation. (A) Schematics of stretching protein TrmD<sub>45,128</sub>. The stretching site is colored in grey. (B) Representative force–distance curves of TrmD<sub>45,128</sub>. Consecutive stretching–relaxation curves revealed that the protein can always fold back to its native state. (C) WLC fits to the force–extension relationship of TrmD<sub>45,128</sub> gave a persistence length of  $0.8 \text{ nm}$  and  $\Delta L_c$  of  $27.6 \text{ nm}$ . (D) Unfolding and refolding force histograms of TrmD<sub>45,128</sub> at a pulling speed of  $50 \text{ nm s}^{-1}$ . The inset shows the force-dependency of the unfolding and folding rate constants. Fitting the experimental data to the Bell–Evans model yielded the intrinsic unfolding/folding rate constants,  $\alpha_0 = (7.9 \pm 2.0) \times 10^{-2} \text{ s}^{-1}$  and  $\beta_0 = (1.0 \pm 0.1) \times 10^4 \text{ s}^{-1}$ , as well as the unfolding/folding distance,  $\Delta x_u = 0.7 \pm 0.2 \text{ nm}$  and  $\Delta x_f = 4.9 \pm 0.6 \text{ nm}$ .



of the unfolding signatures of the correctly folded and knotted TrmD  $\Delta$ helix<sub>45,128</sub> would indicate the successful knotting and folding of the protein. As shown in Fig. 4D, the percentage of successful knotting and folding events of TrmD  $\Delta$ helix<sub>45,128</sub> increased with the increasing of waiting time  $\Delta t$ . Assuming a first-order kinetics, we estimated a rate of knotting of TrmD  $\Delta$ helix of  $3.8 \pm 1.0 \times 10^{-2} \text{ s}^{-1}$ . This rate constant is considerably smaller than the folding rate of TrmD with a pre-formed knotting loop, suggesting that the protein knotting is indeed the rate-limiting step for the folding of the TrmD  $\Delta$ helix. Considering the depth of the trefoil knot in wt TrmD, it is likely that the knotting step of wt TrmD is significantly slower than that of TrmD  $\Delta$ helix, raising an interesting question if the folding of TrmD *in vivo* involves any molecular chaperone.

It is also important to note that the folding of the unfolded TrmD<sub>45,128</sub> also showed significant difference from the folding behavior of the unfolded TrmD<sub>9,246</sub>. The force–distance relationship of the unfolded TrmD<sub>45,128</sub> largely followed the WLC model, and no loosening process of the tightened trefoil knot was observed. This behavior can be readily understood by the unfolding behavior of TrmD<sub>45,128</sub> schematically shown in Fig. 5A. In this pulling geometry, the stretching force will not untie the trefoil knot (as occurred for TrmD  $\Delta$ helix<sub>45,128</sub>) due to the long C-terminal knot tail or tighten the trefoil knot (as occurred for wt TrmD<sub>9,246</sub>) after the unfolding of TrmD<sub>45,128</sub>. Instead, the stretching force acting on TrmD<sub>45,128</sub> will only expand the trefoil knot structure, leading to the absence of a tightened knot conformation in the unfolded TrmD<sub>45,128</sub>. Thus, the folding of the unfolded TrmD<sub>45,128</sub> does not involve the loosening up of the knot structure, but largely reflects the folding of the three dimensional structure of TrmD.

## Discussion

The deep trefoil knot structure is a common and unique feature of the SPOUT family of RNA methyltransferases (MTases), which catalyze the methylation of the base or ribose moiety of ribosomal RNA (rRNA) or transfer RNA (tRNA). TrmD is a conserved SPOUT MTase in bacteria and responsible for the methylation of tRNA.<sup>26–31</sup> It was proposed that the trefoil knot in TrmD is critical for its function, as the trefoil knot is responsible for accommodating the adenosine moiety of the methyl donor of AdoMet in the methylation process by loosening and retightening of the trefoil knot.<sup>27</sup> Thus, correct folding of TrmD into its knotted structure is crucial.

