



## Structural polymorphism and substrate promiscuity of a ribosome-associated molecular chaperone

Chih-Ting Huang<sup>1</sup>, Yei-Chen Lai<sup>2</sup>, Szu-Yun Chen<sup>1</sup>, Meng-Ru Ho<sup>1</sup>, Yun-Wei Chiang<sup>2</sup>, Shang-Te  
5 Danny Hsu<sup>1,3,\*</sup>

1. Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan

2. Department of Chemistry, National Tsing Hua University, Hsichu 30013, Taiwan

3. Institute of Biochemical Sciences, National Taiwan University, Taipei 106, Taiwan

10 Correspondence to: Shang-Te Danny Hsu ([sthsu@gate.sinica.edu.tw](mailto:sthsu@gate.sinica.edu.tw))

**Abstract.** Trigger factor (TF) is a highly conserved multi-domain molecular chaperone that exerts its chaperone activity at the ribosomal tunnel exit from which newly synthesized nascent chains emerge. TF also displays promiscuous substrate binding for a large number of cytosolic proteins independent of ribosome binding. We asked how TF recognizes a variety of substrates while existing in a monomer-dimer equilibrium. Paramagnetic NMR, electron spin resonance spectroscopy and chemical  
15 crosslink show that dimeric TF displays a high degree of structural polymorphism in solution. A series of peptides has been generated to quantify their TF binding affinities in relation with their sequence compositions. The results confirmed a previous predication that TF preferentially binds to peptide fragments that are rich in aromatic and positively charged amino acids. NMR paramagnetic relaxation enhancement analysis showed that TF utilizes multiple binding sites, located in the chaperone domain and part of the prolyl trans/cis isomerisation domain, to interact with these peptides. Dimerization of TF effectively  
20 sequesters most of substrate binding sites, which are expected to become accessible upon binding to the ribosome as a monomer. As TF lacks ATPase activity, which is commonly used to trigger conformational changes within molecular chaperones in action, the ribosome-binding-associated disassembly and conformational rearrangements may be the underlying regulatory mechanism of its chaperone activity.

### 1. Introduction

25 Molecular chaperones are pivotal in facilitating protein folding and maintaining proteostasis *in vivo* (Hartl, 2016; Hartl and Hayer-Hartl, 2002). In prokaryotes, trigger factor (TF) is a highly conserved multidomain molecular chaperone, consisting of a ribosome binding domain (RBD), substrate binding domain (SBD) and a prolyl peptidyl trans-cis isomerization domain (PPI) (Hesterkamp and Bukau, 1996; Hoffmann *et al.*, 2010). TF binds to the ribosomal protein L23 through RBD in a 1:1 stoichiometry at the exit of the ribosomal tunnel where newly synthesized nascent polypeptide chains emerge during translation (Ferbitz *et al.*, 2004; Lakshmiathy *et al.*, 2007; Merz *et al.*, 2008; Rutkowska *et al.*, 2008). Unlike most molecular chaperones,  
30 which display ATPase activity to confer chaperone activities, TF does not have ATPase activity. Instead, TF forms a dragon-like cradle at the ribosomal tunnel exit to sequester emerging nascent chains and continues to hold onto its substrates after being released from the ribosome until folding is complete (Ferbitz *et al.*, 2004; Hesterkamp and Bukau, 1996). Therefore, TF is considered as a holdase to delay protein (mis)folding events. TF binds to the ribosome with a dissociation constant ( $K_d$ ) of ca. 1  $\mu\text{M}$ , and free TF self-dimerizes in solution with a comparable  $K_d$ . Given that the cellular concentrations of the ribosome and TF are ca. 15 and 50  $\mu\text{M}$ , respectively, most ribosomes are likely to be occupied by one TF molecule leaving ca. 35  $\mu\text{M}$  of free TF in monomer-dimer equilibrium. In the presence of ribosome-bound nascent chains, the binding affinity of TF to the



ribosome can be enhanced by up to two orders of magnitudes; the off-rate of TF from the ribosome is also markedly slowed down remarkable (Kaiser *et al.*, 2006).

40 Although TF primarily acts on the ribosome in a co-translational manner, TF has been shown to exhibit chaperone activity in the absence of the ribosome to promiscuously facilitate folding of cytosolic proteins. The crystal structure of a thermophilic TF in complex with the ribosomal protein L9 shows that TF binds to its substrate in a 2:2 stoichiometry, demonstrating the multifaceted substrate recognition modes of TF (Martinez-Hackert and Hendrickson, 2009). In contrast to the bacterial chaperonin GroEL/S and DnaK/J systems, essential for many proteins to fold into native conformations, only a  
45 handful of *E. coli* proteins display TF-obligatory solubility (Niwa *et al.*, 2012). Nevertheless, TF plays an important role in working with SecA and SecB to regulate the membrane protein secretory pathway, as nascent membrane proteins need to be corrected sorted by TF and SecA/B as they emerge from the ribosome (Buskiewicz *et al.*, 2004; Gelis *et al.*, 2007; Huang *et al.*, 2016). It therefore raises the question as to how TF recognizes specific substrates when faced with a multitude of sequence variations of the bacterial proteome during co-translational and post-translational folding. To this end, an empirical scoring  
50 function for predicting TF binding motifs has been proposed based on peptide array analyses: a putative TF binding motif should be at least eight amino acids in length and contain both aromatic (phenylalanine, tyrosine and tryptophan) and positively charged lysine or arginine residues (Patzelt *et al.*, 2001). In a landmark study, Kalodimos and co-workers demonstrated that it requires three TF molecules to bind to one fully unfolded PhoA, which contains multiple TF binding sites with low  $\mu\text{M}$  binding affinities (Saio *et al.*, 2014). TF exhibits multiple substrate binding sites in SBD and PPI, which are evolutionarily conserved.  
55 The additivity of substrate binding affinities in the multiple binding sites on SBD and PPI results in much higher binding affinity and specificity for long PhoA fragments with an apparent dissociated constant ( $K_d$ ) in the nM range, rendering the ability to determine the solution structures of TF in complex with three different PhoA fragments by solution state nuclear magnetic resonance (NMR) spectroscopy. The structural information of TF in complex with various PhoA fragments indicated that TF preferentially binds to aromatic residues as well as large hydrophobic residues. There is no indication of the preference  
60 for positively charged lysine or arginine residues as previously predicted.

In this study, we sought to evaluate the predictive power of the empirical scoring function for TF binding motifs proposed by Bukau and co-workers (Deuerling *et al.*, 2003; Patzelt *et al.*, 2001). Combining methyl NMR and electron spin resonance (ESR) spectroscopy, we revealed the dynamic and polymorphic nature of the TF dimer. We generated a collection of fluorescein isothiocyanate (FITC) labeled peptide to validate by fluorescence polarization (FP) the predictive power of the  
65 proposed TF binding scoring function. Two TF binding peptides were subsequently spin-labeled for paramagnetic relaxation enhancement (PRE) measurements to identify multiple substrate binding sites within the SBD and PPI. Importantly, we demonstrated that dimerization of TF can sequester these binding motifs, and that the dynamic equilibrium between monomer and dimer is essential for substrate recognition. Collectively, our findings illustrated the functional importance of TF dimerization in the context of co-translational folding.

## 70 2. Materials and Methods

### 2.1 Purification of recombinant TF variants

The open reading frame of *E. coli* TF was obtained from the Nara *E. coli* ORF collection (<http://ecoli.aistnara.ac.jp/Resource/ResourceManage.jsp>) and had been subcloned into a pET-21d plasmid with a His<sub>6</sub>-tag at the N-terminus. In addition, the ribosome binding domain (RBD) corresponding to residues 1-117 was obtained by introducing  
75 a stop codon at position 118 through site-directed mutagenesis as described previously (Barbet-Massin *et al.*, 2015; Hsu and Dobson, 2009). The ORF of PPI corresponding to residues 150-250 was obtained by PCR and subsequently was subcloned into a pET-21d plasmid. The constructs of the SBD and RBD-truncated TF construct (TF<sub>113-432</sub>) were kind gifts from Prof. H. Jane Dyson (Scripps Institute, U.S.A.). The single cysteine mutants, Arg14Cys (14C), Thr150Cys (150C), Glu326Cys (326C),



and Ser376Cys (376C), (Kaiser *et al.*, 2006) were kind gifts from Prof. F. Ulrich Hartl (Max Planck Institute for Biochemistry, Germany). The plasmids of all TF variants were amplified using an *E. coli* DH5 $\alpha$  strain (Sigma-Aldrich, U. S. A.) with appropriate antibiotics selection, and their sequences were subsequently confirmed by standard DNA sequencing (Genomics, Taipei, Taiwan).

