Supporting Information - Trigger factor slows co-translational folding through kinetic trapping while sterically protecting the nascent chain from aberrant cytosolic interactions

Edward P. O'Brien,[1] John Christodoulou,[2] Michele Vendruscolo,[1] and Christopher M. Dobson[1]

[1]Department of Chemistry, University of Cambridge, Cambridge, UK

[2]Institute of Structural and Molecular Biology, University College London and Birkbeck College, UK

Correspondence: cmd44@cam.ac.uk

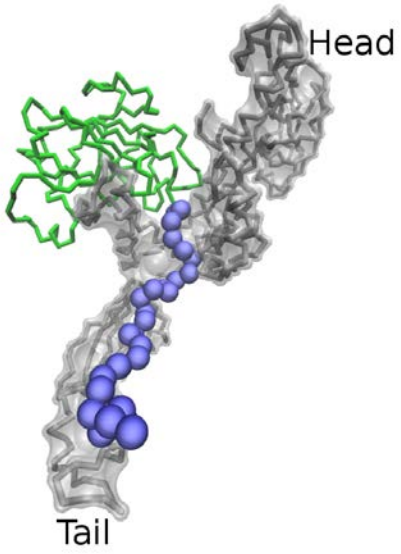

Figure S1, related to Figure 5d. Tertiary folding of β-galactosidase outside of TF's cradle requires an unfolded nascent chain segment to traverse back to the exit tunnel opening. In this structure a portion of β-galactosidase (green) is shown acquiring native tertiary structure outside the cradle of TF (gray); taken from RNC simulations at 310 K at a nascent chain length of 247 residues. The nascent chain segment that is unfolded is shown in blue. The perspective of this structure is from that of looking out of the exit tunnel vestibule into the cradle of TF. Residues 1 through 214 of the nascent chain are shown, with residue 214's position being at the start of the exit vestibule.

MODELS AND METHODS

Force-field for the ribosome nascent chain complex interacting with trigger factor. Protein molecules, including the ribosomal proteins, nascent chains, and trigger factor, were modeled based on a variant[1] of the coarse-grained C_α protein representation of Karanicolas and Brooks.[2] In this model each amino-acid is represented by one interaction site that is located at the C_α backbone position. The associated Hamiltonian consists of bond, bond-angle, dihedral-angle, electrostatic, and native and non-native Lennard-Jones energy terms. Except for some of the Lennard-Jones energy terms, all other energies are non-Go and transferable between different protein molecules. The energy E of a protein configuration in this model is calculated as:

$$
\begin{aligned}
E = & \sum_i k_b (r_i - r_0)^2 + \sum_i \sum_j^5 k_{\varphi,ij}\left(1 + \cos\left[j\varphi_i - \delta_{ij}\right]\right) \\
& + \sum_i \exp[-\gamma k_\alpha (\theta_i - \theta_\alpha)^2 + \varepsilon_\alpha] + \exp\left[-\gamma k_\beta \left(\theta_i - \theta_\beta\right)^2\right] \\
& + \sum_{ij} \frac{q_i q_j e^2}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \exp\left[-\frac{r_{ij}}{\mathrm{l_D}}\right] \\
& + \sum_{ij \in \{Native\}} (\Theta_{\mathrm{NC}} + \Theta_{\mathrm{TF}}) \varepsilon_{ij} \left(13\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 18\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10} + 4\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \\
& + \sum_{i \in \{TF\}} \sum_{j \in \{NC\}} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) + \sum_{i \in \{TF\}} \sum_{j \in \{R\}} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \\
& + \sum_{ij \notin \{Native\}} \varepsilon_{ij} \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}
\end{aligned}
$$

From left to right, these energy terms correspond to summations over the proteins' C_α- C_α bond lengths, dihedral-angles, bond-angles, electrostatic interactions, native Lennard-Jones (LJ) interactions within the nascent chain (NC) and within TF, LJ interactions between TF and the nascent chain, native LJ interactions between TF and the ribosomal binding site (R), and finally the last summation is over non-native interactions between the components. The parameters for the first four energy terms have been described in detail elsewhere.[1,2] For the Debye-Huckel treatment of electrostatic interactions we used a Debye length of $\mathrm{l_D} = 10$ Å, which corresponds to that in the *E. coli* cytoplasm,[3] and a dielectric of 78.5. Lysine and arginine interaction sites were assigned charges of q=+1*e*, and glutamine and aspartate residues were assigned charges of q=-1*e*. All other protein interaction sites had a charge of zero.