By using single-molecule OT, we have investigated the mechanical unfolding, tightening and untying of the trefoil knot, as well as the refolding of TrmD starting from its unfolded polypeptide with or without a pre-formed knot structure. Our results provided some new insights into the unfolding and folding mechanisms of TrmD.

After the mechanical unfolding of TrmD, we observed a mechanical compaction process of the trefoil knot at low forces (<6 pN), which led to the lengthening of contour length of the unfolded polypeptide chain and tightening of the trefoil knot. This is the first time that such a compaction process is observed in a simple trefoil knot. The fact that the compaction

process was only observed at low forces suggested that the compaction process may have completed in the taut unfolded polypeptide chain when TrmD was unfolded at higher forces, as in curves (1) and (2) in Fig. 2A. Moreover, after the compaction process had completed, the length of the unfolded polypeptide chain remained largely unchanged upon further stretching, implying that the knot size remained largely constant. And further tightening and even “jamming” of the trefoil knot (*i.e.* shrinking the knot size) will require much higher forces than that the OT can apply (typically <100 pN) (Fig. S5†). Thus, the stretching of the trefoil knot in the OT experiments most likely only resulted in a limited further tightening.

When the force is lower than  $\sim 7$  pN during the relaxation process, the thermal energy is sufficient to loosen up the knot tightened by a force at least up to 50 pN (Fig. S5†) and the knot can then expand. The observed compaction and loosening of the trefoil knot at low forces revealed the dynamic nature and structural plasticity of the trefoil knot structure in TrmD. It is worth noting that this compaction and loosening of the trefoil knot occurs when TrmD is unfolded. Further experiments to demonstrate that similar processes can occur in the folded TrmD will be critical for a deeper understanding of the function mechanism of TrmD, as crystallography studies suggested that loosening and tightening of the knot are critical for accommodating the adenosine moiety of the methyl donor of AdoMet in the methylation process by TrmD.<sup>27</sup>

It is also of note that such a compaction and loosening process of the trefoil knot was not observed in a trefoil knot prepared from UCH-L1 in a previous OT study.<sup>8</sup> Whether such a compaction and loosening process is related to the protein sequence and/or residual secondary structure remains to be elucidated.

Our results showed that the folding of TrmD with a pre-formed knot is fast and robust, and the knotting is the rate-limiting step for the overall folding of TrmD. These findings complement the previous results obtained on knotted proteins (a 5<sub>2</sub> knotted UCH-L1 and trefoil-knotted methyltransferases YibK and YbeA),<sup>8,54</sup> and raise the question if this feature is common among all knotted proteins.

It is important to point out that, due to the pulling geometry in TrmD<sub>45,128</sub>, the knotting can only occur by threading the C-terminal end in our OT experiment. In comparison, the depth of the trefoil knot in wild type TrmD prevented us from preparing an untied wild type TrmD, and thus it is not possible to directly study the knotting and folding process of wt TrmD using OT. Given the depth of the trefoil knot in wt TrmD, it is most likely that the folding of TrmD *in vivo* is different from what we observed in TrmD<sub>45,128</sub>. More studies, likely involving molecular chaperones or folding on ribosomes, will be needed to elucidate the folding mechanism of wt TrmD *in vivo*.

## Conclusions

Combining optical tweezers and protein engineering techniques, we investigated the mechanical unfolding and folding behavior of a knotted protein TrmD. Our results showed that TrmD can be mechanically unfolded along a direction defined





by two specific residues. Depending on the pulling geometry, the deep trefoil knot of TrmD can take different conformations. It can be tightened, untied, or expanded in the unfolded state by the stretching force. After the mechanical unfolding of TrmD, a compaction process of the trefoil knot was observed at low forces. The folding of TrmD strictly depends on the conformation of the unfolded TrmD. If the unfolded TrmD contains a pre-formed knot (tightened or not), the folding of TrmD is fast and robust. If the unfolded TrmD does not contain a pre-formed knot, the knotting is the rate-limiting step for the folding. Our study provided detailed mechanistic insights into the folding and unfolding of TrmD, and revealed the intricate interplay between the folding and knotting process of TrmD.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work is supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. H. W. acknowledges the fellowship support from the NSERC NanoMat CREATE program.

