





Transmembrane protein rotaxanes reveal kinetic traps in the refolding of translocated substrates

Jianfei Feng¹, Pablo Martin-Baniandres¹, Michael J. Booth ¹, Gianluca Veggiani², Mark Howarth ², Hagan Bayley ¹✉ & David Rodriguez-Larrea ³✉

Understanding protein folding under conditions similar to those found in vivo remains challenging. Folding occurs mainly vectorially as a polypeptide emerges from the ribosome or from a membrane translocon. Protein folding during membrane translocation is particularly difficult to study. Here, we describe a single-molecule method to characterize the folded state of individual proteins after membrane translocation, by monitoring the ionic current passing through the pore. We tag both N and C termini of a model protein, thioredoxin, with biotinylated oligonucleotides. Under an electric potential, one of the oligonucleotides is pulled through a α -hemolysin nanopore driving the unfolding and translocation of the protein. We trap the protein in the nanopore as a rotaxane-like complex using streptavidin stoppers. The protein is subjected to cycles of unfolding-translocation-refolding switching the voltage polarity. We find that the refolding pathway after translocation is slower than in bulk solution due to the existence of kinetic traps.

¹Department of Chemistry, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, UK. ²Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. ³Department of Biochemistry and Molecular Biology (UPV/EHU), Biofisika Institute (CSIC, UPV/EHU), Barrio Sarriena s/n, 48940 Leioa, Spain. ✉email: hagan.bayley@chem.ox.ac.uk; david.rodriguez@ehu.es

Protein folding has attracted much attention during the last 50 years, because it is central to cell function and disease. Early studies examined how proteins folded *in vitro*, typically after chemical or thermal denaturation^{1,2}. More recently, technical developments have allowed a deeper understanding of protein folding under experimental conditions closer to those found *in vivo*^{3–7}. A relevant event within the cell is the emergence of a polypeptide from the ribosome exit tunnel^{5,8}. The nascent chain emerges N terminus-first and, before the C terminus is complete, the protein may start folding^{9,10}, which is a strikingly different process to experimental refolding in bulk solution⁴. Second, during or after synthesis by ribosomes, proteins may be transported between compartments in the cell¹¹ or into the extracellular medium¹². During these processes, they must cross membranes by using one of a diverse set of translocons. For example, in bacteria alone, there are at least ten secretion systems^{13,14} and translocated substrates include many of the virulence factors secreted by pathogens¹⁴. Usually, secretion systems translocate unfolded polypeptides, and therefore the substrate must refold after or during the translocation process^{15–17}.

Currently, pulse-chase experiments, which have a temporal resolution of minutes, are often used to study the folded state of proteins following membrane translocation in cells^{18,19}. Similar temporal resolution is available when experiments are performed *in vitro* with reconstituted translocon components²⁰. Further, these ensemble experiments frequently fail to resolve the elementary steps of complex processes²¹. Studies of translocation to the periplasm of substrates that contain disulfides have been a remarkable exception¹². By using a mutant of DsbA, a disulfide-bond-forming enzyme, the formation of covalent mixed-disulfide complexes between DsbA and a polypeptide substrate was slowed, which allowed the resolution of intermediates in the folding process. However, this approach is limited to disulfide-containing substrates in bacteria.

We recently introduced a method to study co-translocational protein unfolding at the single-molecule level^{22–25}, which avoids the requirement to synchronize an ensemble measurement. To achieve this, a single α -hemolysin pore (α HL) is inserted into a lipid bilayer. The α HL pore forms a conduit of ~ 2 nm diameter (~ 1.3 nm at its narrowest point), which can accommodate alpha helices but forces a protein to pass unfolded²⁶. A model protein, thioredoxin (Trx), derivatized at one of its termini with a negatively-charged oligonucleotide, is driven into the pore by an electric field, which causes unfolding and translocation of the protein^{22,23}. Steps in the ionic current carried by the pore represent intermediates in the translocation process²⁷. In an alternative approach, a signal sequence is placed at the C terminus of a polypeptide and recognized by the molecular motor ClpX, which drives translocation^{24,25}. Both approaches allow real-time tracking of the unfolding and translocation process. However, following translocation, the protein substrate is released into the *trans* compartment and cannot be further examined to determine whether it has refolded.

Here, we generate a rotaxane-like structure that prevents release of the protein substrate into the medium²⁸ and allows examination of the folding status of the translocated substrate. The overall process involves unfolding, translocation and refolding of the protein substrate. Devoid of molecular motors and chaperones, our model system constitutes a simplification of the more complex *in vivo* process but has the vectorial nature of protein folding as *in vivo*. We first attached 3'-biotinylated oligonucleotides to each terminus of Trx. The oligonucleotide-Trx-oligonucleotide conjugate with one end bound to monovalent streptavidin (mSA)²⁹ is driven into the nanopore by an applied potential. The direction of translocation (N-to-C or C-to-N) is ascertained by using distinguishable oligonucleotides. The

oligonucleotide that reaches the *trans* side of the bilayer is captured by a second streptavidin (SA). The interlocked architecture of the threaded complex bound to one streptavidin at each side of the membrane constitutes a transmembrane protein rotaxane and is stable until either streptavidin dissociates. The protein substrate can then be driven from one side of the membrane to the other by repeatedly switching the polarity of the voltage between positive and negative. Our approach allows an examination of the evolution of the folding landscape of Trx, beginning within milliseconds after translocation.

Results

Identification of the terminus leading protein translocation.

We used a mutant of *E. coli* thioredoxin 2 (PDB code 2TRX) lacking the natural disulfide bond (C32S-C35S), with three mutations that increase stability (A22P-I23V-P68A)³⁰ and both an N-terminal and a C-terminal cysteine (S1C and the added C109). This mutant, V5, was selectively modified on the N-terminal cysteine by using a click reaction with a 2-cyanobenzothiazole (CBT)³¹ to form a stable thiazoline ring³². To introduce the CBT functionality into a thiol-modified oligonucleotide, we coupled the amino group of 6-amino-2-CBT to *p*-maleimidophenyl isocyanate (Supplementary Fig. 1a, b). The product, a maleimide-CBT crosslinker, was reacted with the 5'-thiol of a 3'-biotinylated oligonucleotide composed of 30 cytosines, and the resultant 5'-CBT-oligo(dC)₃₀-biotin-3' was purified by ion-exchange chromatography (Fig. 1a and Supplementary Fig. 1c–f).

Following bacterial expression and purification of S1C-V5-C109, more than half of the N-terminal cysteines were found modified with pyruvate³³, which was removed by incubation with methoxyamine exposing the 1,2-aminothiols (Supplementary Fig. 2). Subsequent incubation of the protein, reduced with TCEP, with 5'-CBT-oligo(dC)₃₀-biotin-3', followed by ion-exchange chromatography, produced a product that gave a single band of ~ 20 kDa upon electrophoresis (Supplementary Fig. 3a, d, f), consistent with a single oligonucleotide coupled to the N terminus of S1C-V5-C109³¹. Because CBT reacts site-specifically with 1,2-aminothiols, the C-terminal cysteine was not altered. The conjugate was then reacted at the C-terminal cysteine with the 5'-thiol of a 3'-biotinylated oligonucleotide, composed of 30 adenines, that had been activated with 2,2'-dipyridyl disulfide (Supplementary Fig. 3b, c)³⁴. The purified product produced a band of ~ 30 kDa upon electrophoresis (Supplementary Fig. 3f), and the native mass spectrometry (native MS) was consistent with S1C-V5-C109 derivatized with different oligonucleotides at the N and C termini (hereafter V5 stands for the S1C-A22P-I23V-P68A-C32S-C35S-C109 mutant, and oligo(dC)₃₀-V5-oligo(dA)₃₀ for the mutant modified with the oligonucleotides, Supplementary Fig. 3g). One of the amide bonds of the maleimide group was hydrolyzed to completion during the synthesis of oligo(dC)₃₀-V5-oligo(dA)₃₀, which produced a +18 peak in the MS but left the second amide and the linkage intact.

We added oligo(dC)₃₀-V5-oligo(dA)₃₀ to the *cis* side of a phospholipid bilayer containing a single α HL pore. Under an applied potential of +70 to +140 mV, the construct was driven into the pore, by either of the two oligonucleotides, causing the unfolding and translocation of the protein. Previously we have shown that an oligonucleotide tag at the C-terminus does not modify the stability of the native state as judged by the similar urea induced unfolding curve, but it only serves to drive the translocation²². The ionic current passing through the pore revealed the various molecular steps of the process²³ (Fig. 1b–d). First the leading oligonucleotide, oligo(dC)₃₀ for N terminus-first or oligo(dA)₃₀ for C terminus-first, threaded into the pore (level 2 in Fig. 1e). The ionic current differed depending on the

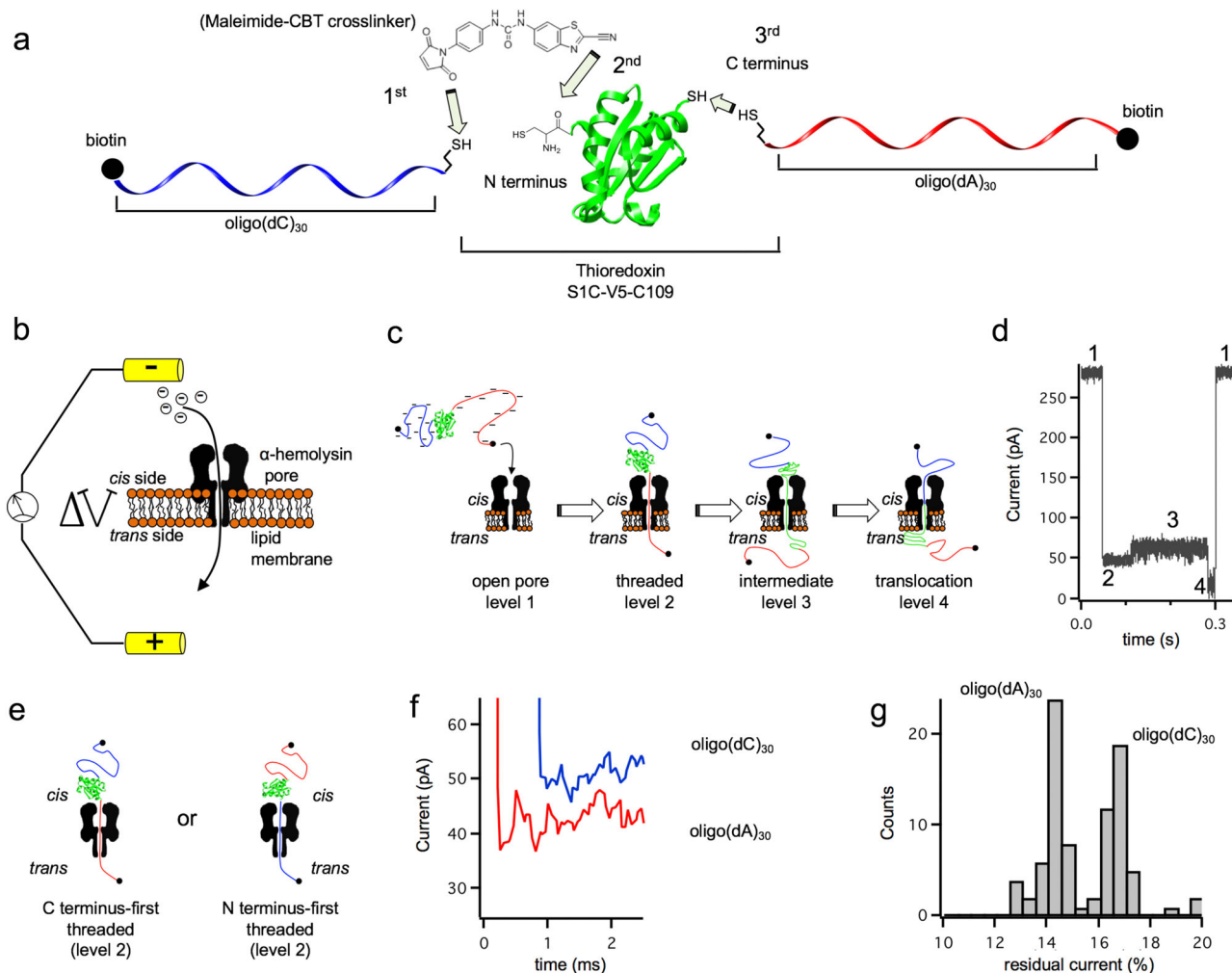
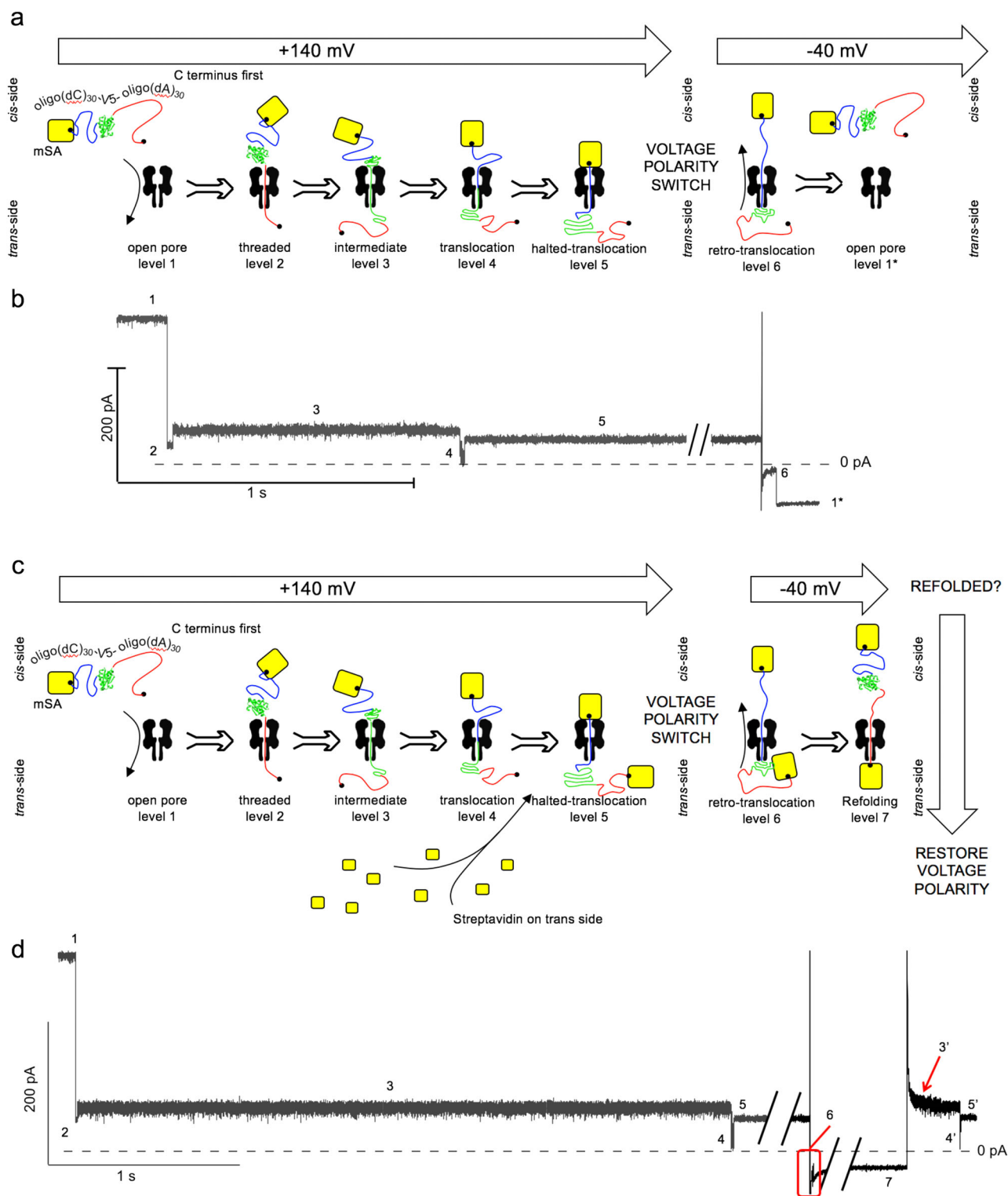