Intra-protein LJ interactions were calculated in the fifth summation term. Θ_{NC} and Θ_{TF} are step functions that have a value of 1 when interaction sites *i* and *j* are both within the nascent chain or TF,

respectively, and 0 otherwise. The Lennard-Jones well-depths between interaction sites *i* and *j*, ε_{ij}, that make contacts in the crystal structure are proportional to the Bentancourt-Thirumalai statistical potential [4]. The set of ε_{ij} values were scaled by a multiplicative factor as described in the next section. Collision diameters σ_{ij} between the C_α interaction sites were set equal to their distance in the crystal structure divided by $2^{1/6}$.

Inter-protein TF, nascent chain LJ interactions were calculated in the sixth energy term, where the summation i is over TF interaction sites and summation j is over the nascent chain interaction sites. The LJ interactions between these two components are transferable. The ε_{ij} values are proportional to the Betancourt-Thirumalai statistical potential as described below. The collision diameters for interactions between TF and the nascent chain were calculated as $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, where σ_i and σ_j are defined in Table S2.

Native inter-protein LJ interactions were calculated between TF and the ribosomal binding site in the seventh summation term. The LJ well-depths, ε_{ij} are uniform and determined as described in the next section. The collision diameters for interactions between TF and the ribosome-binding site were set to the distance between them in the crystal structure divided by $2^{1/6}$. The set of TF-ribosome interaction sites and their distances in the bound state are listed in Table S1. These were identified by first aligning the partially solved L23 protein X-ray structure in PDB 2VRH, which also contains a bound TF, to the L23 ribosomal protein in the full 50S subunit structure in PDB 2WDJ. Using the newly aligned TF and 50S structure we determined the native contacts between these two components in this bound state (see below).

All non-native interactions, whether intra-protein, inter-protein, or between protein and ribosomal RNA molecules, were treated as short-range and repulsive as defined in the last summation term. As in the original Karanicolas-Brooks model, $\varepsilon_{ij} = 0.000132$ kcal/mol, σ_{ij} for intra-protein non-native interactions were calculated as defined previously,[2] and σ_{ij} for inter-protein and protein-RNA interactions were calculated as $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, where σ_i and σ_j are defined in Table S2. We note that interactions between TF and the nascent chain were treated in a transferable manner; therefore this energy term was not applied to those interactions.

Ribosomal RNA was modelled as described previously[5]. Briefly, nucleotides containing pyrimidines and purines were represented as 3 and 4 interaction sites, respectively, with one interaction located at the phosphate position and having a q=-1*e* charge, another at the centroid of the ribose ring, and one at the centroid of each conjugated ring in the base. In the simulations the ribosome was held rigid. Therefore, the intra and inter-RNA interactions were not calculated and no functional forms for the corresponding energy terms were defined in this force-field. The electrostatic component of RNA

interactions with protein molecules were modelled using Debye-Huckel theory, and excluded volume interactions were modelled as $\varepsilon_{ij}\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$, with $\varepsilon_{ij} = 0.000132$ kcal/mol and σ_{ij} calculated from Table S2.

The coarse-grained coordinates for the 50S ribosomal subunit, TF, protein G, and β-galactosidase were obtained, respectively, from the PDBs 2WDJ, 2VRH, 1GB1, and 1JZ7. PDB 2WDJ is a ribosome structure from the organism *Thermus thermophilus*, rather than *E. coli*. This structure was chosen because it was the highest resolution crystal structure available of a bacterial ribosome with the fewest missing ribosomal protein molecules and residues at the start of this study. A comparative study of bacterial ribosome 50S structures reveals they differ by no more than ≈1.5 Å root-mean-squared-deviation (RMSD). Therefore, we expect the results of this study to hold for the *E. coli* ribosome as well. Residues 1-275 were extracted from PDB 1JZ7, as this range contained the N-terminal domain of β-galactosidase. Before the coarse-graining procedure was applied, the program CHARMM[6] was used to place any residues or atoms missing in the PDB files. Intra-protein native contacts were identified as those residue pairs that have any side-chain heavy atoms within 4.5 Å of each other in their crystal structure; all other residue pairs were classified as non-native.

Parameterizing the Lennard-Jones parameters by reproducing experimentally measured free-energies. Protein G's intra-protein Lennard-Jones well-depths for native interactions were chosen such that they reproduce the experimentally measured native stability, with respect to the denatured state, of -4.7 kcal/mol at 295 K in free solution.[7] This was achieved by uniformly scaling the BT interaction matrix by a multiplicative factor, denoted λ_{BT}, of 1.475. The N-terminal domain of β-galactosidase is comparatively large and has a highly complex topology, which we found prevents equilibrium folding/unfolding from being achieved on the time scales we simulated. Such an extremely long equilibration phase has also been observed in coarse-grained simulations of Green Fluorescence Protein,[8] which is of similar size and topological complexity. Therefore, to set a realistic intra-protein energy scale for this domain we determined a generic λ_{BT} scaling factor by matching the calculated stability of Barnase to its experimentally measured stability,[9] and found a λ_{BT} value of 1.50 was required. We then averaged these two λ_{BT} values (1.475 and 1.50) and used the result it to set the intra-protein Lennard-Jones well-depths for native interactions in β-galactosidase.

