Supporting Information

Yassin et al. 10.1073/pnas.1017425108

SI Results and Discussion

Structural Similarity Between the $IF2_{mt}$ Insertion and *Thermus thermophilus* IF1. The additional residues in $IF2_{mt}$, as compared to *Methanobacterium thermoautotrophicum* IF2, could perform the biochemical function of bacterial IF1 by adopting a similar structure. The factors in opposition to this possibility are the lack of sequence similarity, a different predicted secondary structure using Jpred (1) and the shorter length of the insert in comparison to the structurally characterized 72 residue *T. thermophilus* IF1 (2). Nevertheless, the insert region could still mimic smaller key structural parts of IF1 critical to performing its biochemical function. To explore this possibility, we generated vacuum trajectories of *T. thermophilus* IF1 being forced out of its ribosomal binding site in 8 different directions using Miscellaneous Mean Field Potential (MMFP) center-of-mass restraints in the CHARMM program (3). In all these trajectories, the final region of IF1 that lost contact with the ribosomal binding site was at the N-terminal end, specifically a 4-residue stretch (Lys-Glu-Lys-Asp). Although this observation could be due to topology rather than to the importance of this sequence for the interaction of IF1 with the ribosome, a similar sequence (Lys-Asp-Arg-Glu) is present in the 49 residue sequence, which allowed us to postulate a possible gapped sequence alignment of the bovine mitochondrial insert with *T. thermophilus* IF1 (Fig. S5*A*). Using this gapped sequence alignment, we constructed a homology model for the 49-residue sequence that mimics the structure of the first 50 residues of *T. thermophilus* IF1 [Protein Data Bank (PDB) ID 1HR0, chain W] using MODELLER (4). The N- and C-terminal ends of this shortened IF1-like structure were quite far from one another (17 Å), which is inconsistent with the proximity of the residues on the N- and C-terminal ends of the insert in the composite $IF2_{mt}$ model. When this 49-residue homology model was rigidly fit into the same orientation with respect to the ribosome as IF1 (Fig. S5*B*), its N-terminal region was outside the observed cryo-EM map, with its N- and C-terminal ends substantially separated from their connections to the rest of the flexibly fit IF2 model (Fig. S5*C*). This necessitated refolding of larger sections of IF2 by assigning more residues to the linker regions to get a continuous model. Even if the IF1-like orientation of this insert model was not preserved, flexible fitting into its binding site density using the SCX functionality in the Yup program (5) yielded a model with a dearth of secondary structure elements (Fig. S5*D*). This model was therefore considered less likely than the reported model with a compact α-helical rich insert. These observed geometric restrictions imposed by the necessity to have a continuous structure that fits into the experimentally observed cryo-EM map suggest that the $IF2_{mt}$ insert adopts a different internal structure and/or interacts with the ribosome in a slightly different orientation than that observed for *T. thermophilus* IF1.

1. Cuff JA, Barton GJ (1999) Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* 34:508–519.
2. Carter AP, et al. (2001) Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* 291:498–501.
3. Brooks BR, et al. (2009) CHARMM: The biomolecular simulation program. *J Comput Chem* 30:1545–1614.
4. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815.
5. Tan RKZ, Petrov AS, Harvey SC (2006) YUP: A molecular simulation program for coarse-grained and multiscaled models. *J Chem Theory Comput* 2:529–540.