Fig. 1 Differential tagging and co-translocational unfolding of Trx V5. **a** Synthesis of oligo(dC)₃₀-V5-oligo(dA)₃₀. After each of the three steps, the product was purified and characterized by SDS-PAGE and MS (Supplementary Figs. 1–3). **b** An electric potential across a lipid bilayer containing a single α HL pore (*cis* at ground). In response to an applied positive potential, anions and cations are driven through the pore, which is recorded as a positive current. **c** When oligo(dC)₃₀-V5-oligo(dA)₃₀ is added to the *cis* compartment at +140 mV, one of the negatively-charged oligonucleotides is pulled into the α HL pore (level 2). The electrophoretic force acting on the oligonucleotide partly unfolds the protein (level 2→level 3). A relatively long-lived intermediate is generated in which the terminus of the polypeptide chain threaded through the pore (level 3). At this point, the remaining folded portion of Trx is unstable and, after a waiting time (level 3→level 4), it spontaneously unfolds and diffuses through the pore (level 4). **d** Electrical recording of a single translocation event showing levels 1–4. **e** Oligo(dC)₃₀-V5-oligo(dA)₃₀ can thread either N or C terminus-first. **f** Ionic current signals of level 2 for entry with the N terminus-first (blue) and C terminus-first (red). **g** Event histogram of the residual currents ($I_{res\%}$) in level 2 during multiple translocation events of oligo(dC)₃₀-V5-oligo(dA)₃₀ through a single α HL pore showing two populations. The $I_{res\%}$ in level 2 was used to define the direction of translocation. The result was confirmed with two additional pores.

composition of the oligonucleotide, with oligo(dC)₃₀ giving a higher residual conductance value than oligo(dA)₃₀ ($I_{res\%} = 16.5 \pm 0.5\%$ ($n = 36$) and $I_{res\%} = 14.2 \pm 0.2\%$ ($n = 43$) at +140 mV, respectively, Fig. 1f, g), in accord with previous studies with immobilized oligonucleotides^{35,36}. The ionic current during level 2 can therefore be used to determine the orientation in which oligo(dC)₃₀-V5-oligo(dA)₃₀ threads through the pore for each event that is registered. After threading, the protein unfolded by way of a long-lived intermediate (level 3 in Fig. 1c, d) that spontaneously unfolded and then passed through the pore by diffusion²² (level 4 in Fig. 1c, d). Once the leading oligonucleotide has entered the *trans* compartment and left the electric field, there is no longer a pulling force on the protein. Following the translocation of the remaining unfolded protein, the pore remained open until another oligo(dC)₃₀-V5-oligo(dA)₃₀ entered. The terminal oligonucleotide translocated too rapidly to generate a distinguishable ionic current

level. The dwell time and ionic current of level 3 also depended on the direction of translocation, being longer and noisier if co-translocational unfolding is initiated through the N terminus, as already reported²³. These features can also be used to determine the direction in which oligo(dC)₃₀-V5-oligo(dA)₃₀ initiates translocation, and the directionality of pore entry based on the $I_{res\%}$ values in level 2 are in agreement with the features observed in level 3. We study the C-first translocation reaction because when the protein is pulled back the protein translocates N-terminus first and therefore vectorial refolding is as in most biological systems. For these molecules, the dwell time in level 2 reflects the unfolding kinetics of a C-terminal region^{22,23}, and the dwell time in level 3 reflects the unfolding kinetics of the partly unfolded intermediate (Fig. 1c, d). The rate constants determined here ($k_{2\rightarrow3} = 6.9 \pm 0.5 \text{ s}^{-1}$ and $k_{3\rightarrow4} = 1.0 \pm 0.1 \text{ s}^{-1}$ at +100 mV, $n = 160$, Supplementary Fig. 8a, b) are in good agreement with



those reported²² for the construct C1S-V5-C109-oligo(dC)₃₀ ($k_{2 \rightarrow 3} = 6.0 \pm 1.0 \text{ s}^{-1}$ and $k_{3 \rightarrow 4} = 1.5 \pm 0.1 \text{ s}^{-1}$ at +100 mV). These results suggest that simultaneous N-terminal and C-terminal modification have a minimal effect on the kinetics of co-translocational protein unfolding.

Formation of a transmembrane protein rotaxane. Oligo(dC)₃₀-V5-oligo(dA)₃₀ present in the *cis* compartment was then treated

with a sub-stoichiometric amount of mSA. Each of the two oligonucleotides carried a terminal biotin and, therefore, a significant fraction of the construct bound only a single mSA²⁹, leaving the other terminus free. When these complexes were pulled into the pore by voltage (+100 to +140 mV), they produced the characteristic ionic current signature that represents threading, unfolding and translocation of the protein (Fig. 2a, b). However, following translocation, the ionic current signal does not recover to the open pore level but remains at a new level (level 5,

Fig. 2 Formation of an oligo(dC)₃₀-V5-oligo(dA)₃₀/αHL rotaxane. **a** Prevention of release into the *trans* compartment. Oligo(dC)₃₀-V5-oligo(dA)₃₀ (biotinylated oligonucleotides) was treated with a sub-stoichiometric amount of mSA. A complex with one of the two oligonucleotides bound to mSA threads through the pore at +140 mV with the C terminus free oligonucleotide first. After unfolding and translocation of the protein (Trx V5), the translocated substrate remains arrested in the pore by the mSA. Reversal of the voltage polarity (−40 mV) drives the complex back into the *cis* compartment (retro-translocation), where it is released to the solution. The pore is left at level 1* (the open state at a negative potential). **b** Ionic current signal corresponding to the events shown in 'a'. Level 1, open αHL pore; level 2, oligonucleotide threaded into the pore; level 3, unfolding intermediate; level 4, translocation (largely diffusive) of the unfolded protein; level 5, second oligonucleotide threaded into the pore; level 6, retro-translocation; level 1*, open pore (−40 mV). **c** Formation of a transmembrane protein rotaxane. The steps described in 'a' were carried out with SA to the *trans* compartment. After the protein translocation, an approximately 5 min delay allowed SA to bind to the biotinylated oligonucleotide. Then, after a voltage step to −40 mV, the protein was retro-translocated but remained associated with the αHL pore as a rotaxane. After a delay, a subsequent step to +140 mV can be used to probe the folded state of the Trx V5 in the *cis* compartment. **d** Ionic current signal characteristic of rotaxane formation. Level 1, open αHL pore (+140 mV); level 2, oligonucleotide threaded into the pore; level 3, unfolding intermediate; level 4, translocation of the unfolded protein; level 5, second oligonucleotide threaded into the pore; level 6, retro-translocation; level 7, initial leading oligonucleotide threaded into the pore. During level 7 (here 0.49 s) the protein refolds in the *cis* compartment and when the potential is stepped back to +140 mV, levels 2', 3', 4', and 5' appear. The prime symbol signifies that the levels are generated not from the native Trx, but from a refolded form.

in Fig. 2a, b). The residual current in level 5 also depended on whether the process was initiated N or C terminus-first, with $I_{\text{res}\%} = 17.8 \pm 0.1\%$ ($n = 33$, +140 mV) and $I_{\text{res}\%} = 17.4 \pm 0.1\%$ ($n = 32$, +140 mV), respectively (Supplementary Fig. 4). During level 5, mSA on the *cis* side atop the pore stops the movement of the construct, with the terminal oligonucleotide threaded into the pore³⁵. The immobilized oligonucleotides were in the 5'-end-first orientation in level 5, and the higher $I_{\text{res}\%}$ values for the oligo(dA)₃₀ blockades than for those arising from the oligo(dC)₃₀ blockades, opposite to the result in level 2 (3'-end-first), is in agreement with previous work^{35,36}.

The protein was hence in the *trans* compartment, and when the potential was changed from +140 mV to −40 mV the complex was pulled in the opposite direction, from *trans* to *cis*, i.e., it was forced to retro-translocate. At the beginning of this voltage step, the ionic current signal moved transiently to a new level 6, before the characteristic ionic current level of the open pore at −40 mV was obtained (level 1* in Fig. 2a, b, the asterisk signifies the negative potential). Level 6 therefore represents the unfolding and translocation in the *trans* to *cis* direction.

Alternatively, after oligo(dC)₃₀-V5-oligo(dA)₃₀ threaded C terminus-first, instead of stepping the voltage to −40 mV we added mSA to the *trans* compartment, which bound to the biotinylated oligo(dA)₃₀ that had led the translocation process (Fig. 2c). Hence, a rotaxane-like structure was generated (level 5 in Fig. 2c, d)²⁸. In this state, a voltage step from +140 to −40 mV did not produce the current level of the open pore (level 1*), but instead the signal assumed a new level 7 in which the translocating Trx V5 is presumed to have moved completely back into the *cis* compartment, and the C-terminal oligonucleotide is threaded through the pore. After a selected time in level 7, the potential was switched back to +140 mV, and we observed an ionic current pattern similar to that produced during the initial translocation, with levels 2', 3', 4', and 5' (Fig. 2d). Importantly, the time provided in level 7 was time available to allow the protein to refold, and the dwell times during subsequent forward re-translocation (levels 2' and 3') revealed the kinetic stability of the refolded state. Finally, and in agreement with the proposed model, when the voltage was switched from +80 mV (a lower voltage to produce an extended level 2') to −40 mV during residence in level 2' or 3' (i.e., while the protein was still or largely still in the *cis* compartment), level 6' was not observed, but rather level 7 reappeared (Supplementary Fig. 5a, c). Similarly, if the voltage was switched from −40 mV to +80 mV during level 6' (i.e., while the protein was largely in the *trans* compartment), levels 2', 3' and 4' were never observed (Supplementary Fig. 5a, d).

These results are consistent with the formation of a transmembrane protein rotaxane. In this state, the protein can be translocated forwards and backwards across the pore by changing the polarity of the applied potential (Fig. 2c) and thereby subjected to multiple cycles of unfolding, translocation and refolding, observable with millisecond time resolution. Yet higher temporal resolution is challenging, because of capacitive transients after the voltage steps. In addition, the use of monovalent traptavidin (mTA)³⁷ in the *cis* compartment, which dissociates from biotin more than ten-times slower than mSA, and traptavidin (TA) in the *trans* compartment, allowed over 24 h of recording on the same individual rotaxane.

The refolding landscape of a translocated protein substrate. We next aimed to define the folding landscape of Trx after translocation through the αHL pore, by subjecting the protein in the rotaxane to cycles of unfolding, translocation and refolding (Fig. 3a, b). In this work, we analyze constructs that enter C terminus-first from the *cis* compartment. Therefore, the refolding we investigate is initiated during or after retro-translocation N terminus-first from the *trans* into the *cis* compartment.

The protein may refold outside the pore during level 7' (Fig. 3a). We held the protein at level 7' for a selected period (the refolding time) and then switched to a positive potential (+100 mV) to probe the state of the protein³⁸, which at this point could have returned to the native state, remained in an unfolded state or be in an intermediate state. By comparing the dwell times in levels 2' and 3' to the corresponding dwell times during the initial unfolding of the native protein (levels 2 and 3), we analyzed the stability of individual refolded Trx molecules. The dwell time in level 4', which represents diffusive translocation of the remaining unfolded polypeptide, was unaffected by the refolding time and was similar to that of level 4 of the native protein (Supplementary Fig. 6). We allowed the protein to refold at level 7' for 0.09 s, 0.49 s, 0.99 s, 1.49 s, 10 s, 100 s, or 1000 s, and probed the state of individual Trx molecules hundreds of times under each condition.