Unlabeled, uniformly  $^{15}\text{N}$  labeled, or uniformly  $^{15}\text{N}/^{13}\text{C}$  labeled protein samples were expressed by growing the transformed cells in Luria-Bertani (LB) medium or M9 minimal medium containing  $^{15}\text{NH}_4\text{Cl}$  (1 g/L) and  $^{13}\text{C}$  D-glucose (2 g/L) for uniformly  $^{15}\text{N}/^{13}\text{C}$  labeling in the presence of kanamycin or ampicillin for antibiotics selection. Selective  $^{13}\text{C}$  and  $^1\text{H}$  labeling at methyl groups of isoleucine ( $\delta 1$ ), leucine, valine, methionine and/or Ala  $\beta$  positions – U-[ $^{15}\text{N}$ ,  $^2\text{H}$ ], Ile-[ $\delta 1$ - $^{13}\text{C}_m$ ,  $^1\text{H}_m$ ], [Leu/Val-[ $^{13}\text{C}_m$ ,  $^1\text{H}_m$ ], Met-[ $^{13}\text{C}$ ,  $^1\text{H}$ ] and/or Ala-[ $\beta$ - $^{13}\text{C}_m$ ,  $^1\text{H}_m$ ] – was achieved by growing *E. coli* culture in perdeuterated M9 medium containing 99.9% D $_2\text{O}$ ,  $^{15}\text{NH}_4\text{Cl}$  (1 g/L) and  $^2\text{H}$  D-glucose (2 g/L) followed by addition of selectively  $^{13}\text{C}$  and  $^1\text{H}$  labeled metabolic precursors, and 100 mg/L  $^{13}\text{C}$ -labeled methionine (Cambridge Isotope Laboratory, U. S. A.) 30 min prior to IPTG induction as described previously. For selective methyl group-labeled samples, protein over-expression was carried out at 37 °C for four hours after the addition of IPTG. For the other samples, the overexpression was induced by the addition of 0.5 mM IPTG when the cell density reached OD $_{600}$  of 0.6–0.8 followed by overnight growth at 16 °C.

The cells were harvested by centrifugation using a Beckmann J20XP centrifuge with a JLA 8.1K rotor for 30 min with 6000 rpm at 4 °C and resuspended in buffer containing 50 mM potassium phosphate (pH 8.0) and 300 mM NaCl. The harvested cells were disrupted using a sonicator, and the cell debris and supernatant were separated by a second centrifugation step at 45,000 $\times$ g for 30 min at 4 °C. The supernatant was loaded onto a prepacked 5 ml His-Trap HP column (GE Healthcare Life Science) followed by extensive wash using buffer containing 20 mM imidazole to remove protein impurities and prevent non-specific binding. Target fusion protein was eluted using 250 mM imidazole with the same buffer background. The eluted fractions were pooled and subject to size-exclusion chromatography (SEC; HiLoad 26/60 Superdex 75, GE Healthcare Life Sciences) with 20 mM sodium phosphate (pH 7.4) and 100 mM NaCl to remove impurities to yield a purity of higher than 95% based on visual inspection of the Coomassie Brilliant Blue-stained sodium dodecyl sulfate polyacrylamide gel (SDS-PAGE). The protein solution was aliquoted, flash-frozen by liquid nitrogen and stored at -80 °C until further use. Unless otherwise specified, the SEC buffer was used for all the biophysical characterizations described herein.

## 2.2 Fluorescence polarization analysis of FITC-labeled peptides

Five peptides corresponding to fragments of ICDH (Table 1) were synthesized in-house. A fraction of all these peptides were subsequently labeled with FITC at the N-terminus for fluorescence polarization (FP) measurements. All peptides (with and without FITC labeling) were purified by high performance liquid chromatography (HPLC) and validated by MALDI-TOF mass spectrometry against their expected molecular weights. For FP analysis, FITC-labeled peptides were dissolved in DMSO to yield a stock solution of 4 M. They were subsequently diluted by 20mM Tris (pH 7.4) and 100mM NaCl to yield a molar concentration of 100  $\mu\text{M}$ . 200  $\mu\text{l}$  of TF variants were transferred into a 96 well plate followed by serial dilution by the same buffer in a 1:1 dilution ratio. The final protein concentrations were between 0.1  $\mu\text{M}$  and 1000  $\mu\text{M}$ . After serial dilution, 50  $\mu\text{l}$  of protein solutions of various protein concentrations were transferred to new wells by an eight-channel and were mixed with FITC-labeled peptide solution to yield a peptide concentration of 1  $\mu\text{M}$ . FP measurements of these samples were carried out using a plate reader (Paradigm, Molecular Device, U. S. A.) with an excitation wavelength of 485 nm and an emission wavelength of 535nm. The integration time was set to 250 ms. The observed FP values as a function of protein concentration were fit to a one-site binding model using the software Prism (GraphPad, U. S. A.) to extract apparent association constants associated with different combinations of FITC-labeled peptides and TF constructs (Lou *et al.*, 2014).



### 2.3 NMR paramagnetic relaxation enhancement analysis

One mg of IcdH2 and IcdH3 peptides were individually dissolved in 1 ml deionized water and pH adjusted to 7.6. A stock solution of *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonylthioate (MTSL) was prepared by dissolving MTSL powder in dimethyl sulfoxide (DMSO) to reach 150 mM. 10-fold MTSL was added to the peptide solution for overnight reaction at 4 °C in the dark. MTSL-labeled peptides were purified by high performance liquid chromatography (HPLC) and validated by MALDI-TOF mass spectrometry against their expected molecular weights. The eluents were lyophilized and resuspended in the SEC buffer to reach a concentration of 20 mM. U-[<sup>15</sup>N,<sup>2</sup>H], Ile-[δ1-<sup>13</sup>C<sub>m</sub>,<sup>1</sup>H<sub>m</sub>], Leu/Val-[<sup>13</sup>C<sub>m</sub>,<sup>1</sup>H<sub>m</sub>], Met-[<sup>13</sup>C, <sup>1</sup>H], Ala-[β-<sup>13</sup>C<sub>m</sub>,<sup>1</sup>H<sub>m</sub>] TF<sub>113-432</sub> and full-length TF were used for PRE measurements by recording the backbone <sup>15</sup>N-<sup>1</sup>H transverse relaxation optimized spectroscopy (TROSY) and side-chain methyl <sup>13</sup>C-<sup>1</sup>H band-selective optimized flip angle short transient heteronuclear multiquantum correlation (SOFAST-HMQC) spectroscopy in oxidized and reduced states. The NMR spectra were collected by using NMR spectrometers operating at a proton Larmor frequency of 850 MHz or 600 MHz, equipped with a cryogenic triple resonance TCI probe (Bruker, Germany), processed by NMRPipe and analyzed by NMRFAM-SPARKY (<https://nmrfam.wisc.edu/nmrfam-sparky-distribution/>). The nitroxide of MTSL was reduced by adding an aliquot of ascorbic acid to yield a final concentration of 1 mM. The observed PREs were expressed as the ratio of the peak intensities of the oxidized (paramagnetic state) over the reduced (diamagnetic state) state ( $I^{ox}/I^{red}$ ).