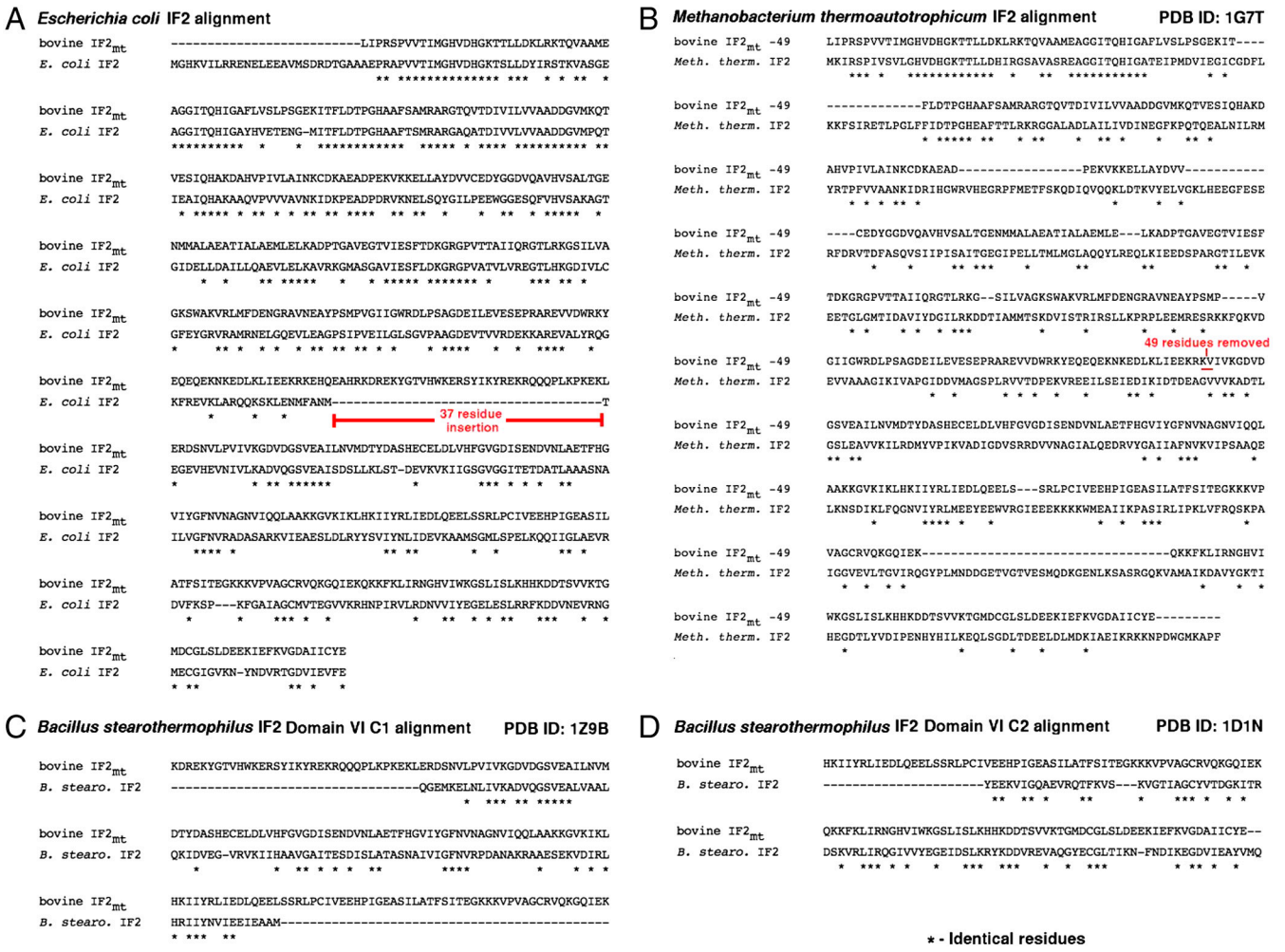

Fig. S1. Sequence alignment of $IF2_{mt}$ with relevant homologous sequences. (*A*) Alignment with *Escherichia coli* IF2, showing 37 amino-acid residue insertion; (*B*) alignment of $IF2_{mt}$ (with 49 amino-acid residues removed between lysine and valine where indicated in red) with *M. thermoautotrophicum* IF2; (*C*) alignment of VI C1 domain regions of $IF2_{mt}$ and *Bacillus stearothermophilus* IF2; and (*D*) alignment of VI C2 domain regions of $IF2_{mt}$ and *B. stearothermophilus* IF2. The PDB IDs of existing experimental structures corresponding to sequence regions are shown on the right.