When we allowed the translocated Trx V5 protein molecule to refold for 90 ms, levels 2' and 3' were observed in only 99 out of 365 re-translocation events (Fig. 3c). In the remaining cases, level 2' (which has been previously related to the folding of ~40 aminoacids at the C terminus^{22,23}), level 3' (previously related to the fold of the remainder of the polypeptide chain^{22,23}) or both were absent (Supplementary Fig. 7). The absence of these levels either represents a lack of folding or the formation of weakly folded regions that unfold faster than the detection limit (~0.2 ms). Importantly, we observed that the longer the refolding time,

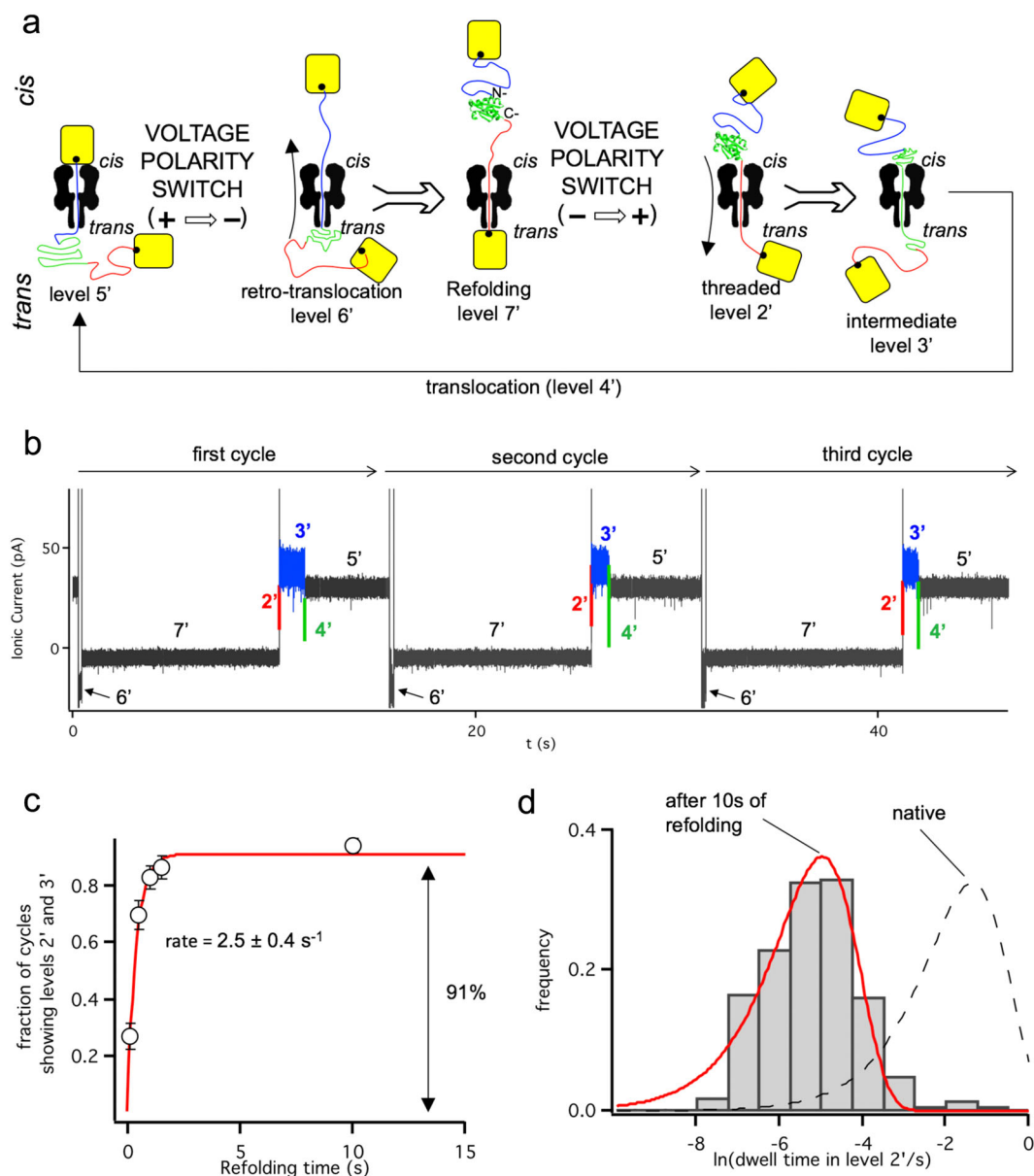


Fig. 3 Employing transmembrane protein rotaxanes to analyze protein refolding. **a** Experimental configuration. Once the rotaxane has been formed (level 5'), the protein is moved from the *trans* to the *cis* side by applying a negative potential (level 6' → level 7'). The retro-translocated substrate remains in level 7' for a specified refolding time, which is computer controlled. Subsequently, the potential is switched to +100 mV, and the co-translocational unfolding of the refolded protein is recorded (level 2' → 3' → 4' → 5'). This process is then repeated. **b** Ionic current signal from three cycles as described in 'a' with a refolding time of 10 s. **c** The fraction of cycles in which the refolded substrate shows both levels 2' and 3', characteristic of refolding to a native-like state, increases with refolding time. The increase of the folded fraction fits a single exponential (red) yielding a rate constant of $2.5 \pm 0.4 \text{ s}^{-1}$. Each point represents data from ~300 translocation events acquired with three different pores. **d** Frequency histogram of the dwell times in level 2' observed during 345 cycles with a refolding time of 10 s (data collected with three different pores). The data are well described by an exponentially-distributed population of dwell times (red solid line, $k_{2 \rightarrow 3} = 145 \pm 10 \text{ s}^{-1}$; black dashed line: probability density function for the native state, $k_{2 \rightarrow 3} = 6.9 \pm 0.5 \text{ s}^{-1}$).

the more frequently the signal contained both levels 2' and 3' (Fig. 3c). The time course of this folding process (measured as the presence of both levels 2' and 3') was not limited by the translocation of the unfolded polypeptide (that is ~50 fold slower than the folding event, Supplementary Fig. 6) and well described by a single exponential function, yielding a rate of $2.5 \pm 0.4 \text{ s}^{-1}$.

After 10 s of refolding, a plateau was reached where ~90% of the re-translocation events showed both levels 2' and 3'. Nonetheless, the co-translocational unfolding of the refolded Trx V5 was faster than that of the Trx V5 that was unfolded for the first time (Fig. 3d). We observed a shorter dwell time in level 2', which converted to level 3' with a rate constant of

$k_{2 \rightarrow 3} = 250 \pm 10 \text{ s}^{-1}$ (at +100 mV, $n = 345$), compared to the value for the native Trx V5 of $k_{2 \rightarrow 3} = 6.9 \pm 0.5 \text{ s}^{-1}$ (at +100 mV, $n = 160$). The dwell time in level 3' was also shorter, but the difference was less obvious (less than 7-fold reduction, Supplementary Fig. 8).

These results imply that Trx V5 had not reached the native state after 10 s of refolding. The folding kinetics of Trx are complicated because there is an invariant *cis* proline, Pro-76, buried in the native structure. In bulk solution, when Trx is placed under denaturing conditions for a few seconds, proline residues do not have time to isomerize and refolding is fast (sub-millisecond)^{39,40}. When Trx is left under denaturing conditions

for hundreds of seconds, proline isomerization occurs and Trx then refolds slowly ($t_{1/2} \sim 500$ s)⁴¹. In this respect, the unfolding and translocation of Trx V5 through the pore typically takes less than ~ 1.5 s. This gives little time for proline isomerization to occur (proline isomerization takes from 10^3 s *-cis* to *-trans* to 10^4 s *-trans* to *-cis*⁴², roughly once every 1000 cycles), and we therefore expected refolding to occur in the sub-millisecond regime. Given that 10 s after translocation the refolded Trx V5 remains in a state which unfolds faster than the native Trx V5, we conclude that the folding mechanism differs from the fast folding route that occurs in bulk solution.

We next used oligo(dC)₃₀-V5-oligo(dA)₃₀ to explore the folded state of Trx V5 after prolonged folding times. We first confirmed that Trx V5 mutant is a slow folder in bulk solution by measuring the recovery of tryptophan fluorescence after dilution from 4 M guanidinium chloride. Refolding proceeded at $2.3 \times 10^{-3} \text{ s}^{-1}$, a similar rate to that of wild-type Trx^{30,39-41}, showing that in these cases folding is rate-limited by proline isomerization (we could not use the tagged substrate because we do not generate enough sample). By contrast, after retro-translocation through the nanopore and a 1000 s delay for refolding, we only observed $\sim 15\%$ of the molecules unfolding with dwell times in levels 2' and 3' that were characteristic of the native state (Supplementary Fig. 8). This yields a refolding rate of $1.4 \times 10^{-4} \text{ s}^{-1}$, >15 -fold slower than that observed in bulk solution, which unlike refolding upon translocation requires proline isomerization. In atomic force microscopy and optical tweezer experiments, tethering to a

surface, a cantilever or a bead slows down folding due to dragging forces⁴³. We expect this effect to be negligible in our case: one terminus of the protein is anchored to the pore embedded in the lipid bilayer and the other is tethered to an oligonucleotide-streptavidin complex, neither of which produces a strong drag. On the other hand, the N-terminal and C-terminal of Trx are fully exposed to the solvent, suggesting that the addition of the tags should have a minor effect on the kinetics. Therefore, refolding after protein translocation conforms with neither the fast folding nor the slow folding pathways observed for Trx in bulk solution.

An advantage of our approach is that a given refolded state can be characterized with at least as many variables as the number of observable unfolding steps, which allows the use of bivariate scatter plots to improve the ability to distinguish different populations and better resolve the folding landscape. We display dwell times in levels 2' and 3' for the refolding times of 0.09 s, 0.49 s, 0.99 s, 1.49 s, 10 s, 100 s, and 1000 s acquired from three different rotaxanes and the superimposed data to better visualize the identified populations (Fig. 4a). The plots are snapshots of how the protein moves through the folding landscape. Trx V5 emerged unfolded from the pore, and after 0.09 s, $\sim 73\%$ of the events examined lacked any detectable folding (level 2' or 3'). The remainder ($\sim 27\%$) had entered one of two states, population A or population B, both less stable than the native state based on the shorter dwell times in levels 2' and 3'. During the first second (0.99 s), the fraction of events with undetectable folding was

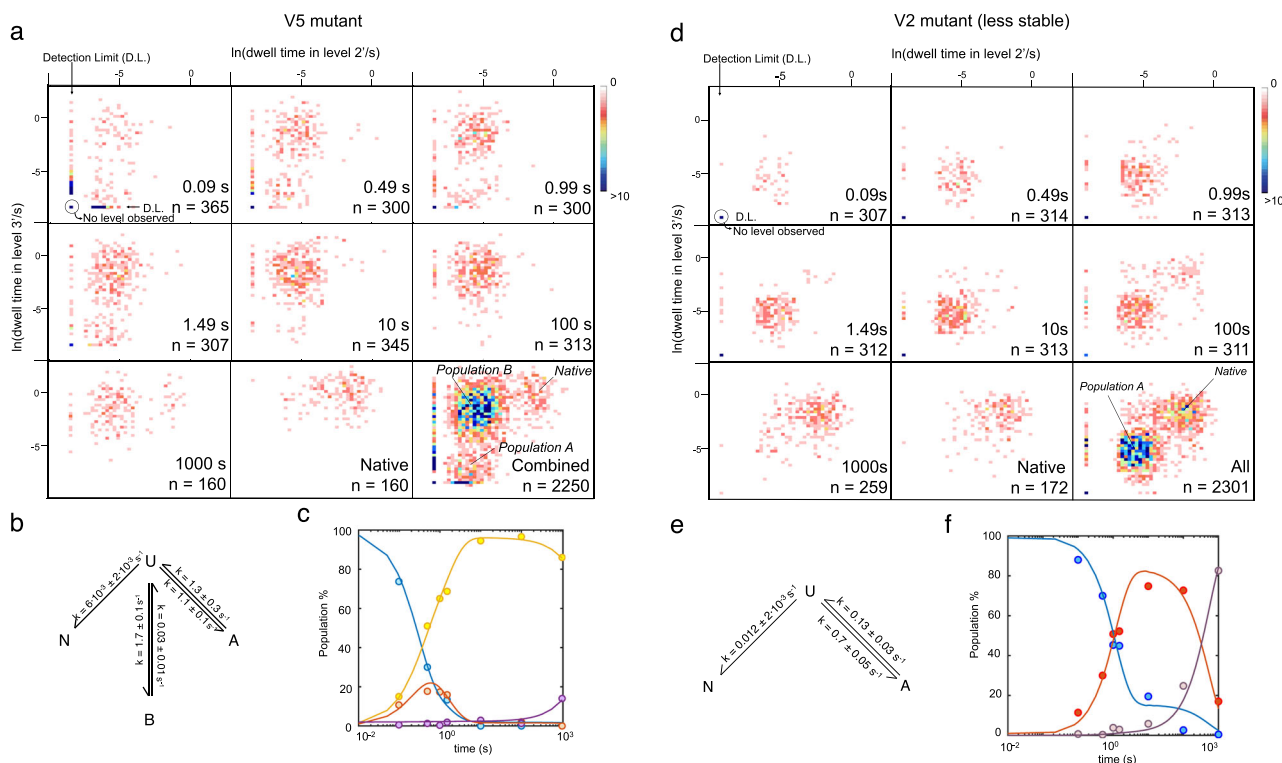


Fig. 4 Folding landscapes of translocated protein substrates. **a** Bivariate histograms representing the dwell times in levels 2' and 3' for oligo(dC)₃₀-V5-oligo(dA)₃₀ threaded C terminus-first. Data were acquired with three different pores. Each panel corresponds to the indicated refolding time (n: number of events). When level 2' or 3' was not observed, a dwell time of 0.2 ms, the temporal detection limit, was assigned. The "Native" panel records values obtained from Trx V5 that had never been unfolded. The "All" panel displays the data from all the other panels in a single plot to provide clearer visualization of three populations. **b** Kinetic model giving the lowest AIC and BIC values with the kinetic constants derived from the fit. **c** Time-course evolution of population A (red circles), population B (yellow circles), native Trx V5 (purple circles) and unfolded Trx V5 (blue circles). The solid lines represent the best fits to the model with the lowest AIC and BIC values. **d** Three destabilizing mutations were introduced into Trx V5 to form Trx V2. A rotaxane was formed, and the refolding was examined as in 'a'. When level 2' or 3' was not observed, a dwell time of 0.2 ms was assigned. **e** Kinetic model giving the lowest AIC and BIC values with the kinetic constants derived from the fit. **f** Time-course evolution of population A (red circles), native Trx V2 (purple circles) and unfolded Trx V2 (blue circles). The solid lines represent the best fits.

reduced drastically to ~17%, while population A remained constant (from ~10% at $t = 0.09$ s to ~17% at $t = 0.99$ s) and population B increased (from ~15% at $t = 0.09$ s to ~65% at $t = 0.99$ s). As the population with undetectable folding disappeared, so did population A. After 10 s of refolding, most of the events were found within population B, and the native state became slowly populated, reaching ~15% at 1000 s.