Likewise, inter-protein Lennard-Jones well-depths between TF and the nascent chain were set by scaling the BT interaction matrix by a multiplicative factor, $\lambda_{TF\text{-}NC}$, of 0.20. This value was found to reproduce the experimentally measured fraction of TF bound to a reduced carboxymethylated α-lactalbumin[10] (Figure 1B). It is important to note that these interactions are non-Go. To calculate the fraction of TF bound, umbrella sampling simulations were carried out at the experimental temperature of

298 K with 16 different Hamiltonians. The Hamiltonians differed by a harmonic restraint ($V = K_i(d - d_{i,0})^2$) that was applied to the center-of-mass distance, d, between TF and α-lactalbuminm, which restrained them to a target distance of $d_{i,0}$ using a force-constant K_i for Hamiltonian i. The target distances were equally spaced in the range of 0 to 120 Å inclusive, and K_i equalled 0.005 kcal/(mol Å^2) for all Hamiltonians. The fraction of TF bound when $K_i = 0$ was calculated using the histogram free version of the WHAM equations [11].

The native interactions between the tail of TF and its binding site on the ribosome were set by choosing a Lennard-Jones well-depth that reproduced the experimentally measured fraction of TF-bound to a non-translating ribosome at a simulation concentration of 39 μM (Figure 1C). A uniform well-depth was chosen as both the L23 ribosomal protein and some ribosomal RNA bases are in contact with TF in the aligned structure. Unrestrained Langevin dynamics simulations were found to be unsuitable for precisely calculating the fraction of TF-bound in the simulations due to poor sampling associated with slow translation and rotational time-scales around the 50S ribosomal subunit. Rigid-body Monte Carlo Replica Exchange simulations were therefore used to calculate the fraction of TF bound to its ribosomal binding site [12]; this simulation approach has been used to calculate accurately K_D values from simulations of non-specific protein-protein interactions.[13] In these simulations the ribosome and TF molecule were held rigid, and Monte Carlo (MC) moves performed on TF to allow it to bind and unbind to the ribosomal binding site. The MC move set consisted of rigid body translation, with random displacements of up to 0.75 Å at a time, and rigid body rotation around center-of-mass of TF's, with random rotations of up to 4°. The probability of randomly picking a translation or rotation move was 0.5. Cubic periodic boundary conditions were used with a box length of 350 Å in conjunction with an image interaction cut-off of 80 Å. 16 replicas, ranging in temperature from 293 to 650 K, were used. 350 MC moves were made before an exchange between replicas with neighboring temperatures was attempted. A total of 9×10^4 replica exchanges were made, with the first 3×10^4 exchanges discarded to allow for equilibration. The fraction of successful exchanges was between 0.59 and 0.91.

REFERENCES

(1) Best, R. B.; Chen, Y. G.; Hummer, G. *Structure* **2005**, *13*, 1755.
(2) Karanicolas, J.; Brooks, C. L. *Protein Sci* **2002**, *11*, 2351.
(3) Spitzer, J. J.; Poolman, B. *Trends Biochem Sci* **2005**, *30*, 536.
(4) Betancourt, M. R.; Thirumalai, D. *Protein Sci* **1999**, *8*, 361.
(5) O'Brien, E. P.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J Am Chem Soc* **2011**, *133*, 513.
(6) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J Comput Chem* **2009**, *30*, 1545.
(7) De Sancho, D.; Doshi, U.; Munoz, V. *J Am Chem Soc* **2009**, *131*, 2074.
(8) Andrews, B. T.; Gosavi, S.; Finke, J. M.; Onuchic, J. N.; Jennings, P. A. *P Natl Acad Sci USA* **2008**, *105*, 12283.
(9) Clarke, J.; Fersht, A. R. *Biochemistry-Us* **1993**, *32*, 4322.
(10) Maier, R.; Scholz, C.; Schmid, F. X. *J Mol Biol* **2001**, *314*, 1181.
(11) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J Comput Chem* **1992**, *13*, 1011.
(12) Sugita, Y.; Okamoto, Y. *Chem Phys Lett* **1999**, *314*, 141.
(13) Kim, Y. C.; Hummer, G. *Journal of Molecular Biology* **2008**, *375*, 1416.
(14) O'Brien, E. P.; Hsu, S. T. D.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M. *J Am Chem Soc* **2010**, *132*, 16928.
(15) Hoffmann, A.; Bukau, B.; Kramer, G. *Bba-Mol Cell Res* **2010**, *1803*, 650.
(16) Liang, S. T.; Xu, Y. C.; Dennis, P.; Bremer, H. *J Bacteriol* **2000**, *182*, 3037.
(17) Fluitt, A.; Pienaar, E.; Vijoen, H. *Comput Biol Chem* **2007**, *31*, 335.
(18) Merz, F.; Boehringer, D.; Schaffitzel, C.; Preissler, S.; Hoffmann, A.; Maier, T.; Rutkowska, A.; Lozza, J.; Ban, N.; Deuerling, E.; Bukau, B. *Embo J* **2008**, *27*, 1622.
(19) Heinig, M.; Frishman, D. *Nucleic Acids Res* **2004**, *32*, W500.
(20) Humphrey, W.; Dalke, A.; Schulten, K. *J Mol Graphics* **1996**, *14*, 33.