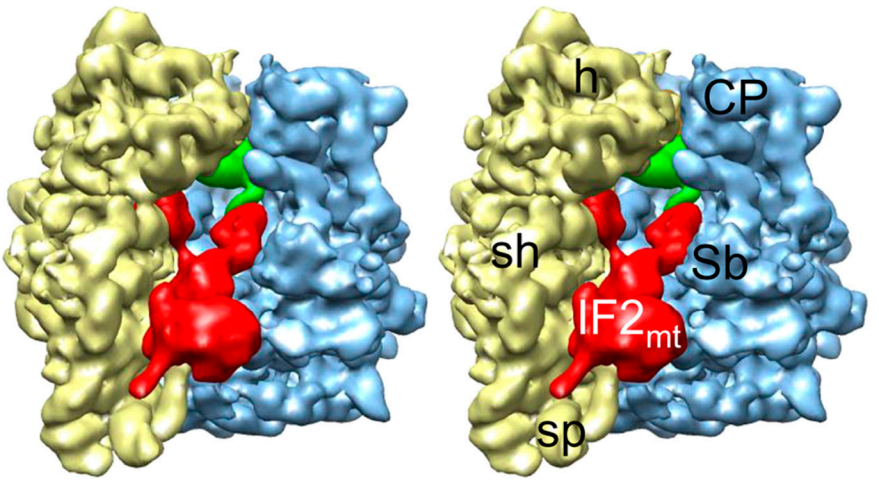

Fig. S2. Stereo-view representation of the segmented cryo-EM structure the 70S • mRNA • fMet-RNA_i^{Met} • $IF2_{mt}$ • GDPNP complex. The cryo-EM map of the *E. coli* 70S ribosome (30S subunit, yellow; 50S subunit, blue) in complex with $IF2_{mt}$ (red), initiator tRNA (green) at the P/I position. Landmarks of the 30S subunit: h, head; sh, shoulder; and b, body. Landmarks of the 50S subunit: CP, central protuberance; and Sb, L7/L12 stalk base.

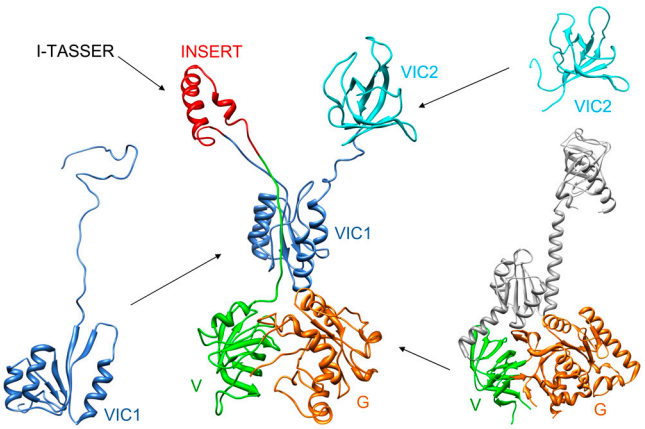

Fig. S3. Schematic to show usage of various template structures to construct a composite $IF2_{mt}$ model that best explains the cryo-EM density. Three separate structures were used. X-ray structure of the *M. thermoautotrophicum* IF2 (PDB entry 1G7T, *Lower Right*), NMR structure of the VI C1 domain of *B. stearothermophilus* IF2 (PDB entry 1Z9B, *Lower Left*), and NMR structure of the VI C2 domain of *B. stearothermophilus* IF2 (PDB entry 1D1N, *Upper Right*).

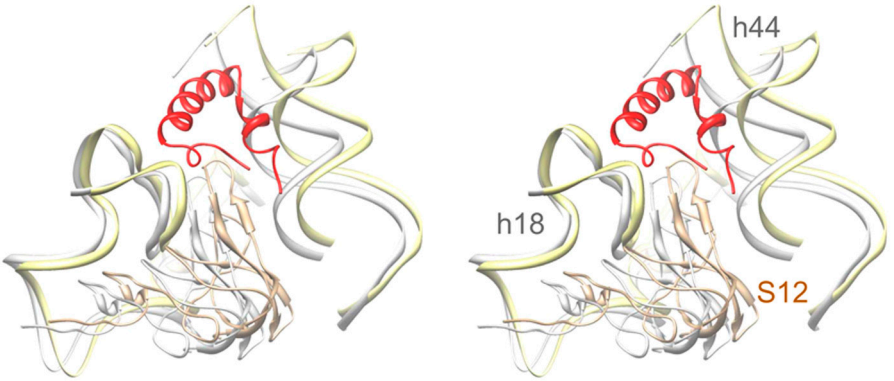

Fig. S4. Stereo-view representation of the conformational changes that occur within the binding pocket upon interaction of the $IF2_{mt}$ insertion domain with the ribosome. Positions of ribosomal components as found in the presence of IF1 (light gray), or in the presence of the $IF2_{mt}$ insert (yellow), are shown in ribbons.