To set our analysis on a more quantitative basis, we used k -means clustering⁴⁴ to classify the events into population A, population B, or the native state (Supplementary Fig. 9). When either level 2', level 3' or both were absent, the events were assigned to an unfolded state. By using this approach, we found that the native state was characterized by $k_{2' \rightarrow 3'} = 6.2 \text{ s}^{-1}$ and $k_{3' \rightarrow 4'} = 1.1 \text{ s}^{-1}$ (at +100 mV, $n = 2251$), which are close to the values obtained by fitting 1D histograms to exponential distributions ($k_{2 \rightarrow 3} = 6.9 \text{ s}^{-1}$ and $k_{3 \rightarrow 4} = 1.0 \text{ s}^{-1}$ at +100 mV, $n = 160$ respectively, Supplementary Fig. 8a, b). Next, the ability of various kinetic models describing the time-dependent evolution of the populations obtained by k -means clustering was evaluated based on the Akaike information criterion (AIC)⁴⁵ and the Bayesian information criterion (BIC)⁴⁶ (Supplementary Fig. 10). These are estimators that allow model selection based on information theory. Remarkably, the best-fitting model was the one that imposed an equilibrium between the unfolded state and both populations A and B (Fig. 4b, c), which suggests that both populations A and B are kinetic traps and that successful folding only occurs from the unfolded state. While we do not know the exact structural nature of these intermediates, the folding funnel hypothesis predicts that along the folding pathway the protein would acquire more native-like structure as the free energy decreases⁴⁷. We therefore take the recovery of mechanical stability as a proxy of nativeness. Similarly, folding upon emergence from the ribosome exit tunnel during protein synthesis may lead to folded but non-native structures⁵. A similar result was reported for the α -lytic protease, which in the absence of its pro-region folds to a native-like structure with its three disulfides in place that lacks the activity and stability of the native state⁴⁸.

To seek support for this interpretation, we performed the same experiments on a Trx mutant of lower thermodynamic stability. We assume that the native state is the deepest well in the free energy landscape. We also assume that intermediates share structural similarities with the native state. A final assumption is that transition states between one state and another share structural features of them both. Based on these assumptions, it follows that destabilization of the native state should cause slower folding kinetics in the presence of on-pathway intermediates^{49,50}. Conversely, destabilization would result in faster folding kinetics if the observed intermediates are off-pathway. We removed the stabilizing mutations A22P-I23V-P68A from Trx V5, which increase the unfolding temperature by 10 °C³⁰ to give Trx V2, and then formed a transmembrane protein rotaxane that was subjected to cycles of unfolding, translocation and refolding. This less stable mutant behaved similarly to Trx V5 but co-translocational unfolding of the native state was faster ($k_{2 \rightarrow 3} = 11.6 \pm 0.4 \text{ s}^{-1}$ and $k_{3 \rightarrow 4} = 7.5 \pm 0.3 \text{ s}^{-1}$ at +100 mV, $n = 172$), as expected²². Refolding, in agreement with the kinetic trap interpretation, was also faster. After 1000 s, Trx V2 had recovered native state stability in ~83% of the cycles examined. The refolding rate was $1.8 \times 10^{-3} \text{ s}^{-1}$, whereas that of Trx V5 was $1.4 \times 10^{-4} \text{ s}^{-1}$. The kinetic model that best fitted the data for Trx V2 again requires that the observed intermediate is off-pathway (Fig. 4d–f and Supplementary Figs. 11, 12). Therefore, in the nanopore system explored here, upon vectorial emergence from the pore, Trx rapidly collapses to non-native state(s) that shows moderate stability and is off-pathway. Due to the presence of both

levels 2' and 3', these states could have similar folds to the native state, nevertheless they must return to the unfolded state to be converted to the native state.

Conclusion

We have presented a single-molecule approach that, through the formation of a protein rotaxane in a transmembrane pore, allows the evaluation of the folded state of a protein at different refolding times after membrane translocation. The approach has millisecond temporal resolution and allows for characterization of the pathway to the refolded state with more than one parameter, which provides a better resolved folding landscape. Another important facet of our study, by comparison with recent single-molecule methods that examine the folding landscape of proteins^{51,52}, is that the protein emerges vectorially from the pore as it does from a ribosome or translocon although in our case the rate may be 2–3 orders of magnitude faster. The approach requires repeated oligonucleotide-driven translocation and retro-translocation of the protein, which cannot be achieved by using a motor protein^{24,25}.

We find that the folding mechanism of Trx following N terminus-first membrane translocation differs from that observed in bulk solution upon the removal of a denaturant such as urea or guanidinium chloride. In the latter, folding proceeds through a short-lived molten globule intermediate, if the protein has been in the unfolded state for only a few seconds⁴⁰. If the protein is unfolded for hundreds of seconds, allowing proline isomerization to occur, folding is far slower and multiple intermediates are observed, all of them on pathways to the native state⁴¹. By contrast, in our nanopore approach, the protein is in the unfolded state only during the period of retro-translocation (tens of milliseconds), which implies that proline isomerization does not occur and hence is unlikely involved in the subsequent refolding process. We therefore would expect the rapid folding behavior: if the protein does not fold vectorially during retrotranslocation, once it completes translocation it should rapidly equilibrate in the unfolded state⁵³ and the rate of folding should approximate that of the fast folding pathway observed in bulk solution. Given that we observed slow folding, the folding mechanism must be different from that in bulk solution, because the refolding rates are more than an order of magnitude slower. We show that this is a consequence of detours to off-pathway intermediates. We therefore suggest that the folding landscape we are observing is caused by the vectorial folding of the polypeptide as it emerges from the pore.

Our results are in line with analyses of T4 lysozyme⁴, HemK⁵, and flavodoxin⁵⁴ which were artificially stalled as they emerged from the ribosome exit tunnel. For these particular cases at least, folding while tethered to the ribosome from the c-terminus is slower than in bulk solution. This suggests that the vectorial nature of folding inherent to protein synthesis and membrane translocation may increase the probability of folding intermediates including off-pathway ones. The generality of this phenomena and how chaperones modulate vectorial folding remain interesting yet unaddressed questions.

Finally, the methodology presented here can find interesting uses not only to study membrane associated processes at the single-molecule level but also in biotechnology applications, providing a novel way to prepare arrays of nanopore sensors.

Methods

Synthesis of maleimide-CBT crosslinker. 17.5 mg of 6-amino-2-cyanobenzothiazole (CBT, Sigma) (0.1 mmol) and 10.7 mg of *p*-maleimidophenyl isocyanate (Sigma) (0.05 mmol) were dissolved in 1 mL of anhydrous dimethyl sulfoxide and stirred for 30 min at room temperature. The product was purified by preparative high-performance liquid chromatography (HPLC) on an Agilent 1260 Infinity system equipped with a Supelcosil PLC-18 12 μm , 250 \times 21.2 mm column. The mobile phase consisted of 0.1% formic acid in H₂O (buffer A), and 0.1% formic

acid in acetonitrile (buffer B). A gradient from 5 to 95% of buffer B in buffer A was run over 30 min at a flow rate of 5 mL min⁻¹. The separation was monitored at 335 nm. The identity of the product was confirmed by mass spectrometry (MS) and nuclear magnetic resonance (NMR) (Supplementary Fig. 1a, b).

Oligonucleotide modifications. 1.0 mg of an oligonucleotide (Biomers) composed of 30 cytosines with a 5'-thiol (hexamethylene linker) and a 3'-biotin (triethylene glycol spacer) was dissolved in 1 mL of 10 mM Tris-HCl, 1 mM EDTA, 5 mM TCEP, pH 7.5. After heating at 65 °C for 1.5 h, TCEP was removed with an Amicon Ultra column (0.5 mL, 3 kDa). The reduced oligonucleotide was then treated with the maleimide-CBT crosslinker (15 equivalents dissolved in a minimal volume of dimethyl sulfoxide) at room temperature for 2 h. The modified oligonucleotide (5'-CBT-oligo(dC)₃₀-biotin-3') was purified by ion-exchange chromatography on a monoQ FF Sepharose column (GE Healthcare) with an AKTApurifier FPLC (GE Healthcare) eluted with 0–1 M KCl in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The identity of the product was confirmed by MS (Supplementary Fig. 1c–f).

5'-S-thiopyridyl-oligo(dA)₃₀-biotin-3' was synthesized by incubating 1.0 mg of a reduced oligonucleotide composed of 30 adenines with a 5'-thiol (hexamethylene linker) and a 3'-biotin (triethylene glycol spacer) with 20 mM 2,2'-dithiodipyridine (Sigma) in acetonitrile for 2 h at room temperature (Supplementary Fig. 3b). The modified oligonucleotide was purified by ion-exchange chromatography on the monoQ FF Sepharose column.

Thioredoxin purification³⁰. The thioredoxin (Trx) mutants S1C-V5-C109 (S1C-A22P-I23V-C32S-C35S-P68A-C109) and S1C-V2-C109 (S1C-C32S-C35S-C109) were each cloned into the pET 30a+ plasmid (TopGene). *E. coli* BL21(DE3) cells (Novagen) were transformed with this plasmid and grown on LB medium supplemented with kanamycin (10 mg L⁻¹) at 37 °C with shaking. During the exponential growth phase (OD_{600nm} = 0.6), protein production was induced with isopropyl-β-D-1-thiogalactopyranoside (0.4 mM, final concentration). After overnight growth, the cells were collected by centrifugation, suspended in TE buffer (30 mM Tris-HCl, 1 mM EDTA, pH 8.3) and lysed by sonication. Cell debris was removed by centrifugation, and the supernatant containing the protein was mixed with one quarter of its volume of 10% streptomycin sulfate (w/v) in TE buffer added drop by drop at 4 °C with stirring. After 2 h, the solution was centrifuged at 20,000 × g for 30 min, and the supernatant was loaded onto a Superdex 75 10/300 column (GE Healthcare). The peak corresponding to Trx was almost pure, but was further purified by ion-exchange chromatography on a monoQ FF column eluted with 0–1 M KCl in TE buffer. The purity of the protein was estimated with SDS-PAGE in 18% Criterion TGX Precast Midi Protein Gels (Bio-Rad) stained with InstantBlue Protein Stain (Expedeon). Protein concentration was estimated spectrophotometrically from the absorbance at 280 nm by using a molar extinction coefficient of 14,100 M⁻¹cm⁻¹³⁰.

Protein-oligonucleotide conjugates. The N-terminal cysteines of Trx as purified are not reactive because they are modified with aldehydes such as pyruvate³³. The cysteines were deprotected by incubation with methoxyamine (0.4 M methoxyamine, 100 mM sodium phosphate, 150 mM NaCl, 5 mM TCEP, pH 7.0) at room temperature overnight. The protein was recovered with buffer exchange on a PD MiniTrap G-25 column (GE Healthcare). To modify the N-terminal cysteine with 5'-CBT-oligo(dC)₃₀-biotin-3', protein in 10-fold excess was mixed and incubated with 0.1–0.2 mg of the oligonucleotide in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) at room temperature overnight. The 3'-biotin-oligo(dC)₃₀-S1C-V5-C109 conjugate was purified by ion-exchange chromatography on a MonoQ FF Sepharose column eluted with 0–1 M KCl in the same buffer. The eluted conjugate was immediately mixed with 5'-S-thiopyridyl-oligo(dA)₃₀-biotin-3' and incubated for 16 h at room temperature²². The product (3'-biotin-oligo(dC)₃₀-S1C-V5-C109-oligo(dA)₃₀-biotin-3') was purified by ion-exchange chromatography, and the mass was verified with SDS-PAGE and native MS (Supplementary Fig. 3).

Purification and formation of heptameric α-hemolysin pores. α-Hemolysin (αHL) monomers were synthesized by in vitro transcription/translation and oligomerized into heptameric pores on rabbit red blood cell membranes²⁷. [³⁵S] Methionine radioactivity was used to identify the oligomeric band upon SDS-PAGE. The heptameric pores were eluted from the gel in a concentration range of 0.1–1 ng μL⁻¹, ready to use in single-molecule experiments.

Streptavidin variants expression and purification. Monovalent streptavidin (mSA) is a tetramer (SAe1D3) containing one streptavidin (SA) subunit with wild-type biotin binding affinity with a C-terminal Glu₆ tag (SAe) and three “dead” subunits (D) with negligible biotin binding affinity^{29,55}. Subunits were expressed as inclusion bodies in *E. coli*, solubilized in guanidinium hydrochloride, refolded by dilution, and the desired heterotetramer was purified by ion-exchange chromatography as described⁵⁵. mSA assembly and composition were validated by mobility on SDS-PAGE with Coomassie Blue (Bio-Rad) staining for samples with or without boiling⁵⁵. Monovalent traptavidin (mTA) is a tetramer (Tre1D3) containing one traptavidin (TA) subunit with a C-terminal Glu₆ tag (Tre) and three “dead” subunits (D)^{37,56}. mTA was purified and validated by using the same

procedure as for mSA. Tetravalent SA had the subunit composition SAe4. Tetravalent TA had the subunit composition Tre4. Tetravalent forms were expressed and refolded as above, with purification by iminobiotin-Sepharose as described⁵⁵.

Single-molecule measurements²⁷. A bilayer of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids) was made by using the Müller-Montal method on a 100 μm-diameter aperture made in a Teflon film (25 μm thick, Goodfellow) that separated two compartments (*cis* and *trans*). Each compartment was filled with electrolyte solution (10 mM HEPES, 2 M KCl, pH 7.2) and connected to the headstage of an Axopatch 200B amplifier (Molecular Devices) with Ag/AgCl electrodes. A lid was used to cover the compartments to avoid evaporation, and the temperature was controlled with a circulating bath to 23 °C. The signal from the amplifier was stored on a computer by using a Digidata 1440 A digitizer (Molecular Devices). Following membrane formation, heptameric αHL pores (0.2 μL, see above) were added to the *cis* compartment under an applied potential of +100 mV. After a single insertion event (identified by a step increase in the conductance of 1 nS) the *cis* compartment was perfused with fresh buffer by using a push-pull syringe driver PHD 2000 Syringe Pump (Harvard Apparatus) to prevent further insertions. Data were low-pass filtered at 5 kHz, which corresponds to a temporal resolution of ~0.2 ms, and sampled at 20 kHz. Voltage protocols in Clampex (Molecular Devices) were used to automatically change the voltage at defined times. The negative potential at level 7' was maintained for 0.1 s, 0.5 s, 1 s, 1.5 s, 10 s, 100 s, or 1000 s, and therefore the protein was allowed to refold for 0.09 s, 0.49 s, 0.99 s, 1.49 s, 10 s, 100 s, or 1000 s, respectively (the average duration of the retro-translocation, level 6', was deducted to calculate the refolding times at short timescales).