### Continuous-wave (CW) and pulsed ESR measurements

Introduction of MTSL into single cysteine TF variants, *i.e.*, 14C, 150C, 326C and 376C, was achieved by incubating the protein samples with 10 mM DTT, which was removed by using a desalting column (PD-10, GE Healthcare, USA). 10-fold molar excess of MTSL was added immediately after the removal of DTT and the mixtures were incubated overnight at 4 °C in the dark. Free MTSL was subsequently removed by using the same desalting column and complete MTSL incorporation was confirmed by mass spectrometry (Rezwave, Taiwan). A Bruker ELEXSYS E580-400 X-band cw/pulsed spectrometer, equipped with a split-ring resonator (EN4118X-MS3) and a helium gas flow system (4118CF and 4112HV), was used. CW ESR spectra were recorded at temperature 310 K, with an operating frequency of 9.4 GHz, 100-kHz field modulation and 1.5 mW incident microwave power. 0.25-0.6 mM TF variants in deuterated buffer were loaded in 3 mm (O.D.) quartz tubes. d8-glycerol was supplemented to achieve a final glycerol concentration of 30% (v/v). The total volume is approximately 20 μL. For the electron spin echo (ESE) measurements, sample tube was plunge-cooled in liquid nitrogen and then transferred into the ESR probe head, which was precooled to 50 K. ESE experiments were performed using the 2-pulse Hahn echo sequence, consisting of a  $\pi/2$  pulse along the x-axis followed by a delay  $\tau$  and a train of  $\pi$  pulses, separated by interpulse delays  $2\tau$  (Lai *et al.*, 2013; Zecevic *et al.*, 1998). The field was adjusted to optimize the spin echo, and the duration times of  $\pi/2$  and  $\pi$  pulses were set to 16 and 32 ns. As previously described [2], the ESE signals were fitted to a stretched exponential function to extract  $T_2$  values from the ESE data using the MATLAB software.

### Chemical crosslink

Bis [sulfosuccinimidyl] suberate (BS<sup>3</sup>; ThermoFisher, U. S. A.) was dissolving in 20 mM NaPi (pH 7.4), 100 mM NaCl to yield a stock solution of 20 mM. 100 μM of TF was crosslinked by 10-fold BS<sup>3</sup> for 30 min at room temperature. The reaction was quenched by 20 mM Tris-HCl (pH7.5) for 10 min. The crosslinked TF was analyzed by a 12% SDS-PAGE to confirm the level of crosslink.

### Biolayer interferometry (BLI)

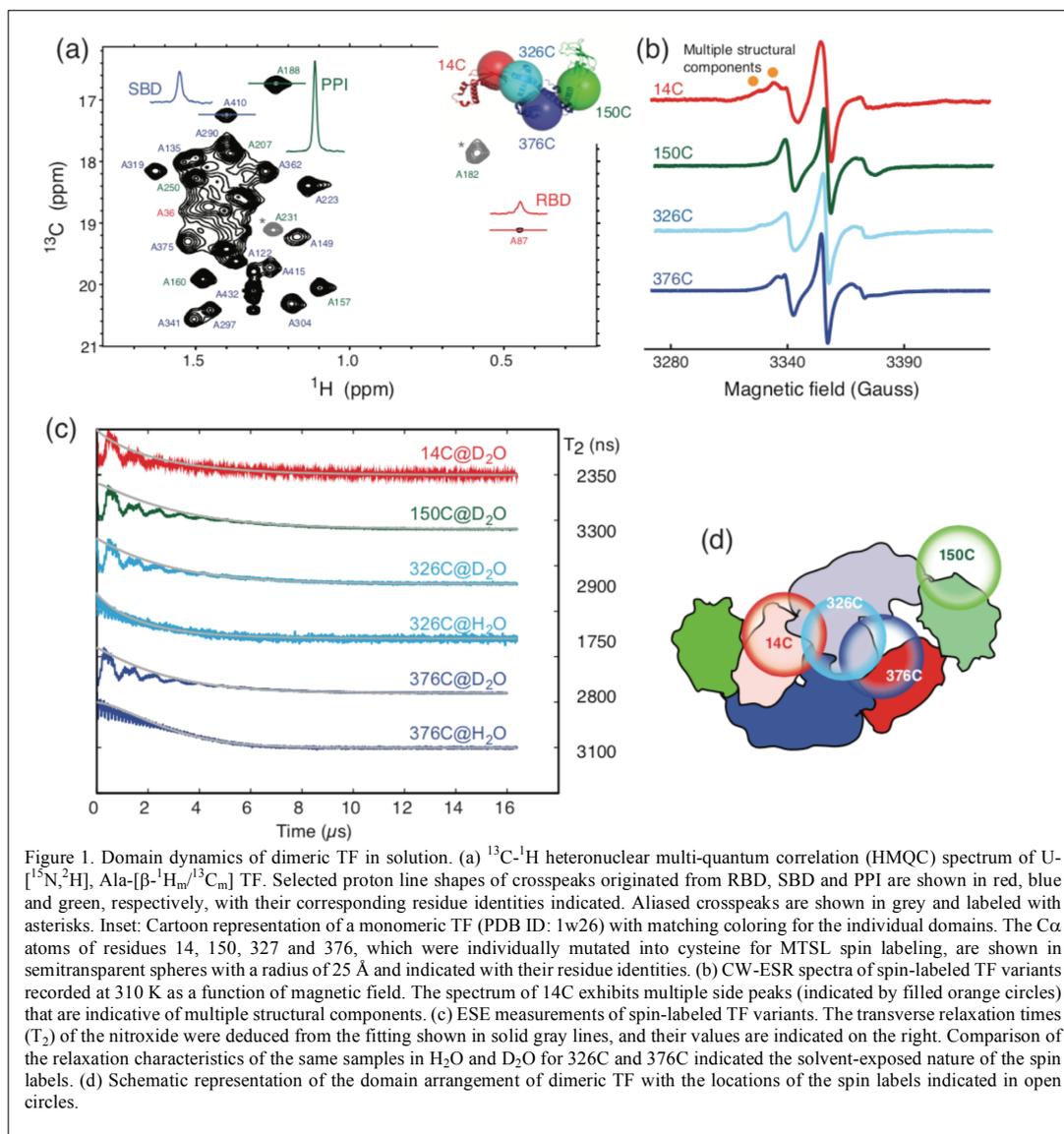
BLI measurements were carried out as described previously. Briefly, recombinant precursor maltose-binding protein (preMBP) was cloned from the Nara *E. coli* ORF collection and overexpressed in *E. coli* BL21 DE3 strain following the procedure as described previously (Weiss and Bassford, 1990). PreMBP was subsequently denatured in 8 M urea and rapidly



60 diluted to keep it in an unfolded form as described previously (Crooke *et al.*, 1988). Native and BS<sup>+</sup> crosslinked TF were immobilized onto the high precision streptavidin (SAX) biosensors (Satorius, U. S. A.) for binding to preMBP using a BLI instrument (Octet Red96, ForteBIO, U. S. A.). Global fitting to the kinetic traces and the steady state responses was achieved by using the built-in software of the BLI instrument.

## Results

65 TF exists in solution in a monomer-dimer equilibrium with a low  $\mu\text{M}$   $K_d$  (Kaiser *et al.*, 2006). RBD is responsible for the dimerization (Patzelt *et al.*, 2002). Isolated RBD forms a dimer in solution but complete NMR assignments could be accomplished with uniformly  $^{13}\text{C}$  and  $^{15}\text{N}$  labeling without deuteration despite the moderate molecular weight (28 kDa as a dimer) (Hsu and Dobson, 2009). The assignments were used for studying how RBD binds to the *E. coli* 70S ribosome using proton-detected solid state NMR spectroscopy (Barbet-Massin *et al.*, 2015). To facilitate further NMR-based analysis of full-length TF, we employed a divide-and-conquer assignment strategy to overcome the spectral overlap issue (Supplement Fig.





70 S1). In addition to RBD, we have previously reported the NMR assignments of PPI (Huang and Hsu, 2016). Dyson and  
workers have reported the NMR assignments of SBD and TF<sub>113-432</sub> (Yao *et al.*, 2008), and kindly shared the constructs for this  
study. In the absence of RBD, TF<sub>113-432</sub> (SBD+PPI) and SBD exhibited sharp line widths and homogeneous peak intensities in  
the 2D <sup>15</sup>N-<sup>1</sup>H backbone amide and <sup>13</sup>C-<sup>1</sup>H side-chain methyl correlation spectra under perdeuteration backgrounds, indicating  
75 that the molecules were monomeric and conformationally homogeneous. As reported by Dyson and co-workers, most of the  
backbone amide and side-chain methyl crosspeaks observed in SBD can be superimposed to those of TF<sub>113-432</sub>, indicating the  
modular nature of the domain structures (Supplement Fig. S2a). In the case of full-length TF, we observed a broad range of  
dynamic range in the NMR peak intensities although most of the resolved crosspeaks could be superimposed to those  
originating from RBD, SBD or PPI. Nevertheless, it was possible to assign most of the resolved resonances with the aid of the  
information derived from the individual domains and the combinations thereof. The previously reported NMR assignments of  
the full-length TF were essential for the confirmation of our assignments (Morgado *et al.*, 2017; Saio *et al.*, 2014; Saio *et al.*,  
80 2018).