## Notes and references

- 1 C. Liang and K. Mislow, *J. Am. Chem. Soc.*, 1994, **116**, 11189.
- 2 M. L. Mansfield, *Nat Struct Biol*, 1994, **1**, 213.
- 3 F. Takusagawa and S. Kamitori, *J. Am. Chem. Soc.*, 1996, **118**, 8945.
- 4 W. R. Taylor, *Nature*, 2000, **406**, 916.
- 5 M. Jamroz, W. Niemyska, E. J. Rawdon, A. Stasiak, K. C. Millett, P. Sulkowski and J. I. Sulkowska, *Nucleic Acids Res.*, 2015, **43**, D306.
- 6 D. Bolinger, J. I. Sulkowska, H. P. Hsu, L. A. Mirny, M. Kardar, J. N. Onuchic and P. Virnau, *PLoS Comput. Biol.*, 2010, **6**, e1000731.
- 7 P. Virnau, A. Mallam and S. Jackson, *J. Phys. Condens. Matter*, 2011, **23**, 033101.
- 8 F. Ziegler, N. C. Lim, S. S. Mandal, B. Pelz, W. P. Ng, M. Schlierf, S. E. Jackson and M. Rief, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 7533.
- 9 P. Dabrowski-Tumanski and J. I. Sulkowska, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 3415.
- 10 T. C. Sayre, T. M. Lee, N. P. King and T. O. Yeates, *Protein Eng. Des. Sel.*, 2011, **24**, 627.
- 11 J. Especial, A. Nunes, A. Rey and P. F. Faisca, *Phys. Chem. Chem. Phys.*, 2019, **21**, 11764.
- 12 M. K. Sriramoju, Y. Chen, Y. C. Lee and S. D. Hsu, *Sci. Rep.*, 2018, **8**, 7076.
- 13 J. I. Sulkowska, *Curr. Opin. Struct. Biol.*, 2020, **60**, 131.
- 14 A. L. Mallam and S. E. Jackson, *J. Mol. Biol.*, 2005, **346**, 1409.
- 15 A. L. Mallam, *FEBS J.*, 2009, **276**, 365.
- 16 T. Bornschlogl, D. M. Anstrom, E. Mey, J. Dzubiella, M. Rief and K. T. Forest, *Biophys. J.*, 2009, **96**, 1508.
- 17 A. L. Mallam and S. E. Jackson, *J. Mol. Biol.*, 2006, **359**, 1420.
- 18 M. Rivera, Y. Hao, R. A. Maillard and M. Baez, *Sci. Rep.*, 2020, **10**, 9562.
- 19 S. P. Elmer and V. S. Pande, *J. Chem. Phys.*, 2005, **122**, 124908.
- 20 J. I. Sulkowska, P. Sulkowski and J. Onuchic, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 3119.
- 21 J. K. Noel, J. I. Sulkowska and J. N. Onuchic, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 15403.
- 22 S. a Beccara, T. Skrbic, R. Covino, C. Micheletti and P. Faccioli, *PLoS Comput. Biol.*, 2013, **9**, e1003002.
- 23 J. I. Sulkowska, J. K. Noel and J. N. Onuchic, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 17783.
- 24 Y. C. Lee, C. Y. Chang, S. Y. Chen, Y. R. Pan, M. R. Ho and S. D. Hsu, *Sci. Rep.*, 2017, **7**, 45174.
- 25 Y. C. Lee and S. D. Hsu, *J. Phys. Chem. Lett.*, 2018, **9**, 2433.
- 26 P. A. Elkins, J. M. Watts, M. Zalacain, A. van Thiel, P. R. Vitazka, M. Redlak, C. Andraos-Selim, F. Rastinejad and W. M. Holmes, *J. Mol. Biol.*, 2003, **333**, 931.