Statistics and reproducibility. The state of individual molecules was probed hundreds of times under each condition. These data were collected on at least 3 independent experiments for each condition. Raw data files were first analyzed with pClamp (Molecular Devices) to obtain the current amplitude and dwell time in each level. The event histograms of the natural logarithm of the dwell times were fit to an exponential distribution by Igor Pro (Wavemetrics):

$$f(x) = Ake^{(x-ke^x)}$$

with: *A*, amplitude; *k*, rate constant; *x*, the natural logarithm of the dwell time. The rate constants in the main text are the best-fit values ± 1σ confidence intervals.

The 95% confidence interval in the estimation of the refolded fraction, *p*, (Fig. 3c) was estimated by using:

$$p \pm 1.96 \sqrt{\frac{p(1-p)}{n}}$$

with: *n*, the number of observations.

k-means clustering analysis was carried out by Igor Pro considering only the cases where both levels 2' and 3' were detected. In *k*-means clustering, the centroids define the center of each population (Supplementary Figs. 9, 11). Each data point is assigned to the population with the closest centroid. For Trx V5, a ~5% overestimation of the unfolded population and population A was assumed, due to an overlap with the population B. Therefore, the estimation of each population should be taken as a rough estimate. The different kinetic models used to explain the time-course evolution of refolding were evaluated with the application SimBiology from Matlab (MathWorks), which provided the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Supplementary Figs. 10, 12).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Relevant data and/or materials are available upon reasonable request from D.R.-L. and/or H.B.

Received: 1 October 2019; Accepted: 20 February 2020;

Published online: 03 April 2020

References

1. Anfinsen, C. B., Haber, E., Sela, M. & White, F. H. Jr The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl Acad. Sci. USA* **47**, 1309–1314 (1961).
2. Lipman, E. A., Schuler, B., Bakajin, O. & Eaton, W. A. Single-molecule measurement of protein folding kinetics. *Science* **301**, 1233–1235 (2003).
3. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).

4. Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I. Jr & Bustamante, C. The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727 (2011).
5. Holtkamp, W. et al. Cotranslational protein folding on the ribosome monitored in real time. *Science* **350**, 1104–1107 (2015).
6. Kim, S. J. et al. Protein folding. Translational tuning optimizes nascent protein folding in cells. *Science* **348**, 444–448 (2015).
7. Kowarik, M., Kung, S., Martoglio, B. & Helenius, A. Protein folding during cotranslational translocation in the endoplasmic reticulum. *Mol. Cell* **10**, 769–778 (2002).
8. Nilsson, O. B. et al. Cotranslational protein folding inside the ribosome exit tunnel. *Cell Rep.* **12**, 1533–1540 (2015).
9. Ugrinov, K. G. & Clark, P. L. Cotranslational folding increases GFP folding yield. *Biophys. J.* **98**, 1312–1320 (2010).
10. Fedyukina, D. V. & Cavagnero, S. Protein folding at the exit tunnel. *Annu. Rev. Biophys.* **40**, 337–359 (2011).
11. Okamoto, K. et al. The protein import motor of mitochondria: a targeted molecular ratchet driving unfolding and translocation. *EMBO J.* **21**, 3659–3671 (2002).
12. Kadokura, H. & Beckwith, J. Detecting folding intermediates of a protein as it passes through the bacterial translocation channel. *Cell* **138**, 1164–1173 (2009).
13. Tsirogotaki, A., De Geyter, J., Sostaric, N., Economou, A. & Karamanou, S. Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15**, 21–36 (2017).
14. Costa, T. R. et al. Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat. Rev. Microbiol.* **13**, 343–359 (2015).
15. Junker, M., Besing, R. N. & Clark, P. L. Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Mol. Microbiol.* **71**, 1323–1332 (2009).
16. Li, L. et al. Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature* **531**, 395–399 (2016).
17. Matouschek, A., Pfanner, N. & Voos, W. Protein unfolding by mitochondria. The Hsp70 import motor. *EMBO Rep.* **1**, 404–410 (2000).
18. Wilson, R. et al. The translocation, folding, assembly and redox-dependent degradation of secretory and membrane proteins in semi-permeabilized mammalian cells. *Biochem. J.* **307**(Pt 3), 679–687 (1995).
19. Geiger, R., Gautschi, M., Thor, F., Hayer, A. & Helenius, A. Folding, quality control, and secretion of pancreatic ribonuclease in live cells. *J. Biol. Chem.* **286**, 5813–5822 (2011).
20. Bonardi, F. et al. Probing the SecYEG translocation pore size with preproteins conjugated with sizable rigid spherical molecules. *Proc. Natl Acad. Sci. USA* **108**, 7775–7780 (2011).
21. Ha, T. Single-molecule methods leap ahead. *Nat. Methods* **11**, 1015–1018 (2014).
22. Rodriguez-Larrea, D. & Bayley, H. Multistep protein unfolding during nanopore translocation. *Nat. Nanotechnol.* **8**, 288–295 (2013).
23. Rodriguez-Larrea, D. & Bayley, H. Protein co-translocational unfolding depends on the direction of pulling. *Nat. Commun.* **5**, 4841 (2014).
24. Nivala, J., Mulrone, L., Li, G., Schreiber, J. & Akeson, M. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS nano* **8**, 12365–12375 (2014).
25. Nivala, J., Marks, D. B. & Akeson, M. Unfoldase-mediated protein translocation through an alpha-hemolysin nanopore. *Nat. Biotechnol.* **31**, 247–250 (2013).
26. Song, L. et al. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859–1866 (1996).
27. Maglia, G., Heron, A. J., Stoddart, D., Japrun, D. & Bayley, H. Analysis of single nucleic acid molecules with protein nanopores. *Methods Enzymol.* **475**, 591–623 (2010).
28. Sanchez-Quesada, J., Saghatelian, A., Cheley, S., Bayley, H. & Ghadiri, M. R. Single DNA rotaxanes of a transmembrane pore protein. *Angew. Chem. Int. Ed. Engl.* **43**, 3063–3067 (2004).
29. Howarth, M. et al. A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods* **3**, 267–273 (2006).
30. Rodriguez-Larrea, D. et al. Role of conservative mutations in protein multi-property adaptation. *Biochem. J.* **429**, 243–249 (2010).
31. Ren, H. et al. A biocompatible condensation reaction for the labeling of terminal cysteine residues on proteins. *Angew. Chem. Int. Ed. Engl.* **48**, 9658–9662 (2009).
32. Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **13**, 697–705 (2017).
33. Gentle, I. E., De Souza, D. P. & Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug. Chem.* **15**, 658–663 (2004).
34. Iyer, M., Norton, J. C. & Corey, D. R. Accelerated hybridization of oligonucleotides to duplex DNA. *J. Biol. Chem.* **270**, 14712–14717 (1995).
35. Stoddart, D., Heron, A. J., Mikhailova, E., Maglia, G. & Bayley, H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc. Natl Acad. Sci. USA* **106**, 7702–7707 (2009).
36. Purnell, R. F., Mehta, K. K. & Schmidt, J. J. Nucleotide identification and orientation discrimination of DNA homopolymers immobilized in a protein nanopore. *Nano Lett.* **8**, 3029–3034 (2008).
37. Chivers, C. E. et al. A streptavidin variant with slower biotin dissociation and increased mechanostability. *Nat. Methods* **7**, 391–393 (2010).
38. Kiefhaber, T. Kinetic traps in lysozyme folding. *Proc. Natl Acad. Sci. USA* **92**, 9029–9033 (1995).
39. Kelley, R. F., Wilson, J., Bryant, C. & Stellwagen, E. Effects of guanidine hydrochloride on the refolding kinetics of denatured thioredoxin. *Biochemistry* **25**, 728–732 (1986).
40. Georgescu, R. E., Li, J. H., Goldberg, M. E., Tasayco, M. L. & Chaffotte, A. F. Proline isomerization-independent accumulation of an early intermediate and heterogeneity of the folding pathways of a mixed alpha/beta protein, *Escherichia coli* thioredoxin. *Biochemistry* **37**, 10286–10297 (1998).
41. Roderer, D. J., Scharer, M. A., Rubini, M. & Glockshuber, R. Acceleration of protein folding by four orders of magnitude through a single amino acid substitution. *Sci. Rep.* **5**, 11840 (2015).
42. Reimer, U. et al. Side-chain effects on peptidyl-prolyl cis/trans isomerisation. *J. Mol. Biol.* **279**, 449–460 (1998).
43. Berkovich, R. et al. Rate limit of protein elastic response is tether dependent. *Proc. Natl Acad. Sci. USA* **109**, 14416–14421 (2012).
44. Lloyd, S. P. Least Squares Quantization in PCM. In *IEEE Transactions on Information Theory* **IT-28**, 9 (1982).
45. Akaike, H. A New Look at the Statistical Model Identification. In *IEEE Transactions on Automatic Control* **AC-19**, 8 (1974).
46. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **6**, 4 (1978).
47. Leopold, P. E., Montal, M. & Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl Acad. Sci. USA* **89**, 8721–8725 (1992).
48. Baker, D., Sohl, J. L. & Agard, D. A. A protein-folding reaction under kinetic control. *Nature* **356**, 263–265 (1992).
49. Baldwin, R. L. On-pathway versus off-pathway folding intermediates. *Fold. Des.* **1**, R1–R8 (1996).
50. Matouschek, A., Kellis, J. T. Jr, Serrano, L., Bycroft, M. & Fersht, A. R. Transient folding intermediates characterized by protein engineering. *Nature* **346**, 440–445 (1990).
51. Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C. & Rief, M. The complex folding network of single calmodulin molecules. *Science* **334**, 512–516 (2011).
52. Pirchi, M. et al. Single-molecule fluorescence spectroscopy maps the folding landscape of a large protein. *Nat. Commun.* **2**, 493 (2011).
53. Zwanzig, R. Two-state models of protein folding kinetics. *Proc. Natl Acad. Sci. USA* **94**, 148–150 (1997).
54. Houwman, J. A., Andre, E., Westphal, A. H., van Berkel, W. J. & van Mierlo, C. P. The ribosome restrains molten globule formation in stalled nascent flavodoxin. *J. Biol. Chem.* **291**, 25911–25920 (2016).
55. Fairhead, M. et al. Plug-and-play pairing via defined divalent streptavidins. *J. Mol. Biol.* **426**, 199–214 (2014).
56. Chivers, C. E. et al. How the biotin-streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. *Biochem. J.* **435**, 55–63 (2011).

Acknowledgements

The authors thank Ellina Mikhailova for α -hemolysin prepared by in vitro transcription and translation. We thank Dr. Jonathan Hopper and Professor Carol Robinson for the MS of the construct oligo(dC)₃₀-V5-oligo(dA)₃₀. H.B. was funded by the NIH and Oxford Nanopore Technologies and M.H. by the Biotechnology and Biological Sciences Research Council (BBSRC, grant BB/M02122X/1). J.F. was supported by the China Scholarship Council and G.V. by a Medical Research Council studentship and Merton College Oxford. D.R.-L. is a recipient of a Ramón y Cajal Fellowship (RYC-2013-12799). D.R.-L. was funded by MINECO grants BIO2017-88946-R and BFU2016-81754-ERC (FEDER funds). This work was supported in part by the Fundación Biofísica Bizkaia and the Basque Excellence Research Centre (BERC) program of the Basque Government.

Author contributions

J.F., D.R.-L., and H.B. planned the research. J.F., P.M.-B., and D.R.-L. performed nanopore experiments and data analysis. J.F. and M.J.B. synthesized the CBT-maleimide linker. G.V. and M.H. purified, validated, and provided the mSA and mTA. J.F., D.R.-L., and H.B. wrote the paper. All authors reviewed the paper.

Competing interests

H.B. declares no competing non-financial interests but the following competing financial interest: H.B. is the Founder of, a consultant for and a share-holder of Oxford Nanopore Technologies, a company engaged in the development of nanopore sensing and

sequencing technologies. The remaining authors declare no financial or non-financial interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-020-0840-5>.