To probe the dynamics TF, we selectively protonated and <sup>13</sup>C-label the methyl groups alanine residues of SBD, TF<sub>113-  
432</sub> and full-length TF under a perdeuteration background, and examined the line widths of their <sup>13</sup>C-<sup>1</sup>H side-chain methyl  
correlations. In contrast to the uniform line widths of monomeric SBD and TF<sub>113-432</sub> (Supplement Fig. S2a), full-length TF  
exhibited a very broad dynamic range (Supplement Fig. S2b). In particular, the resonances originated from RBD were so  
85 broadened that they became very challenging to detect even with the aid of perdeuteration and favorable dynamics of methyl  
NMR spectroscopy (Fig. 1a). Nevertheless, it was possible to demonstrate that the dimerization of TF made the most impact  
on the dynamics of RBD and to a lesser extent on SBD. In contrast, the alanine methyl resonances of PPI remained very sharp,  
suggesting that the dynamics of PPI was not significantly restricted by the dimerization. The same finding could be deduced  
for other methyl groups in isoleucine, leucine and valine residues.

90 If the dimer interface of TF were jointly formed by RBD and SBD, why would the methyl groups in RBD exhibit  
much broader NMR lines than those of SBD? To address this question, we individually introduced a spin label to one of the  
four sites in TF, namely residue 14 on RBD (14C), residue 150 on PPI (150C), and residues 326 and 376 on SBD (326C and  
376C), by covalently attaching a MTSL to the mutated cysteine side-chain (despite the long sequence of TF, it is void of  
cysteine, which greatly helped the MTSL labeling process). The spin-labeled TF variants were analyzed by ESR to probe the  
95 domain dynamics manifested in the line shapes and spin relaxation dynamics. Comparative analysis of the CW-ESR spectra  
the TF variants showed distinct side bands for 14C at 310K, suggesting the presence of multiple conformations (Fig. 1b). In  
contrast, the CW-ESR spectrum of 326C showed minor signals that were similar to those of 14C, but 376C did not exhibit the  
same signals, implying that the conformational heterogeneity of RBD is more pronounced than that of SBD. ESE analysis was  
subsequently used to deduce the transverse relaxation time ( $T_2$ ) of the free radical, *i.e.*, the nitroxide of MTSL, at individual  
100 sites (Fig. 1c). In line with the methyl NMR line width analysis, the  $T_2$  of 14C was the shortest (2350 ns), followed by 376C  
(2800 ns), 326C (2900 ns), and that of 150C was the longest (3300 ns). Further comparison of the time domain spin-echo ESR  
spectra of 326C and 376C in H<sub>2</sub>O and D<sub>2</sub>O showed a clear impact of solvent on the relaxation of the spin labels. The results  
indicated that both spin labels were solvent exposed despite their implication in dimer formation. Collectively, the NMR and  
ESR analyses suggested distinct domain dynamics of a dimeric TF with PPI being the least restricted, and RBD being the most  
05 heterogeneous. Although SBD also forms part of the dimer interface, its dynamics is less restricted than that of RBD.

Having established the ground work of characterizing the dynamics of TF in its apo form, we next set to characterize  
how TF recognizes its substrates. According to the peptide array study based on the sequence of isocitrate dehydrogenase  
(ICDH), several surface-immobilized peptides showed prominent TF binding (Deuerling *et al.*, 2003). Together with the results  
derived from other peptide arrays, an empirical scoring function for predicting the potential TF binding site along a given  
10 protein sequence was proposed (Deuerling *et al.*, 2003; Patzelt *et al.*, 2001). Nevertheless, the predictive power of such a  
scoring function has not been experimentally verified thus far. According to the prediction, an ideal TF binding motif should



be at least eight residues long, and rich in aromatic residues and positively charged lysine or arginine. The requirement for the coexistence of hydrophobic and charged residues is an intriguing feature. We plotted the predicted TF binding score as a function of ICDH sequence (Fig. 2a) and synthesized five segments from the ICDH sequence that fulfilled the requirement of peptide length and composition (Table 1). Their binding affinities were determined by fluorescence polarization (FP) using FITC labeling, which ranged between low  $\mu\text{M}$  to low mM, spanning more than two orders of magnitudes (Fig. 2b). Importantly, the natural logarithms of the observed dissociation constants ( $K_d$ ) showed a good correlation with the predicted binding score (an  $R^2$  value of 0.88 was obtained from the linear regression), demonstrating the predictive power of the empirical scoring function (Fig. 2c).

20 **Table 1. Synthetic peptides used for TF binding analyses**

Peptide name	Sequence	Length (residue number)	Predicted score	Length-normalized score	$K_d$ ( $\mu\text{M}$ )
IcdH1	KAYKGERKISWMEIYT	16	-8.9	-0.56	28.2 $\pm$ 1.0
IcdH2	YICLRPVRYQGT	13	-14.0	-1.08	8.6 $\pm$ 0.3
IcdH3	KFLREEMGVKKIRFPEHC	18	-9.9	-0.55	132 $\pm$ 9
IcdH4	KGNIMKFTGAFK	14	-5.6	-0.40	293 $\pm$ 49
IcdH5	GTAPKYAGQDK	12	-1.6	-0.13	1466 $\pm$ 2117

To further examine the structural basis of substrate recognition by TF, we chose IcdH2 and IcdH3, which had an endogenous cysteine residue that can be spin labeled with MTSL, for paramagnetic relaxation enhancement (PRE) measurements. We first used U- $^{15}\text{N}$ ,  $^2\text{H}$ ], Ile- $[\delta\text{-}^{13}\text{C}_m, ^1\text{H}_m]$ , [Leu/Val- $^{13}\text{C}_m, ^1\text{H}_m]$ , Ala- $[\beta\text{-}^{13}\text{C}_m, ^1\text{H}_m]$ , Met- $^{13}\text{C}$ ,  $^1\text{H}$ ] TF<sub>113-432</sub> to collect 2D  $^{15}\text{N}$ - $^1\text{H}$  backbone amide and  $^{13}\text{C}$ - $^1\text{H}$  side-chain methyl correlation spectra under oxidized (paramagnetic) and reduced (diamagnetic) states to determine the PREs defined by the resonance intensity ratios between the oxidized and reduced states,  $I^{\text{ox}}/I^{\text{red}}$  (Fig. 3). The observed PREs were mapped onto the structure of TF<sub>113-432</sub>, which revealed multiple hotspots within SBD and one cluster within PPI that showed strong PREs (Fig. 4 and Fig. 5). Although the binding affinity of IcdH3 to full-length TF ( $K_d = 132 \pm 9 \mu\text{M}$ ) is weaker than that of IcdH2 ( $K_d = 8.6 \pm 0.3 \mu\text{M}$ ), the observed PREs in IcdH3 were more prominent than that of IcdH2 when TF<sub>113-432</sub>, which is a truncated and monomeric form of TF (Supplement Fig. S3), was used in the NMR PRE analysis (Fig. 4a and Fig. 5a). When full-length TF was used for the same NMR PRE analysis under the TF concentration that it is predominantly dimeric, the PREs were significantly reduced, and the remaining PREs were mostly localized within PPI that is not part of the TF dimer interface (Fig. 4b and Fig. 5b). The loss of PRE was much more pronounced for IcdH3 compared to that of IcdH2, in line with the FP analysis that showed a weaker TF binding for IcdH3 compared to