- 27 T. Ito, I. Masuda, K. Yoshida, S. Goto-Ito, S. Sekine, S. W. Suh, Y. M. Hou and S. Yokoyama, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E4197.
- 28 H. J. Ahn, H. W. Kim, H. J. Yoon, B. I. Lee, S. W. Suh and J. K. Yang, *EMBO J.*, 2003, **22**, 2593.
- 29 M. Redlak, C. Andraos-Selim, R. Giege, C. Florentz and W. M. Holmes, *Biochemistry*, 1997, **36**, 8699.
- 30 S. Goto-Ito, T. Ito and S. Yokoyama, *Biomolecules*, 2017, **7**, 32.
- 31 Y. M. Hou, R. Matsubara, R. Takase, I. Masuda and J. I. Sulkowska, *Enzymes*, 2017, **41**, 89.
- 32 C. He, G. Lamour, A. Xiao, J. Gsponer and H. Li, *J. Am. Chem. Soc.*, 2014, **136**, 11946.
- 33 C. He, S. Li, X. Gao, A. Xiao, C. Hu, X. Hu, X. Hu and H. Li, *Nanoscale*, 2019, **11**, 3945.
- 34 H. Wang, X. Gao, X. Hu, X. Hu, C. Hu and H. Li, *Biochemistry*, 2019, **58**, 4751.
- 35 C. He, G. Z. Genchev, H. Lu and H. Li, *J. Am. Chem. Soc.*, 2012, **134**, 10428.
- 36 M. L. Bennink, O. D. Scharer, R. Kanaar, K. Sakata-Sogawa, J. M. Schins, J. S. Kanger, B. G. de Grooth and J. Greve, *Cytometry*, 1999, **36**, 200.
- 37 C. Cecconi, E. A. Shank, F. W. Dahlquist, S. Marqusee and C. Bustamante, *Eur. Biophys. J.*, 2008, **37**, 729.
- 38 J. C. Gebhardt, T. Bornschlogl and M. Rief, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2013.
- 39 G. Zoldak and M. Rief, *Curr. Opin. Struct. Biol.*, 2013, **23**, 48.
- 40 X. Zhang, K. Halvorsen, C. Z. Zhang, W. P. Wong and T. A. Springer, *Science*, 2009, **324**, 1330.
- 41 Z. Xu, A. Colosimo and D. C. Gruenert, *Methods Mol. Biol.*, 2003, **235**, 203.
- 42 Y. Cao and H. Li, *Nat. Mater.*, 2007, **6**, 109.
- 43 H. Lei, C. He, C. Hu, J. Li, X. Hu, X. Hu and H. Li, *Angew. Chem. Int. Ed.*, 2017, **56**, 6117.
- 44 S. B. Smith, Y. Cui and C. Bustamante, *Methods Enzymol.*, 2003, **361**, 134.
- 45 <http://tweezerslab.unipr.it/cgi-bin/mt/home.pl>.
- 46 K. Liu, X. Chen and C. M. Kaiser, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 25641.
- 47 L. Oberbarnscheidt, R. Janissen and F. Oesterhelt, *Biophys. J.*, 2009, **97**, L19.



- 48 J. Brujic, R. I. Hermans, S. Garcia-Manyes, K. A. Walther and J. M. Fernandez, *Biophys. J.*, 2007, **92**, 2896.
- 49 E. Evans, *Annu. Rev. Biophys. Biomol. Struct.*, 2001, **30**, 105.
- 50 G. I. Bell, *Science*, 1978, **200**, 618.
- 51 G. Zoldak, J. Stigler, B. Pelz, H. Li and M. Rief, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 18156.
- 52 M. Mickler, R. I. Dima, H. Dietz, C. Hyeon, D. Thirumalai and M. Rief, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 20268.
- 53 Q. Peng and H. Li, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 1885.
- 54 A. L. Mallam and S. E. Jackson, *Nat. Chem. Biol.*, 2011, **8**, 147.