Correspondence and requests for materials should be addressed to H.B. or D.R.-L.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

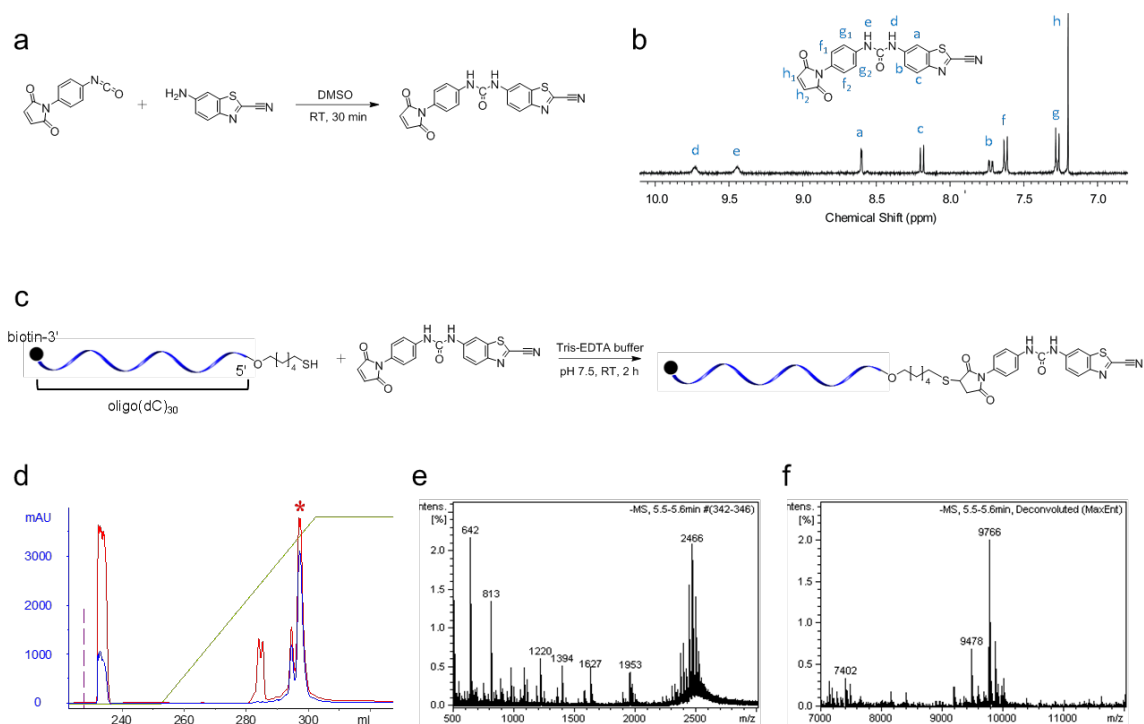


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Supplementary Figures

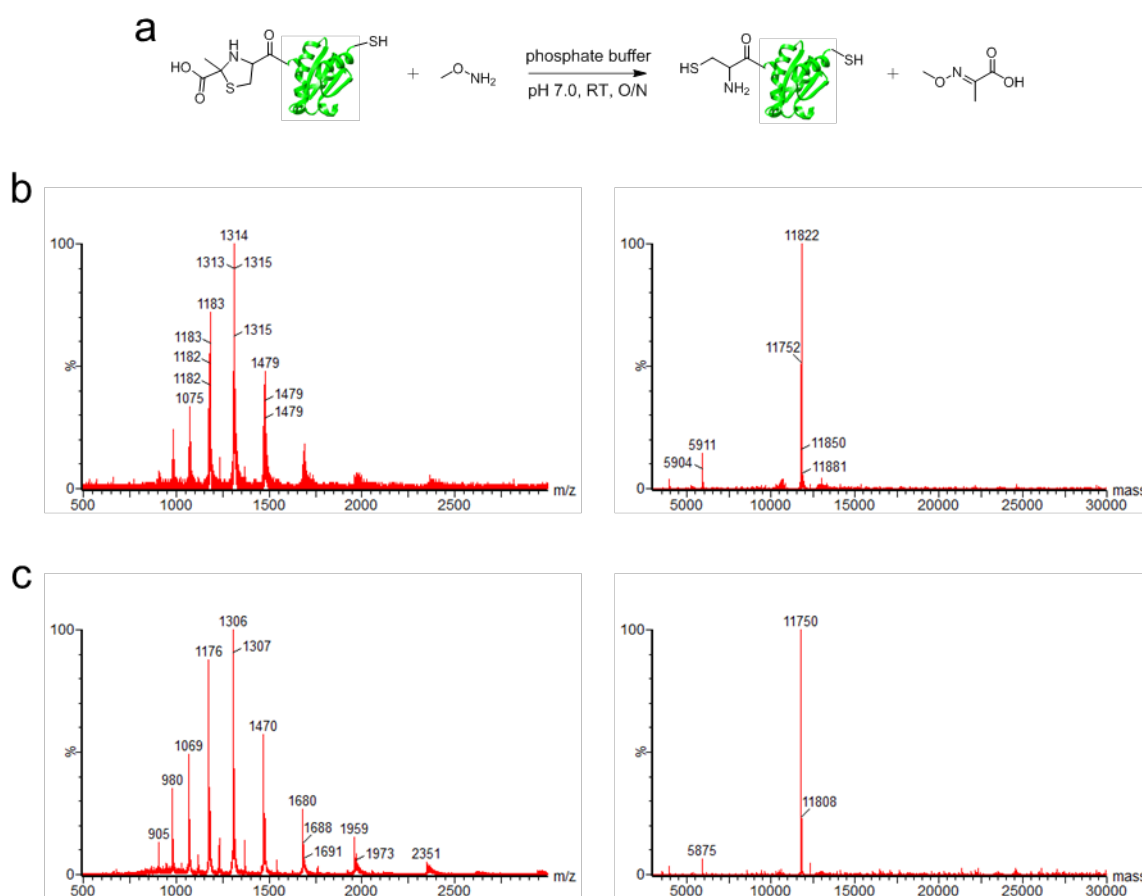
Supplementary Figure 1



Supplementary Figure 1. Synthesis of 5'-CBT-oligo(dC)₃₀-biotin-3'. a

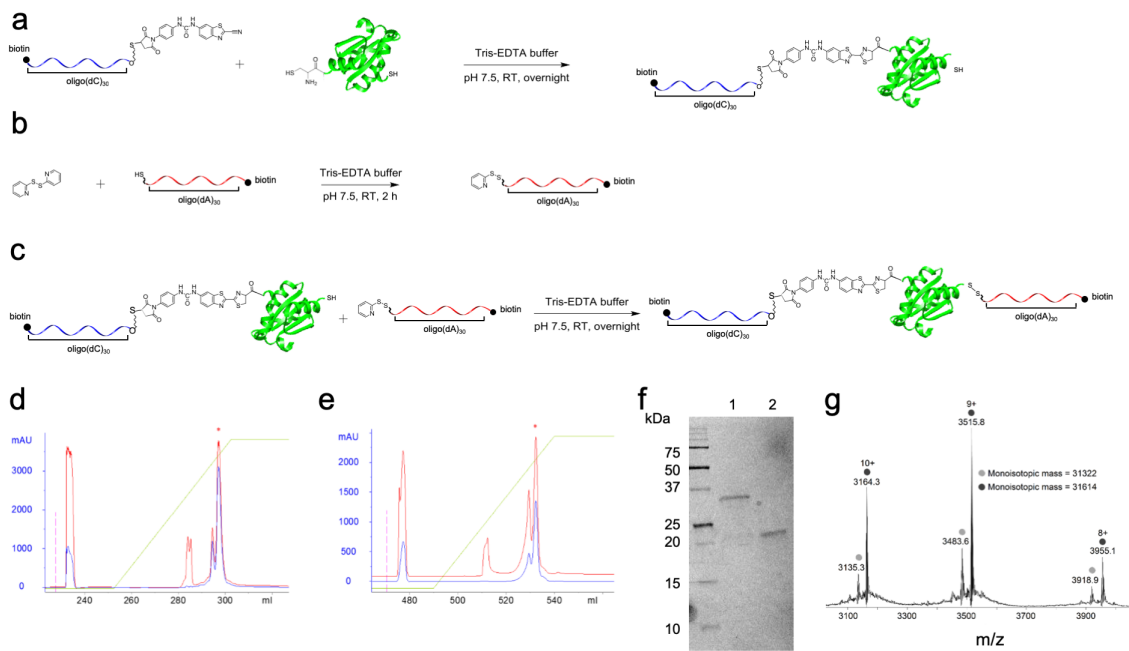
Reaction between 6-amino-2-CBT and *p*-maleimidophenyl isocyanate to form the maleimide-CBT crosslinker. **b** ¹H NMR spectrum of the product (400 MHz, DMSO-d₆): δ 7.20 (s, 2H), 7.26-7.28 (d, 2H, *J* = 8.85), 7.61-7.63 (d, 2H, *J* = 8.85), 7.72-7.74 (dd, 1H, *J* = 9.00, 1.98), 8.18-8.20 (d, 1H, *J* = 9.00), 8.60-8.61 (d, 1H, *J* = 1.98), 9.44 (s, 1H), 9.73 (s, 1H). **c** Reaction of 5'-thiol-oligo(dC)₃₀-biotin-3' with the maleimide-CBT crosslinker. **d** Ion-exchange chromatography of the modified oligonucleotide (absorbance at 280 nm, blue; 254 nm, red). **e** ESI-MS of the major peak (*) from 'd'. **f** Deconvoluted ESI-MS: calculated for 5'-CBT-oligo(dC)₃₀-biotin-3' = 9,769; observed = 9,766.

Supplementary Figure 2



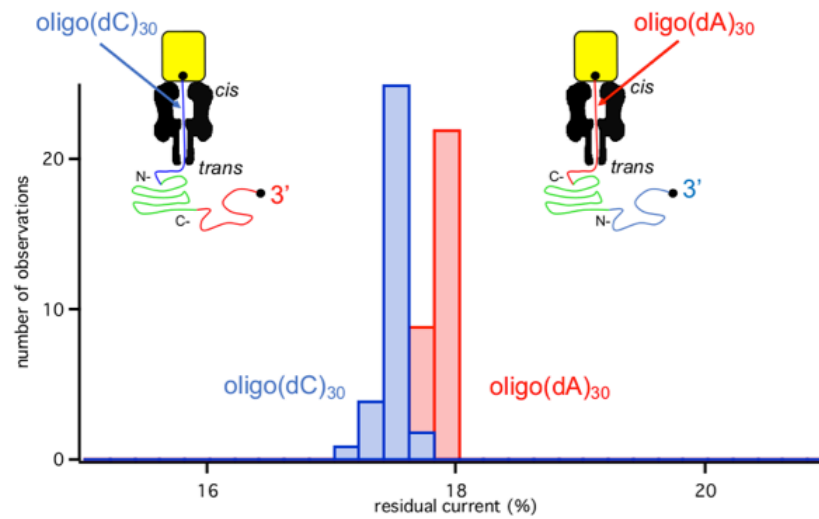
Supplementary Figure 2. Removal of pyruvate from the N-terminal cysteine of Trx S1C-V5-C109. **a** Pyruvate was removed from the N terminus of Trx by incubation with methoxyamine. **b** ESI-MS, and the deconvoluted spectrum, of Trx as purified from *E. coli*. Calculated $[M + H]^+ = 11,749$. The observed mass of 11,822 corresponds to the pyruvate adduct (+73 Da). **c** ESI-MS, and the deconvoluted spectrum, after the incubation with methoxyamine. Observed = 11,750.

Supplementary Figure 3



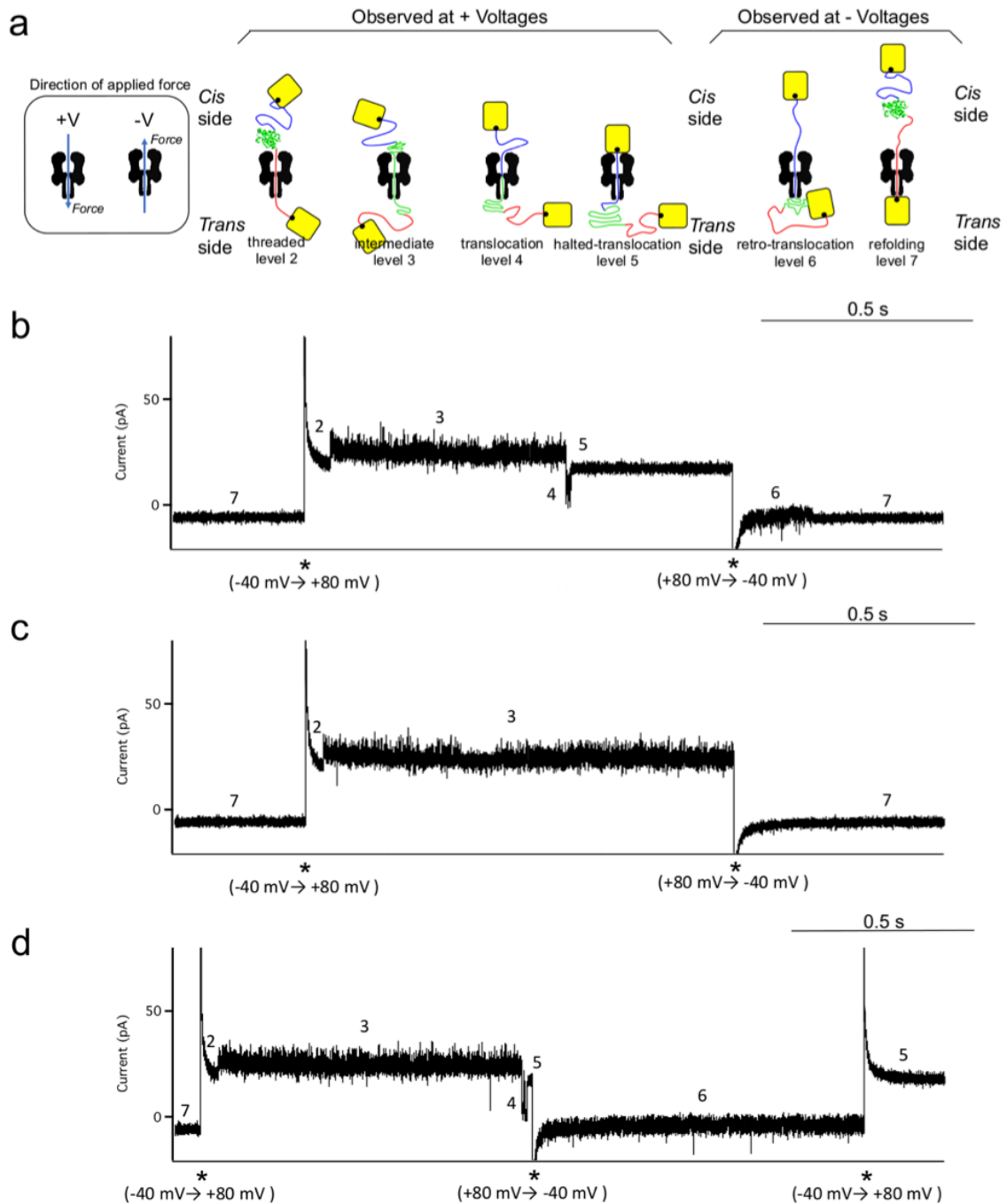
Supplementary Figure 3. Synthesis of 3'-biotin-oligo(dC)₃₀-S1C-V5-C109-oligo(dA)₃₀-biotin-3'. **a** The 30-mer cytosine oligonucleotide (5'-CBT-oligo(dC)₃₀-biotin-3') was coupled to the N-terminal cysteine of Trx S1C-V5-C109. **b** A 30-mer adenine oligonucleotide (5'-thiol-oligo(dA)₃₀-biotin-3') was activated with 2,2'-dipyridyl disulphide. **c** The 3'-biotin-oligo(dC)₃₀-S1C-V5-C109 conjugate produced in 'a' was reacted with the activated oligonucleotide from 'b' to form the 3'-biotin-oligo(dC)₃₀-S1C-V5-C109-oligo(dA)₃₀-biotin-3' conjugate. **d** Ion-exchange chromatography of the reaction products shown in 'a'. The asterisk marks the elution of the 3'-biotin-oligo(dC)₃₀-S1C-V5-C109 conjugate (absorbance at 280 nm, blue; 254 nm, red). **e** Ion-exchange chromatography of the reaction products shown in 'c'. The asterisk marks the elution of the 3'-biotin-oligo(dC)₃₀-S1C-V5-C109-oligo(dA)₃₀-biotin-3' conjugate. **f** SDS-PAGE of the purified products. Lane 1: the 3'-biotin-oligo(dC)₃₀-S1C-V5-C109-oligo(dA)₃₀-biotin-3' conjugate; lane 2: the 3'-biotin-oligo(dC)₃₀-S1C-V5-C109 conjugate. **g** Native MS of the 3'-biotin-oligo(dC)₃₀-S1C-V5-C109-oligo(dA)₃₀-biotin-3' conjugate. Calculated = 31,598; observed = 31,614. The difference observed is due to the opening of the succinimide ring by hydrolysis (+18). The minor peak of 31,322 corresponds to a conjugate missing one cytosine nucleotide.