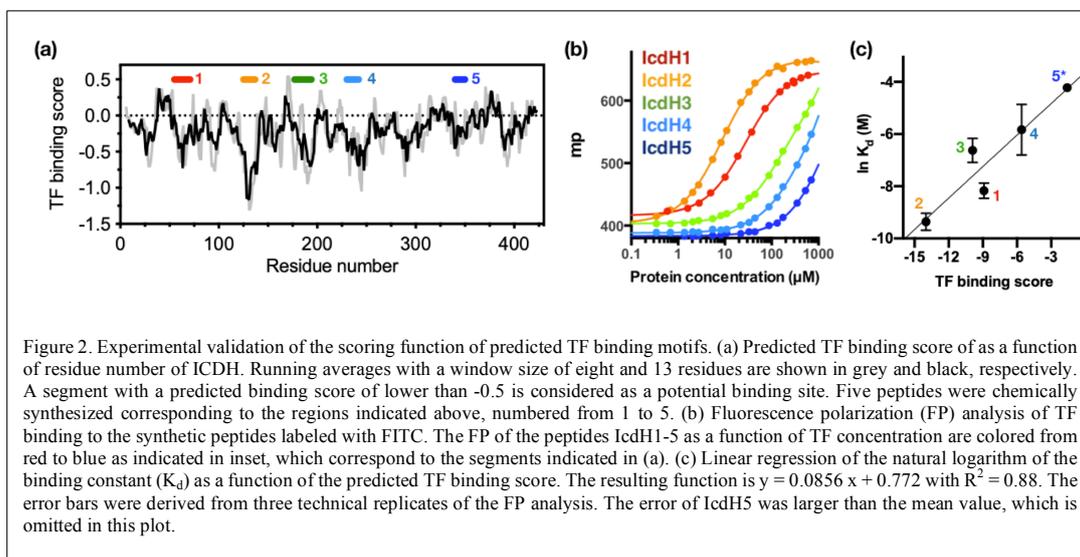
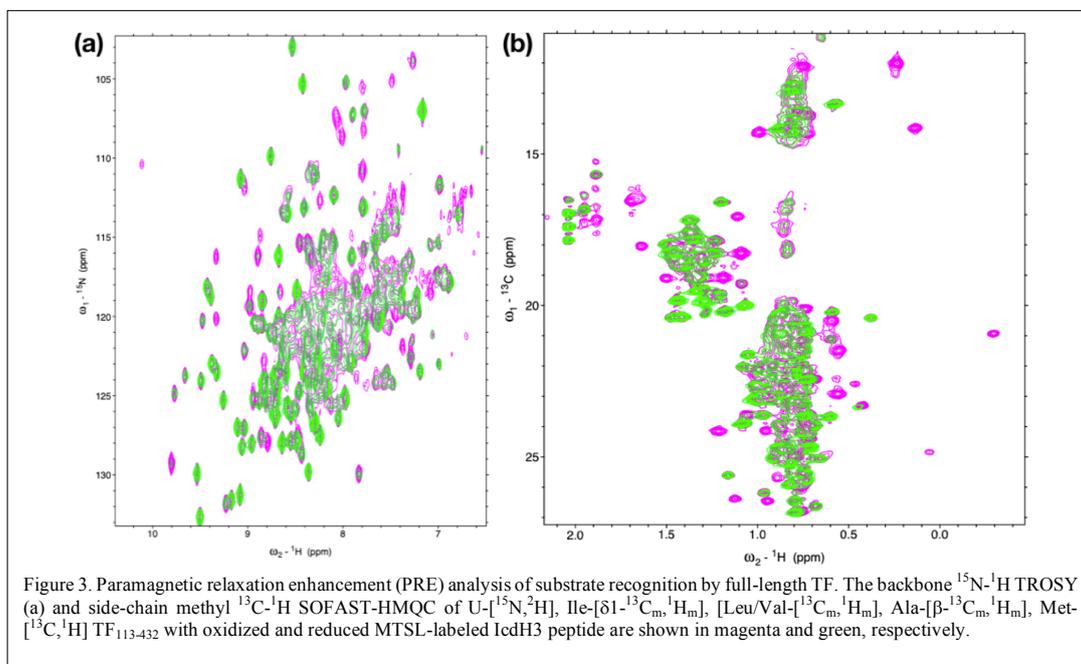
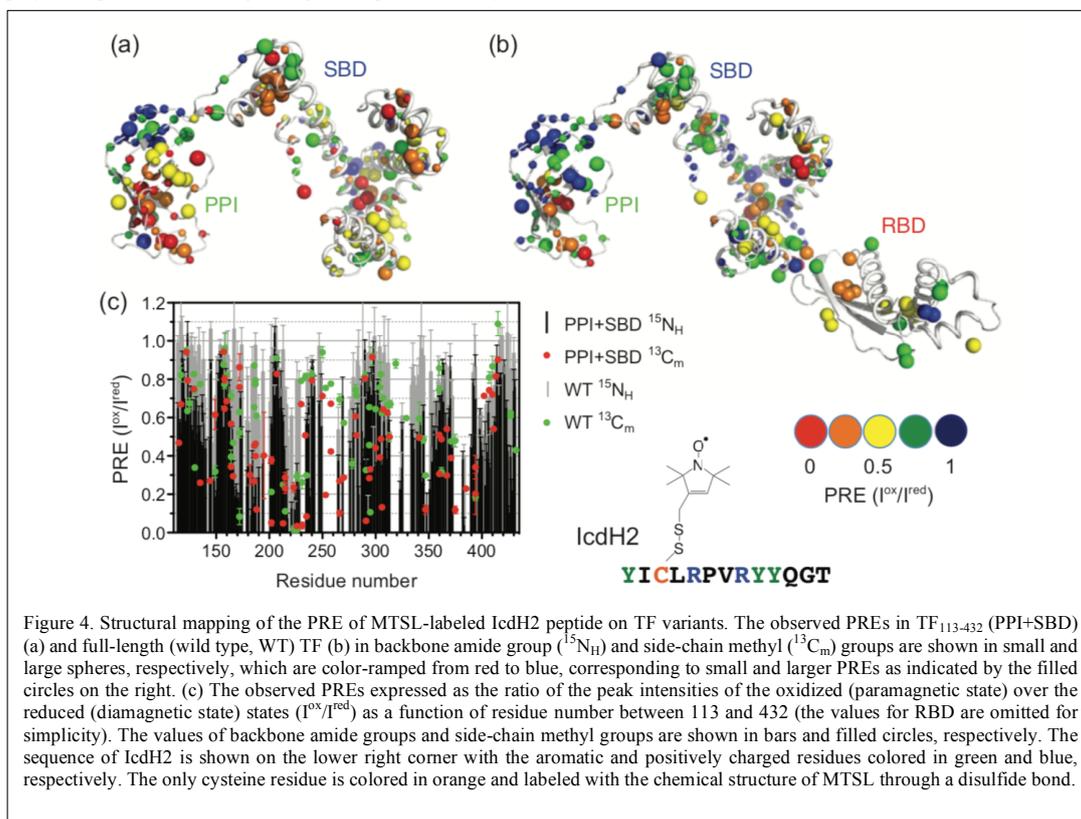


Figure 2. Experimental validation of the scoring function of predicted TF binding motifs. (a) Predicted TF binding score as a function of residue number of ICDH. Running averages with a window size of eight and 13 residues are shown in grey and black, respectively. A segment with a predicted binding score of lower than -0.5 is considered as a potential binding site. Five peptides were chemically synthesized corresponding to the regions indicated above, numbered from 1 to 5. (b) Fluorescence polarization (FP) analysis of TF binding to the synthetic peptides labeled with FITC. The FP of the peptides IcdH1-5 as a function of TF concentration are colored from red to blue as indicated in inset, which correspond to the segments indicated in (a). (c) Linear regression of the natural logarithm of the binding constant ( $K_d$ ) as a function of the predicted TF binding score. The resulting function is  $y = 0.0856x + 0.772$  with  $R^2 = 0.88$ . The error bars were derived from three technical replicates of the FP analysis. The error of IcdH5 was larger than the mean value, which is omitted in this plot.



IcdH2. The implication of this finding is that the dimerization of TF sequesters the substrate binding sites within SBD and to a lesser extent the binding site in PPI. Dynamic equilibrium between the monomeric and dimeric TF is therefore expected to play an important role in regulating its chaperone activity.