Table S1: Collision diameters between TF and its binding site on the ribosome.

TF residue number	Coarse grained ribosomal site	Corresponding all-atom residue in PDB 2WDJ	Collision Diameter (Å)	TF residue number	Coarse grained ribosomal site	Corresponding all-atom residue in PDB 2WDJ	Collision Diameter (Å)
41	R	A, GUA, 1241	7.8	44	P	A, GUA, 1243	8.0
41	SL	X, LEU, 11	8.0	44	SL	X, LEU, 11	7.7
41	SS	X, SER, 12	7.1	44	SS	X, SER, 12	5.0
41	SE	X, GLU, 13	7.4	44	SE	X, GLU, 13	3.3
42	R	A, GUA, 1241	7.9	44	SK	X, LYS, 14	6.6
42	SP	X, PRO, 9	7.4	44	SY	X, TYR, 16	7.9
42	SL	X, LEU, 11	4.7	45	R	A, URA, 1220	6.6
42	SS	X, SER, 12	4.5	45	P	A, ADE, 1221	7.9
42	SE	X, GLU, 13	5.9	45	R	A, ADE, 1221	7.1
42	SY	X, TYR, 16	8.0	45	R	A, GUA, 1242	7.7
42	SL	X, LEU, 93	7.1	45	SS	X, SER, 12	8.0
42	SI	2, ILE, 40	7.4	45	SE	X, GLU, 13	6.2
43	R	A, GUA, 1241	6.1	46	R	A, ADE, 1221	6.6
43	SP	X, PRO, 9	5.4	48	SE	X, GLU, 13	6.8
43	R	A, GUA, 1241	4.8	49	SE	X, GLU, 13	6.3
43	P	A, GUA, 1243	6.0	50	SE	X, GLU, 13	7.6
43	SV	X, VAL, 10	7.8	50	SY	X, TYR, 16	7.5
43	SL	X, LEU, 11	5.0	50	SA	X, ALA, 17	7.9
43	SS	X, SER, 12	2.6	52	SG	X, GLY, 92	7.1
43	SE	X, GLU, 13	4.0	52	SL	X, LEU, 93	7.8
43	SK	X, LYS, 14	6.6	53	SG	X, GLY, 92	6.4
43	SA	X, ALA, 15	7.0	53	SL	X, LEU, 93	5.7
43	SY	X, TYR, 16	7.8	56	SG	X, GLY, 92	7.0
44	R	A, GUA, 1242	6.6	56	SL	X, LEU, 93	7.4

Table S2: Non-bonded collision diameters for protein-protein and protein-RNA interactions.

Interaction site name	Group name	Collision Diameter (Å)
P	RNA Phosphate	7.1
R	RNA Ribose	7.1
BR	RNA Conjugated ring	7.1
SS	Serine	5.2
SF	Phenylalanine	6.3
ST	Threonine	5.6
SN	Asparagine	5.7
SK	Lysine	6.3
SY	Tyrosine	6.4
SE	Glutamic acid	6.0
SV	Valine	5.9
SQ	Glutamine	6.1
SM	Methionine	6.1
SC	Cysteine	5.4
SL	Leucine	6.1
SA	Alanine	5.1
SW	Tryptophan	6.8
SP	Proline	5.5
SH	Histidine	6.1
SD	Aspartic Acid	5.6
SI	Isoluecine	6.1
SR	Arginine	6.6
SG	Glycine	4.5