Supplementary Figure 4



Supplementary Figure 4. The residual current in level 5 depends on the threading direction. Histogram of the residual currents obtained from oligo(dC)₃₀-V5-oligo(dA)₃₀ that threaded N terminus-first (blue bars) or C terminus-first (red bars) in the presence of equimolar amounts of mSA. After translocation of the protein through the pore, oligo(dC)₃₀ is arrested within the pore by mSA if the translocation is N terminus-first. Oligo(dA)₃₀ is arrested within the pore if the translocation is C terminus-first. The different base compositions produce different residual currents. Data were collected with two different α HL pores. The residual currents were $I_{\text{res}\%} = 17.8 \pm 0.1$ for oligo(dA)₃₀ and $I_{\text{res}\%} = 17.4 \pm 0.1$ for oligo(dC)₃₀ (mean \pm standard deviation; best-fit values to Gaussian distributions).

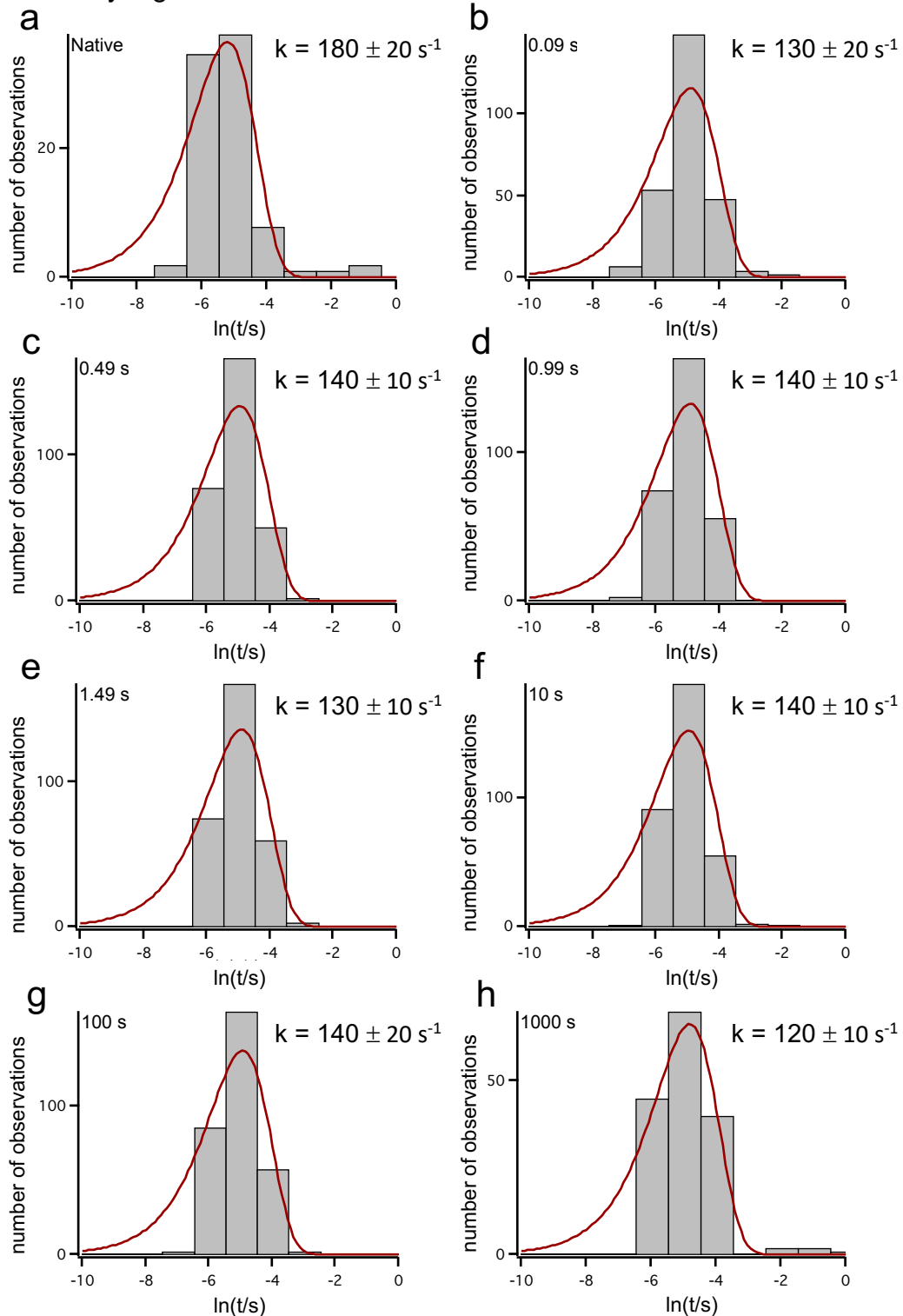
Supplementary Figure 5



Supplementary Figure 5. Ionic current signals during movement of the oligo(dC)₃₀-V5-oligo(dA)₃₀ within the α HL pore. **a** Current levels associated with different states of the rotaxane. Box: change in the direction of the force acting on an oligonucleotide arising from the reversal of voltage polarity. **b** Ionic current signal showing one cycle of unfolding-translocation-refolding. The asterisk indicates a change in the voltage polarity. Initially, under a negative voltage, the protein is located in the *cis* compartment (level 7'). A change in the voltage polarity to +80 mV pulls the protein through the pore towards the *trans* compartment. The protein unfolds (levels 2' and 3') and translocates (level 4'), and oligo(dC)₃₀-V5-oligo(dA)₃₀ is then arrested by mSA with the protein in the *trans* compartment (level 5'). A switch in the voltage polarity pulls the protein

back towards the *cis* compartment. The protein, if refolded, unfolds and retro-translocates from the *trans* to the *cis* compartment (level 6'), and during level 7', it refolds in the *cis* compartment before a new cycle is initiated. **c** If the voltage is switched from positive to negative during level 3' (in which the protein has not reached the *trans* compartment) only level 7' (but not level 6') is observed. **d** If the voltage is switched from negative to positive during level 6', level 5' follows, because the protein has not completed retro-translocation into the *cis* compartment.

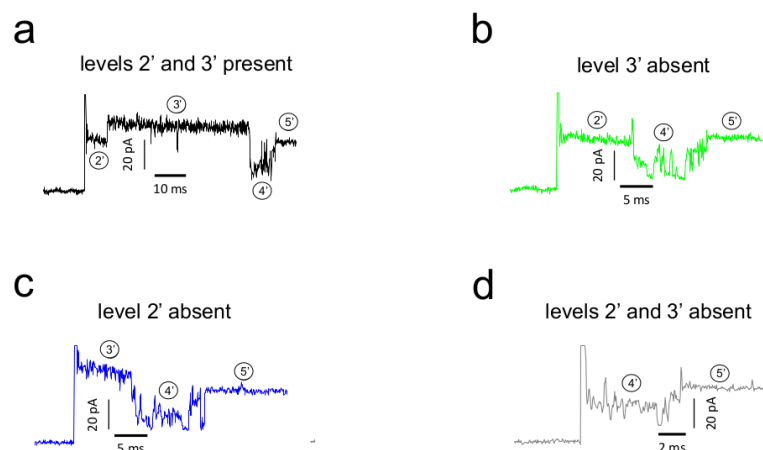
Supplementary Figure 6



Supplementary Figure 6. The translocation time of the unfolded protein (dwell time in level 4') does not depend on the refolding time (dwell time in level 7'). **a** Histogram of the dwell time in level 4' during co-translocational unfolding of native Trx V5 at +100 mV. The red line is the best fit to an exponential distribution. **b** Histogram of the dwell time in level 4' during co-translocational unfolding of refolded Trx V5 after a refolding time of 0.09 s. The red line is the best fit to an exponential distribution. **c** Same as 'b' for a refolding time of 0.49 s. **d** A refolding time of 0.99 s. **e** A refolding time of 1.49 s. **f** A

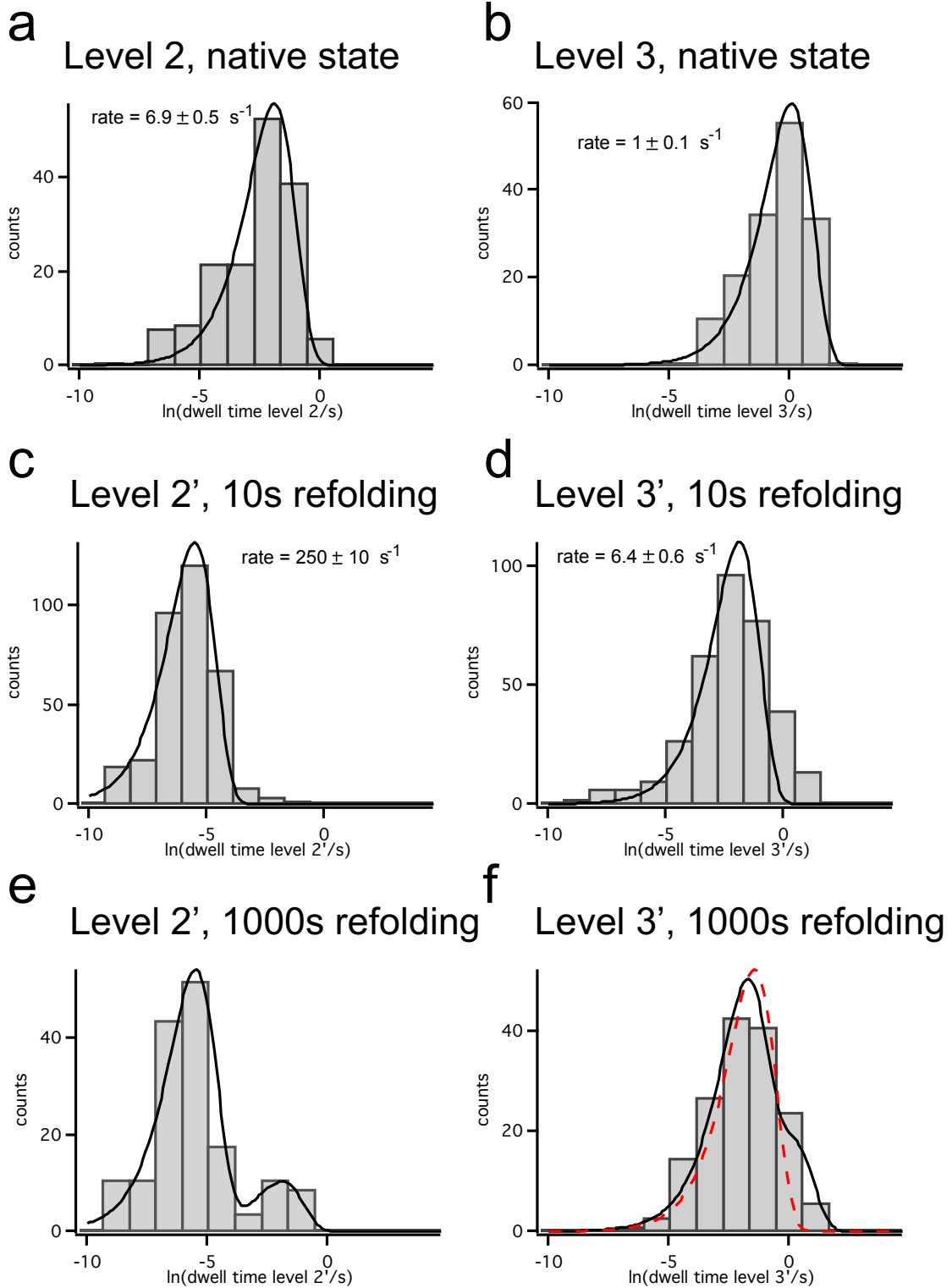
refolding time of 10 s. **g** A refolding time of 100 s. **h** A refolding time 1000 s. Rate constants ($k_{4 \rightarrow 1}$ or $k_{4' \rightarrow 5'}$) are indicated within each panel (best-fit value $\pm 1\sigma$ confidence interval). Histograms are based on data collected from at least three independent experiments.

Supplementary Figure 7



Supplementary Figure 7. Presence and absence of levels 2' and 3' upon co-translocational unfolding of the refolded state. Representative traces of ionic currents obtained from the oligo(dC)₃₀-V5-oligo(dA)₃₀/αHL rotaxane at +100 mV after 1.49 s of refolding during level 7' at -100 mV. **a** both levels 2' and 3' are present. **b** level 3' is absent, which suggests either that the N-terminal domain of the protein did not refold or that it unfolded faster than the detection limit (0.2 ms, a short spike is observed at the end of level 2', which suggests that the second explanation may apply here). **c** level 2' is not observed, which suggests either that the C-terminal domain did not refold or that it unfolded faster than the detection limit (here the capacitive transient after the voltage step increases the detection limit to ~1 ms). **d** neither level 2' nor 3' is observed, which suggests either that the protein remained unfolded or that it unfolded faster than the detection limits.