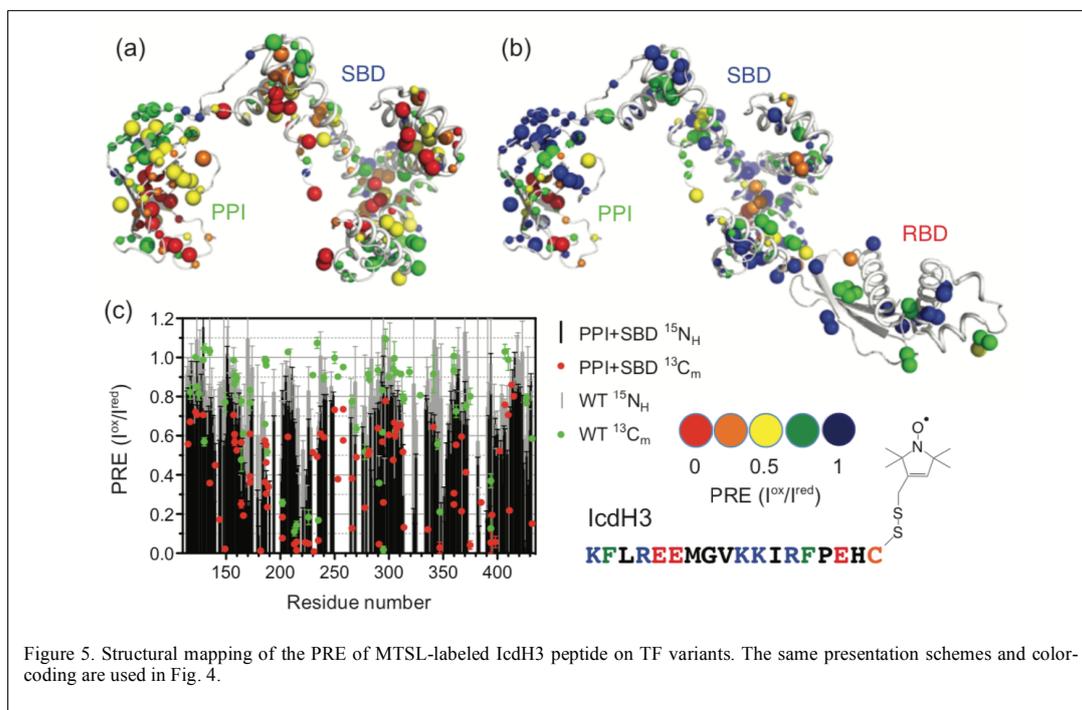


Figure 5. Structural mapping of the PRE of MTSL-labeled IcdH3 peptide on TF variants. The same presentation schemes and color-coding are used in Fig. 4.

40 To test this hypothesis, we used a chemical crosslinker, BS<sup>3</sup>, to lock TF in its dimeric form and examine how the inability  
 to dissociate into monomers may impact on the chaperone activity of TF (Fig. 6a). We used urea-denatured precursor maltose-  
 binding protein (preMBP) as a model substrate to study substrate recognition by TF. Unlike the synthetic peptides that we  
 used in the FP and NMR PRE analyses, preMBP is a large polypeptide chain that affords multiple TF binding sites, which can  
 enhance the binding through avidity. We employed BLI to compare the preMBP binding of native TF and BS<sup>3</sup>-crosslinked  
 45 TF. Due to the complex binding process between TF and preMBP, involving multiple binding sites from both molecules, a  
 simplified one site binding model cannot fully account for the observed sensorgrams (Fig. 6b-e). Nevertheless, the effect of

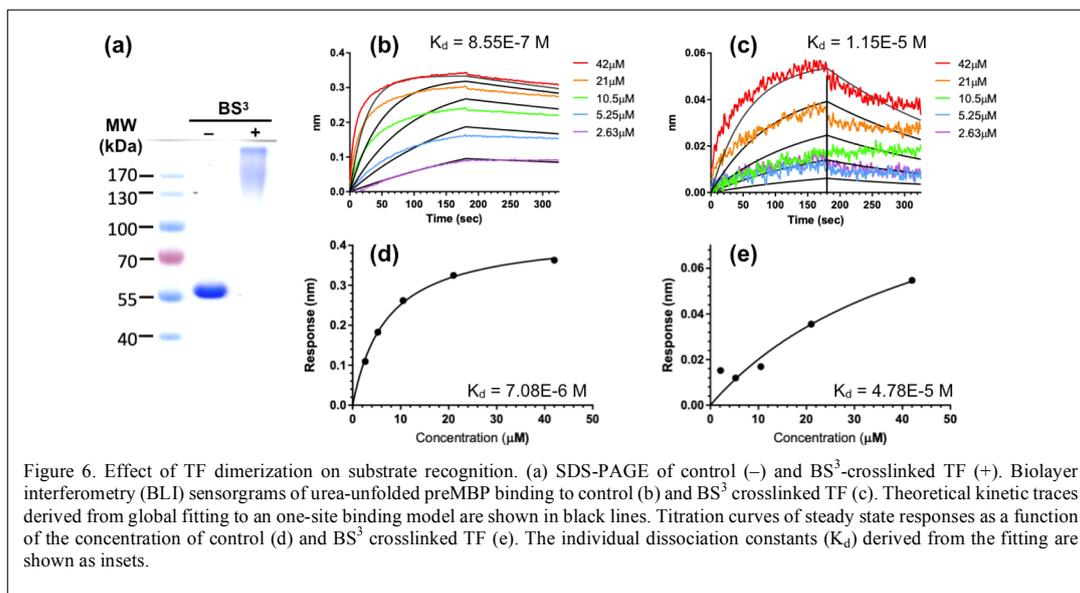


Figure 6. Effect of TF dimerization on substrate recognition. (a) SDS-PAGE of control (-) and BS<sup>3</sup>-crosslinked TF (+). Biolayer interferometry (BLI) sensorgrams of urea-unfolded preMBP binding to control (b) and BS<sup>3</sup> crosslinked TF (c). Theoretical kinetic traces derived from global fitting to a one-site binding model are shown in black lines. Titration curves of steady state responses as a function of the concentration of control (d) and BS<sup>3</sup> crosslinked TF (e). The individual dissociation constants ( $K_d$ ) derived from the fitting are shown as insets.



crosslinking of TF on its substrate recognition was apparent. Native TF bound to urea-denatured preMBP in the sub- $\mu$ M range whereas the amount of substrate-binding competent population in the BS<sup>3</sup>-crosslinked TF was reduced by almost 10-fold, and the corresponding binding affinity was also reduced by almost 10-fold. The results confirmed our hypothesis that substrate binding requires the exposure of the multiple binding sites within SBD that are sequestered upon dimer formation.

### 3. Discussion

TF is a unique molecular chaperone in that it is the first molecular chaperone that all newly synthesized nascent polypeptide chains encounter (Hoffmann *et al.*, 2012; Kaiser *et al.*, 2006). While there is good NMR evidence to demonstrate that a ribosome-bound nascent chain can fold into its native conformation without the aid of TF (Cabrita *et al.*, 2016; Cabrita *et al.*, 2009; Hsu *et al.*, 2007; Waudby *et al.*, 2019), there are indeed a handful of proteins within the *E. coli* proteome that require the contributions of TF to fold correctly (Niwa *et al.*, 2012). Unlike GroEL/GroES and DnaK/DnaJ, which require ATP hydrolysis to confer their chaperone activities, TF lacks the ATPase activity (Hartl and Hayer-Hartl, 2002). Instead, TF acts as a holdase. By holding onto its substrates, TF can remodel the folding pathway of the client proteins (Mashaghi *et al.*, 2013). It can also act in sync with other molecular chaperones, namely SecA and SecB, that target membrane proteins to the appropriate membrane localizations through the secretory pathway. Although the obligatory clients of TF constitute only a small fraction of the *E. coli* proteome (Niwa *et al.*, 2012), the question of how TF selectively recognizes its client proteins is of fundamental importance in light of the biomedical implications of protein misfolding and proteostasis (Balchin *et al.*, 2016; Dobson, 2003; Hartl and Hayer-Hartl, 2002).

In this study, we employed methyl NMR and ESR to characterize the dynamics of TF in its dimeric form. Although TF exists in equilibrium between monomer and dimer, the experimental conditions under which the NMR and ESR experiments were conducted, *i.e.*, protein concentrations above 0.25 mM, ensured that TF is predominantly dimeric. Our findings indicated that the domain dynamics of RBD, which is responsible for the dimerization of TF, is highly restricted. There is also good ESR evidence to suggest that RBD and, to a lesser extent, parts of SBD exist in multiple conformations. In contrast, the domain dynamics of PPI is not constrained by the dimer formation, which is in line with the anti-parallel dimeric assembly of TF as reported previously (Morgado *et al.*, 2017; Saio *et al.*, 2014; Saio *et al.*, 2018).