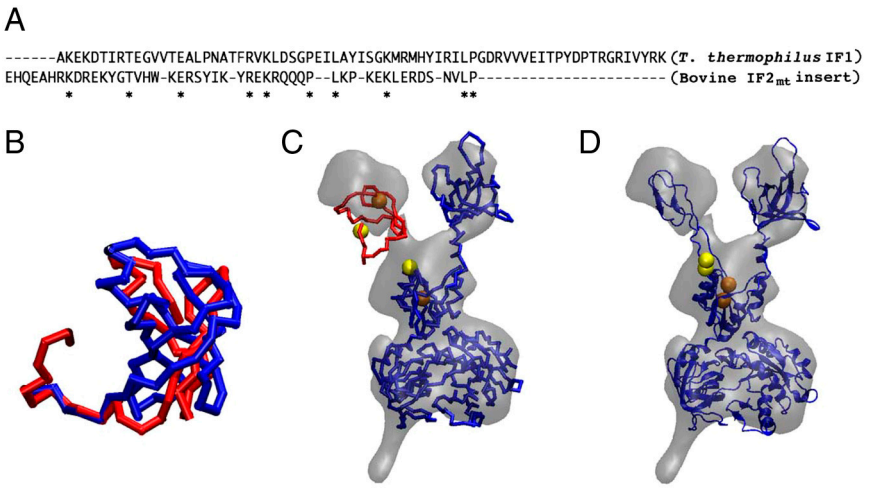

Fig. S5. Exploration of the possibility of structural similarity between the $IF2_{mt}$ insertion and *T. thermophilus* IF1. (*A*) Manually generated gapped sequence alignment of additional 49 residues in $IF2_{mt}$ (compared to *M. thermoautotrophicum* IF2) and *T. thermophilus* IF1, (*B*) homology model of the 49 residue region (red) aligned to *T. thermophilus* IF1 (blue) shown in trace form, (*C*) placement of aligned 49 residue $IF2_{mt}$ model (red) into cryo-EM map along with the flexibly fit $IF2_{mt}$ homology model (blue) in trace form showing that the N-terminal region of insert lies outside of the observed density, and (*D*) complete model of $IF2_{mt}$ (blue) with the IF1 orientation of the insertion not preserved showing dearth of secondary structure in the insert region. In (*C*) and (*D*), residues 468 and 469 are shown in yellow spheres and residues 518 and 519 are shown in orange spheres to indicate points of covalent linkage between the $IF2_{mt}$ insert region and rest of $IF2_{mt}$.

Table S1. Comparison of interactions of IF1 (2) and the 37-aa insertion domain of $IF2_{mt}$ with the 16S rRNA helices h18 and h44, and protein S12

IF1:	Glu3	S12	Thr42
	Asp61		
	Tyr60		Val40
IF1:	Lys2	**h18**:	G530
	Lys39		
	Arg66		
	Gly38		C519
	Tyr35		
	Met42		G530
IF1:	Arg41	**h44**:	A1492
	Arg46		A1493
	Asn19		G1491
			A1492
			G1494
	Ala20		A1493
Insert:	Lys508	**S12**:	Leu52
	Leu509*		Thr42
			Val43
			Thr44
	Arg511*		Arg41
Insert:	Lys497	**h18**:	C518
			G530
	Gln500		G517
			G530
			U531
	Gln501		C519
	Lys504		G517
			C518
			C519
	Pro505		C519
	Lys506		C519
Insert:	Lys476	**h44**:	C1411
	Lys480		A1492
	Gly482		A1492
	Thr483		A1493
	Val484		A1492
			A1493
	Lys487		A1492
	Arg489		C1397
	Glu507		A1492

Amino-acid and nucleotide residues of the ribosomal components involved in interaction with both factors are underlined.
*Residues directly following the insert.

Table S2. Comparison of different composite homology models with respect to their similarity with the $IF2_{mt}$ sequence and the corresponding cryo-EM density

Model number	Template IF2 (PDB ID)	% identity (% similarity)	Alignment score	Cross correlation coefficient (ccc), after flexible fits using	
				Yup (1)	MDFF (2)
1	*M. thermoautotrophicum*(1G7T)	23.4 (37.5)	433	0.70	—
2	Domains IV-VIC1 from *M. thermoautotrophicum*(1G7T) Domain VIC2 from *B. stearothermophilus* (1D1N)	28.2 (43.9)	581	0.67	—
3	Domains IV-V from *M. thermoautotrophicum* (1G7T) Domain VIC1 from *B. stearothermophilus* (1Z9B) Domain VIC2 from *B. stearothermophilus*(1D1N)	28.9 (45.2)	641.5	0.72	0.85

1 Tan RKZ, Petrov AS, Harvey SC (2006) YUP: A molecular simulation program for coarse-grained and multiscaled models. *J Chem Theory Comput* 2:529–540.
2 Trabuco LG, Villa E, Schreiner E, Harrison CB, Schulten K (2009) Molecular dynamics flexible fitting: A practical guide to combine cryoelectron microscopy and X-ray crystallography. *Methods* 49:174–180.