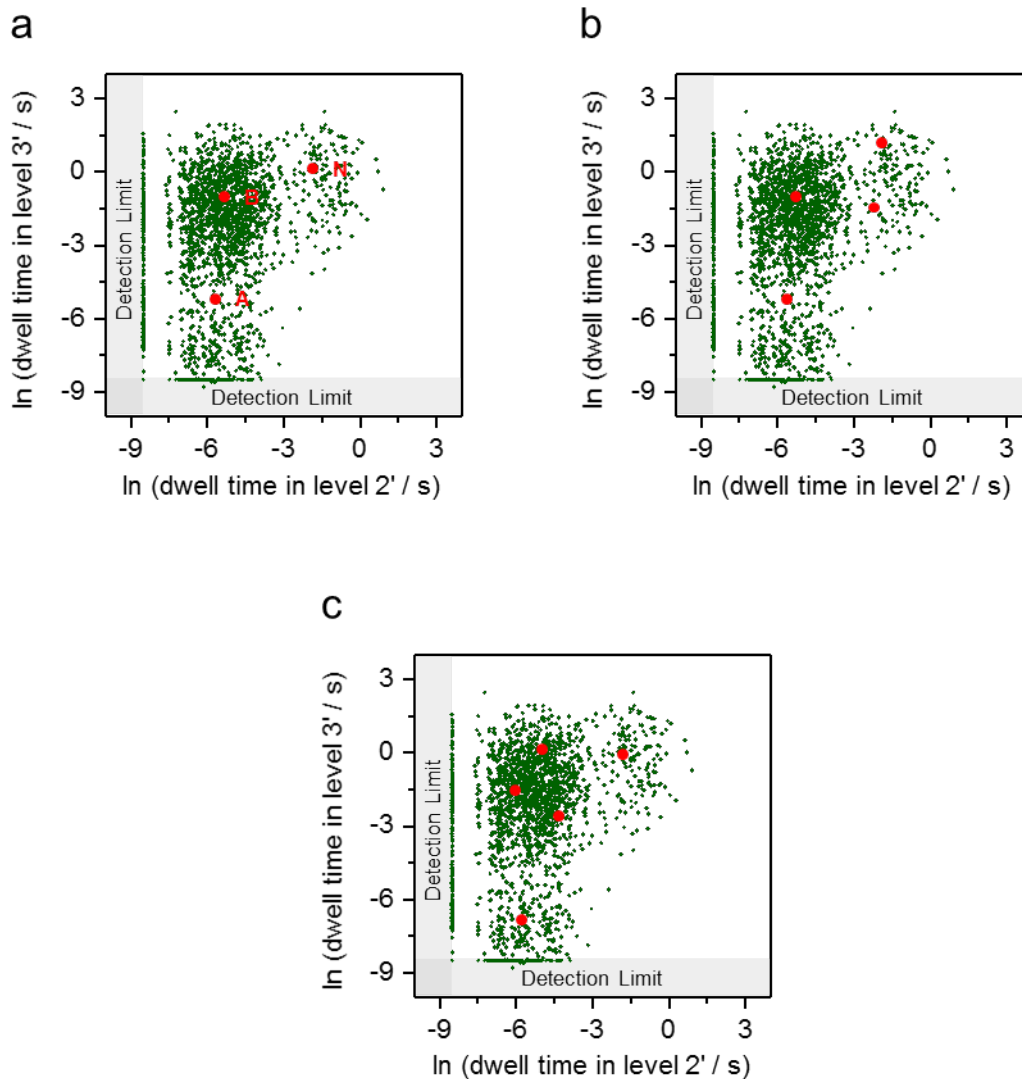
Supplementary Figure 8



Supplementary Figure 8. Dwell time distributions in levels 2' and 3'. Histograms showing the dwell time distributions in levels 2' and 3' for the oligo(dC)₃₀-V5-oligo(dA)₃₀/αHL rotaxane at +100 mV. **a** Dwell time in level 2 for the native state. The solid line is the best fit to an exponential distribution. **b** Dwell time in level 3 for the native state. **c** Dwell time in level 2' after 10 s of

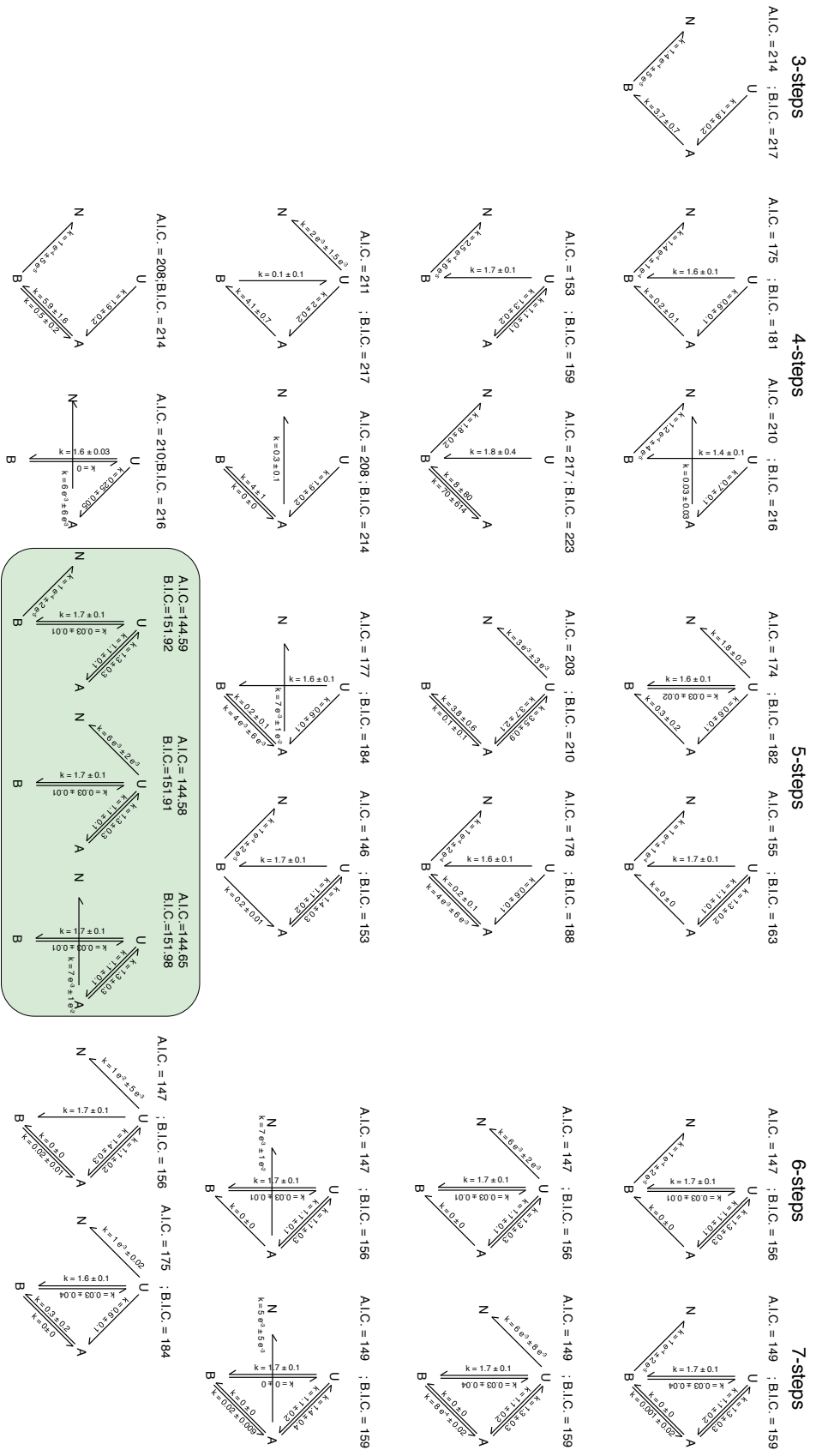
refolding during level 7'. **d** Dwell time in level 3' after 10 s of refolding during level 7'. **e** Dwell time in level 2' after 1000 s of refolding during level 7'. The solid line is the best fit to a double exponential distribution using the rate constants obtained from 'a' and 'c'. **f** Dwell time in level 3' after 1000 s of refolding during level 7'. The solid line is the best fit to a double exponential distribution using the rate constants obtained from 'b' and 'd'. The red dashed line is a fit to a single exponential distribution. The derived rate constants are indicated within each panel (best-fit value $\pm 1\sigma$ confidence interval). Histograms are based on data collected from at least three independent experiments.

Supplementary Figure 9



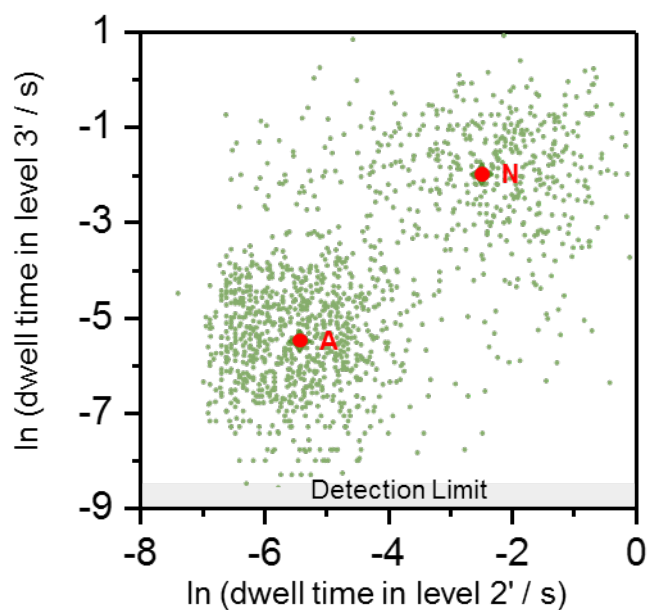
Supplementary Figure 9. Centroids obtained by *k*-means clustering of dwell times with the oligo(dC)₃₀-V5-oligo(dA)₃₀/αHL rotaxane. The natural logarithm of the dwell times in levels 2' and 3' for the various refolding times combined, together with those in levels 2 and 3 for the native protein, are used to obtain the centroids of three different populations. Each data point (green) is assigned to the population with the closest centroid (red). **a** Centroids obtained with *k*-means clustering when three clusters are considered. The lowest centroid is shifted towards the upper cluster because the upper cluster comprises more data points. **b** Considering four clusters does not improve the assignment to the lowest population. **c** Considering five clusters gives centroids that clearly identify the different populations, with the major population having three centroids. Data were obtained from three different pores for the native protein and three for the refolded protein.

Supplementary Figure 10



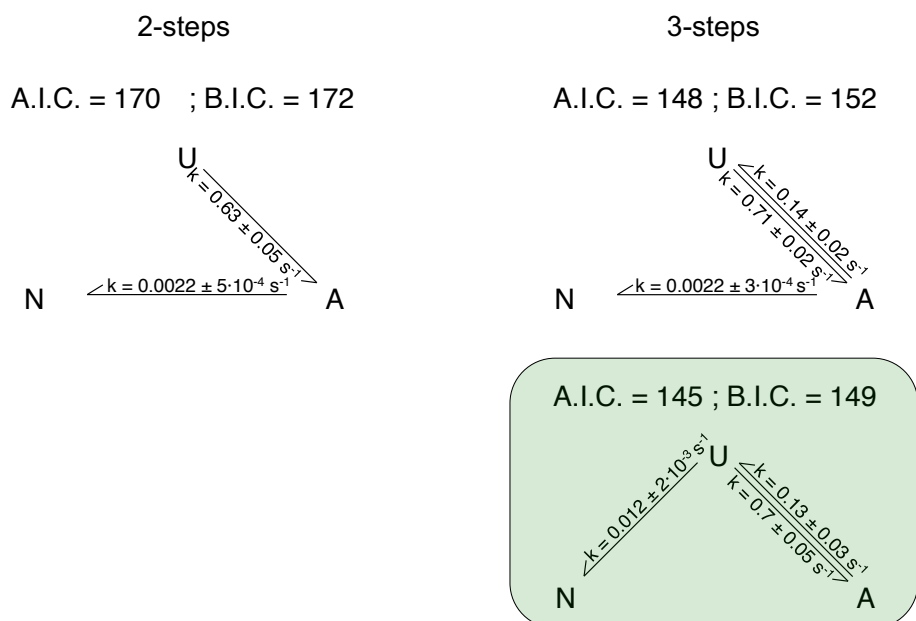
Supplementary Figure 10. The best models describing the folding kinetics of Trx V5 have five kinetic steps and two kinetic traps. The models considered assume that formation of the native state is irreversible and that, at infinite time, all molecules reach the native state (i.e. there is no irreversible formation of the populations A and B). The models displayed fit the time-dependent evolution of populations A and B, the native state and the unfolded state. For each model, the BIC and the AIC are indicated. For both criteria, the lowest value indicates the model that better fits the data with fewer parameters. Three models that consider the populations A and B to be in equilibrium with the unfolded state have the lowest BIC and AIC values (green box). The lowest one -therefore the chosen one-, the formation of the native state only occurs from the unfolded state, and both populations A and B are off-pathway.

Supplementary Figure 11



Supplementary Figure 11. Centroids obtained by *k*-means clustering on dwell times with the oligo(dC)₃₀-V2-oligo(dA)₃₀/αHL rotaxane. The natural logarithm of the dwell times in levels 2' and 3' for the various refolding times combined, together with those in level 2 and 3 for the native state, are used to obtain the centroids (red) of two different populations. Data were obtained from three different pores for the native protein and three for the refolded protein.

Supplementary Figure 12



Supplementary Figure 12. The best model describing the folding kinetics of Trx V2 has three kinetic steps and one kinetic trap. The models considered assume that formation of the native state is irreversible and that, at infinite time, all molecules reach the native state (i.e. there is no irreversible formation of the population A). The presented models were used to fit the time-dependent evolution of population A, the native state and the unfolded state. For each model, the BIC and the AIC are shown. For both criteria, the lowest value indicates the model that better fits the data with fewer parameters. The best-fit model is in the box and indicates that the native state only occurs from the unfolded state and the population A is off-pathway.