We next generated five FITC-labeled peptides derived from ICDH to demonstrate the predicted power of the empirical scoring function for TF binding based on the sequence composition. Two peptides, *i.e.*, IcdH2 and IcdH3, were spin labeled to map their binding sites on TF based on the NMR PREs. We identified three distinct binding sites within SBD and one binding site within PPI (Fig. 4 and Fig. 5). The locations of these binding sites are consistent with a previous study in which four disordered fragments of PhoA are used to map the binding sites on TF (Saio *et al.*, 2014). Note that in the previous study, TF<sub>113-432</sub> was used to determine the solution structures of TF in complex with different PhoA fragments based on intermolecular NOEs, while full-length TF was used to demonstrate that full length PhoA in its unfolded form can be occupied by multiple TF molecules. Crucially, the study also determined the microscopic  $K_d$  values for individual binding sites in the low  $\mu$ M range. The multivalency of substrate recognition significantly increases the binding affinity to a nM range.

According to ESR analysis, the spin labels within RBD and SBD are mostly solvent exposed. Furthermore, the dimer interface appeared to be quite heterogeneous and dynamic, according to methyl NMR and CW-ESR line shaper analyses (Fig. 1). The unique domain architecture of TF suggests that the dimer interface does not form a properly encapsulated cavity to accommodate its substrates. Additionally, the distributions of sparsely negatively charged surfaces surrounded by small patches of neutral (hydrophobic) surfaces within SBD and PPI coincide with the observed peptide binding sites, which could explain why positively charged residues and aromatic residues are both favored for TF binding (Fig. 7). Unlike GroEL/GroES, which has an efficient nucleotide-dependent regulatory mechanism to mechanically control the exposure of its substrate binding sites, TF relies on the self-dimerization to achieve the same regulation (Hartl and Hayer-Hartl, 2002).



Here we showed by comparing the peptide binding induced PREs in TF<sub>113-432</sub> and full-length TF that dimerization of TF effectively sequesters the binding sites within SBD from peptide binding. Although PPI is not involved in dimer formation, the peptide induced PREs in PPI are also diminished potentially due to steric hinderance. Considering that the effective peptide binding affinities are relatively weak compared to the dissociation constant of TF self-association, it is not surprising that the TF dimerization can outcompete peptide binding at a relatively high TF concentration (100  $\mu$ M). Nevertheless, cytosolic TF concentration is estimated to be in the range of 35  $\mu$ M while that of the ribosome is about 1  $\mu$ M. TF binds to the ribosome in a 1:1 stoichiometry, and the associated binding affinity is strongly modulated by the presence and compositions of fledgling nascent chains. Considering that a long nascent chain can be occupied by multiple TF molecules, the effective free TF concentration in the cytosol would be significantly lower than 30  $\mu$ M. In that case, some of the protein sequences could compete with the dimerization of TF thereby enabling TF to exert its holdase activity (Fig. 7).

#### 4. Conclusion

The intricate interplay between TF, nascent chains, and the ribosome can be modulated by the sequence compositions of the nascent chains. Our NMR, ESR and biophysical findings revealed the structural plasticity of individual domains of TF and the transient binding events in the monomeric form of TF, which were lost upon dimer formation. These results led us to propose that the relatively strong ribosome binding affinity serves as the key regulatory mechanism to modulate monomer-dimer equilibrium and therefore the accessibility of the substrate binding sites, which are fully exposed when TF binds to the ribosome through its RBD. The observed binding affinities of the selected peptides from ICDH indeed fit well within the dynamic range of these binding events.

#### Data availability

Data are available upon request

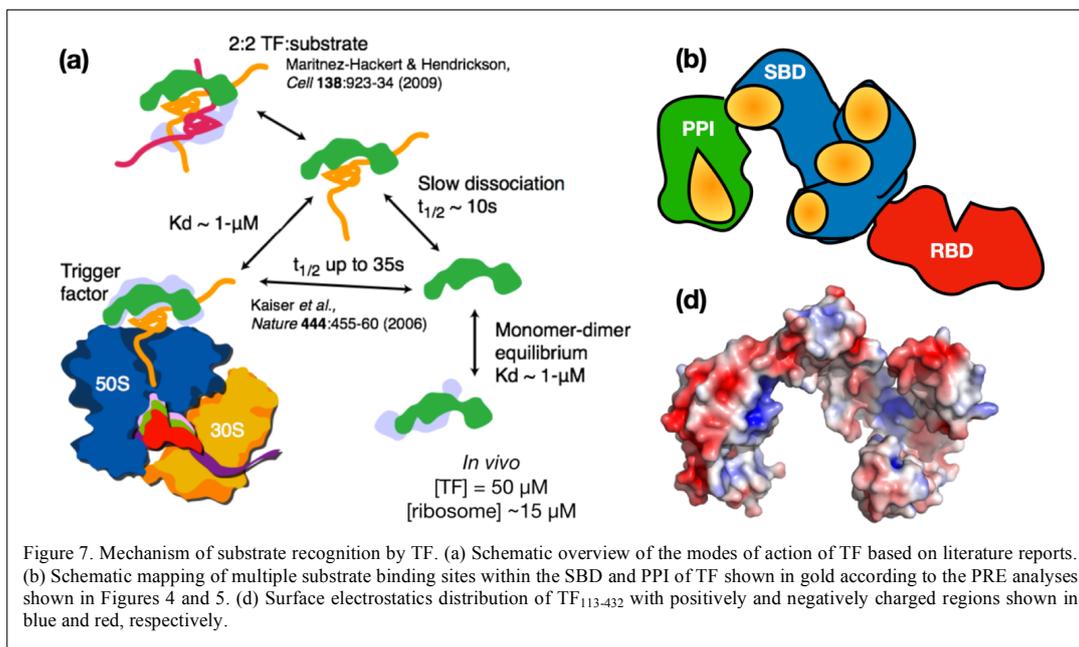


Figure 7. Mechanism of substrate recognition by TF. (a) Schematic overview of the modes of action of TF based on literature reports. (b) Schematic mapping of multiple substrate binding sites within the SBD and PPI of TF shown in gold according to the PRE analyses shown in Figures 4 and 5. (d) Surface electrostatics distribution of TF<sub>113-432</sub> with positively and negatively charged regions shown in blue and red, respectively.



## Supplement

### Author contributions

10 STDH conceived and designed the experiments with contributions from YWC for ESR. CTH prepared the NMR and ESR samples and assisted NMR data collection and analysis. YCL and YWC collected and analyzed the ESR data. CTH and SYC contributed to the FITC-labeled peptide binding analyses supported by MRH. MRH collected and analyzed the BLI data. STDH wrote the manuscript with inputs from all authors.

### Competing interest

15 The authors declare that they have no conflict of interest

### Acknowledgements

This article is dedicated to the 80<sup>th</sup> birthday of Prof. Robert Kaptein as part of the special issue "Robert Kaptein Festschrift". We thank Prof. Ulrich Hartl for stimulating discussions and sharing the constructs of TF variants. The NMR data were collected at the High Field NMR Center at Academia Sinica, the Instrumentation Center at the National Tsing Hua University (NTHU), supported by the Ministry of Science and Technology, Taiwan (MOST) and NTHU, and the Instrument Center at the National Taiwan University. The ESR data were collected at the Instrumentation Center at NTHU. The FITC-labeled peptides were synthesized by and analyzed at the Synthesis Facility and Biophysics Facility, respectively, at the Institute of Biological Chemistry, Academia Sinica. The BLI data were collected at the Biophysics Core Facility, Academia Sinica, funded by Academia Sinica Core Facility and Innovative Instrument Project (AS-CFII108-111).

### 25 Financial support

This research has been supported by MOST, Taiwan (grant numbers 100-2113-M-001-031-MY2 and 102-2113-M-001-017 - MY2 to STDH) and the intramural fund from Academia Sinica, Taiwan, to STDH.

### References

- Balchin, D., Hayer-Hartl, M., and Hartl, F.U., *In vivo* aspects of protein folding and quality control. *Science* 353, aac4354. doi: 10.1126/science.aac4354, 2016.
- 30 Barbet-Massin, E., Huang, C.-T., Daebel, V., Hsu, S.-T.D., and Reif, B., Site-specific solid-state NMR studies of "Trigger Factor" in complex with the large ribosomal subunit 50S. *Angew. Chem. Intl. Ed.* 54, 4367-4369. doi: 10.1002/anie.201409393, 2015.
- Buskiewicz, I., Deuerling, E., Gu, S.Q., Jockel, J., Rodnina, M.V., Bukau, B., and Wintermeyer, W., Trigger factor binds to ribosome-signal-recognition particle (SRP) complexes and is excluded by binding of the SRP receptor. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7902-7906. doi: 10.1073/pnas.0402231101, 2004.
- 35 Cabrita, L.D., *et al.*, A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat. Struc. Mol. Biol.* 23, 278-285. doi: 10.1038/nsmb.3182, 2016.
- Cabrita, L.D., Hsu, S.-T.D., Launay, H., Dobson, C.M., and Christodoulou, J., Probing ribosome-nascent chain complexes produced *in vivo* by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 106, 22239-22244. doi: 10.1073/pnas.0903750106, 2009.
- 40 Crooke, E., Brundage, L., Rice, M., and Wickner, W., ProOmpA spontaneously folds in a membrane assembly competent state which trigger factor stabilizes. *EMBO J.* 7, 1831-1835. doi: 1988.
- Deuerling, E., *et al.*, Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.* 45, 1317-1328. doi: 2003.



- Dobson, C.M., Protein folding and misfolding. *Nature* 426, 884-890. doi: 10.1038/nature02261, 2003.
- Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., and Ban, N., Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature* 431, 590-596. doi: 10.1038/nature02899, 2004.
- Gelis, I., Bonvin, A.M., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A., and Kalodimos, C.G.,  
50 Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* 131, 756-769. doi: 10.1016/j.cell.2007.09.039, 2007.
- Hartl, F.U., Cellular Homeostasis and Aging. *Ann. Rev. Biochem.* 85, 1-4. doi: 10.1146/annurev-biochem-011116-110806, 2016.
- Hartl, F.U., and Hayer-Hartl, M., Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295,  
55 1852-1858. doi: 10.1126/science.1068408, 2002.
- Hesterkamp, T., and Bukau, B., The *Escherichia coli* trigger factor. *FEBS Lett.* 389, 32-34. doi: 1996.
- Hoffmann, A., Becker, A.H., Zachmann-Brand, B., Deuerling, E., Bukau, B., and Kramer, G., Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol. Cell* 48, 63-74. doi: 10.1016/j.molcel.2012.07.018, 2012.
- 60 Hoffmann, A., Bukau, B., and Kramer, G., Structure and function of the molecular chaperone Trigger Factor. *Biochim. Biophys. Acta* 1803, 650-661. doi: 10.1016/j.bbamcr.2010.01.017, 2010.
- Hsu, S.-T.D., and Dobson, C.M., <sup>1</sup>H, <sup>15</sup>N and <sup>13</sup>C assignments of the dimeric ribosome binding domain of trigger factor from *Escherichia coli*. *Biomol. NMR Assign.* 3, 17-20. doi: 10.1007/s12104-008-9130-8, 2009.
- Hsu, S.-T.D., Fucini, P., Cabrita, L.D., Launay, H., Dobson, C.M., and Christodoulou, J., Structure and dynamics of a  
65 ribosome-bound nascent chain by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 104, 16516-16521. doi: 10.1073/pnas.0704664104, 2007.
- Huang, C., Rossi, P., Saio, T., and Kalodimos, C.G., Structural basis for the antifolding activity of a molecular chaperone. *Nature* 537, 202-206. doi: 10.1038/nature18965, 2016.
- Huang, C.-T., and Hsu, S.-T.D., NMR assignments of the peptidyl-prolyl cis-trans isomerase domain of trigger factor from *E. coli*. *Biomol. NMR Assign.*s 10, 149-152. doi: 10.1007/s12104-015-9655-6, 2016.
- 70 Kaiser, C.M., Chang, H.C., Agashe, V.R., Lakshminpathy, S.K., Etschells, S.A., Hayer-Hartl, M., Hartl, F.U., and Barral, J.M., Real-time observation of trigger factor function on translating ribosomes. *Nature* 444, 455-460. doi: 10.1038/nature05225, 2006.
- Lai, Y.C., Chen, Y.F., and Chiang, Y.W., ESR study of interfacial hydration layers of polypeptides in water-filled nanochannels and in vitrified bulk solvents. *PLoS ONE* 8, e68264. doi: 10.1371/journal.pone.0068264, 2013.
- 75 Lakshminpathy, S.K., *et al.*, Identification of nascent chain interaction sites on trigger factor. *J. Biol. Chem.* 282, 12186-12193. doi: 10.1074/jbc.M609871200, 2007.
- Lou, Y.-C., Wang, I., Rajasekaran, M., Kao, Y.-F., Ho, M.-R., Hsu, S.-T.D., Chou, S.-H., Wu, S.-H., and Chen, C., Solution structure and tandem DNA recognition of the C-terminal effector domain of PmrA from *Klebsiella pneumoniae*. *Nucleic Acids Research* 42, 4080-4093. doi: 10.1093/nar/gkt1345, 2014.
- 80 Martinez-Hackert, E., and Hendrickson, W.A., Promiscuous substrate recognition in folding and assembly activities of the trigger factor chaperone. *Cell* 138, 923-934. doi: 10.1016/j.cell.2009.07.044, 2009.
- Mashaghi, A., Kramer, G., Bechtluft, P., Zachmann-Brand, B., Driessen, A.J., Bukau, B., and Tans, S.J., Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* 500, 98-101. doi: 10.1038/nature12293, 2013.
- 85 Merz, F., *et al.*, Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *EMBO J.* 27, 1622-1632. doi: 10.1038/emboj.2008.89, 2008.
- Morgado, L., Burmann, B.M., Sharpe, T., Mazur, A., and Hiller, S., The dynamic dimer structure of the chaperone Trigger Factor. *Nat. Comm.* 8, 1992. doi: 10.1038/s41467-017-02196-7, 2017.



- 90 Niwa, T., Kanamori, T., Ueda, T., and Taguchi, H., Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. U. S. A.* 109, 8937-8942. doi: 10.1073/pnas.1201380109, 2012.
- Patzelt, H., Kramer, G., Rauch, T., Schonfeld, H.J., Bukau, B., and Deuerling, E., Three-state equilibrium of *Escherichia coli* trigger factor. *Biol. Chem.* 383, 1611-1619. doi: 10.1515/BC.2002.182, 2002.
- Patzelt, H., *et al.*, Binding specificity of *Escherichia coli* trigger factor. *Proc. Natl. Acad. Sci. U. S. A.* 98, 14244-14249. doi: 10.1073/pnas.261432298, 2001.
- 95 Rutkowska, A., Mayer, M.P., Hoffmann, A., Merz, F., Zachmann-Brand, B., Schaffitzel, C., Ban, N., Deuerling, E., and Bukau, B., Dynamics of trigger factor interaction with translating ribosomes. *J. Biol. Chem.* 283, 4124-4132. doi: 10.1074/jbc.M708294200, 2008.
- Saio, T., Guan, X., Rossi, P., Economou, A., and Kalodimos, C.G., Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* 344, 1250494. doi: 10.1126/science.1250494, 2014.
- 00 Saio, T., Kawagoe, S., Ishimori, K., and Kalodimos, C.G., Oligomerization of a molecular chaperone modulates its activity. *eLife* 7. doi: 10.7554/eLife.35731, 2018.
- Waudby, C.A., Dobson, C.M., and Christodoulou, J., Nature and regulation of protein folding on the ribosome. *Trends Biochem. Sci.* 44, 914-926. doi: 10.1016/j.tibs.2019.06.008, 2019.
- Weiss, J.B., and Bassford, P.J., Jr., The folding properties of the *Escherichia coli* maltose-binding protein influence its interaction with SecB *in vitro*. *J. Bacteriol.* 172, 3023-3029. doi: 10.1128/jb.172.6.3023-3029.1990, 1990.
- 05 Yao, Y., Bhabha, G., Kroon, G., Landes, M., and Dyson, H.J., Structure discrimination for the C-terminal domain of *Escherichia coli* trigger factor in solution. *J. Biomol. NMR* 40, 23-30. doi: 10.1007/s10858-007-9207-1, 2008.
- Zecevic, A.N.A., Eaton, G.R., Eaton, S.S., and Lindgren, M., Dephasing of electron spin echoes for nitroxyl radicals in glassy solvents by non-methyl and methyl protons. *Mol. Phys.* 95, 1255-1263. doi: 10.1080/00268979809483256, 1998.

10