HHS Public Access

Author manuscript

Comput Biol Med. Author manuscript; available in PMC 2025 June 17.

Published in final edited form as:

Comput Biol Med. 2025 June ; 191: 110188. doi:10.1016/j.compbiomed.2025.110188.

Evolutionary insights into elongation factor G using AlphaFold and ancestral analysis

Shawonur Rahaman[a], **Jacob H. Steele**[b], **Yi Zeng**[a], **Shoujun Xu**[a], **Yuhong Wang**[a,*]

[a]Department of Chemistry, University of Houston, Houston, TX, 77204, USA

[b]Department of Biology and Biochemistry, University of Houston, Houston, TX, 77204, USA

Abstract

Elongation factor G (EF-G) is crucial for ribosomal translocation, a fundamental step in protein synthesis. Despite its indispensable role, the conformational dynamics and evolution of EF-G remain elusive. By integrating AlphaFold structural predictions with multiple sequence alignment (MSA)-based sequence analysis, we explored the conformational landscape, sequence-specific patterns, and evolutionary divergence of EF-G. We identified five high-confidence structural states of wild type (WT) EF-G, revealing broader conformational diversity than previously captured by experimental data. Phylogenetic analysis and MSA-embedded sequence patterns demonstrated that single-point mutations in the switch I loop modulate equilibrium between the two dominant conformational states, con1 and con2, which exhibit distinct functional specializations. Reconstructions of two ancestral EF-Gs revealed minimal GTPase activity and reduced translocase function in both forms, suggesting that robust translocase activity emerged after the divergence of con1 and con2. However, ancestral EF-Gs retained the fidelity of three-nucleotide translocation, underscoring the early evolutionary conservation of accurate mRNA movement. These findings establish a framework for understanding how conformational flexibility shapes EF-G function and specialization. Moreover, our computational pipeline can be extended to other translational GTPases, providing broader insights into the evolution of the translational machinery. This study highlights the power of AlphaFold-assisted structural analysis in revealing the mechanistic and evolutionary relationships involved in protein translation.

1. Introduction

Elongation factor G (EF-G) plays a critical role in ribosomal protein synthesis by facilitating the rapid and accurate translocation of mRNA [1]. This process ensures proper codon

*Corresponding author. ywang60@uh.edu (Y. Wang).

CRediT authorship contribution statement

Shawonur Rahaman: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Data curation. **Jacob H. Steele:** Writing – review & editing, Validation, Funding acquisition, Data curation. **Yi Zeng:** Writing – review & editing, Validation, Investigation, Formal analysis, Data curation. **Shoujun Xu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Yuhong Wang:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

positioning for the incoming aminoacyl tRNA, preventing frameshifting errors that can have cumulative effects. Compared with single amino acid misincorporation, translocation errors occur much less frequently because of EF-G's essential role in ensuring fidelity [2]. EF-G undergoes large conformational changes that coincide with major ribosome rearrangements, coupling energy release to translocation [3–6]. Structurally, EF-G consists of five domains with hinge-like flexibility between domains I-II and III-V, which allows conformational changes. In ribosome-bound structures, EF-G is typically extended, except for one case where domain IV folds back 180° [7]. In solution, EF-G samples a broader conformational range (Table S1, Fig. S1). Prior studies suggest that domain IV insertion into the ribosomal A-site requires pre-translocation conformational changes, but the exact sequence of these structural transitions remains unclear [8]. The AlphaFold prediction tool now offers a more comprehensive analysis of EF-G's conformational landscape than traditional methods do, providing a novel approach to reveal the conformation-function relationship of EF-G [9–12].

In this study, we applied multi-sequence-alignment (MSA) clustering to predict structural ensembles of WT and its T49E/T49V mutants of *E. coli* EF-G [13]. We identified two major conformations, con1 and con2, which correspond to previous sequence-defined EF-G variants, EF-G1 and EF-G2 [14,15]. Our analysis revealed that single-point mutations shift the balance between these conformations, offering insights into their functions. However, more evidence is likely needed to confirm the common ancestral form of con1 and con2. Experimentally, we expressed and functionally characterized two ancestral EF-Gs inferred using maximum likelihood that correspond to con1 and con2. Both lacked GTPase activity but retained the ability to catalyze 3-nucleotide translocation, suggesting that con1 and con2 diverged before the evolution of GTPase activity in modern con1 EF-G. These findings provide an evolutionary framework linking sequence divergence, structural adaptation, and function. Our study demonstrates the power of computational modeling, in conjunction with innovative experimental assays, in uncovering previously unknown correlation between EF-G conformations and their evolutionary trajectory and shows that this method can be extended to other translational factors and biological systems.

2. Materials and methods

All the computational work was streamlined including a Jupyter Notebook available online at https://github.com/ywang6000/EF-G-conformation-fold.

2.1. AFClustering and ColabFold prediction

MSAs were generated using ColabFold with the query sequence of WT *E. coli* EF-G or the T49E/T49V mutated sequences [10,16]. For clarity, the method description and the Jupyter Notebook script use the WT sequence as an example. The MSA was clustered using a modified version of the AFCluster method (implemented in ClusterMSA_notebook.py). The optimized epsilon (eps) value for clustering was determined to be 24 for EF-G with 706 residues [13]. Approximately 50 % of the DBSCAN [17] clustered ".a3m" files were selected for folding with ColabFold.

2.2. Structure clustering

The predicted structures (in PDB format) exceeding the pLDDT threshold were loaded into PyMOL, aligned to a reference structure (EX_175), and saved with updated coordinates. The RMSD values were calculated against two reference structures, EX_175 and EX_276, which were identified through visual inspection of the highest pLDDT structures. The RMSD values were plotted, as shown in Fig. 2, and clustering was determined by visual inspection of the RMSD plots and evaluating the RMSD values (RMSD <10). The corresponding ".a3m" files that generated the clustered PDB structures were pooled and tagged for downstream analysis. For example, WT_list1.txt and WT_list2.txt contain the con1 and con2 clustered PDB IDs. The combined sequences for these clusters, excluding the query sequence, are saved as WT_list1.a3m and WT_list2.a3m. An example ID is "WT-list1-UniRef100-A0A0M1N2R8", where "WT" and "list1" are tags appended to the original UniRef ID to allow tracking.

2.3. Phylogenetic tree construction

The UniRef100 IDs were used to fetch protein IDs from UniProt.org. After the protein names were shortened and the taxonomy information was removed, a new text file was generated with IDs formatted as: "EFG-WT-list1-A0A0M1N2R8," and similar variations. The "EFG" portion of the ID was used to create the circumference rings, whereas the "list1" tag was used to color the leaves. The "WT" tag was utilized for generating the colored leaves when all MSAs were combined.

The MSAs from ColabFold were prealigned. For cases where alignment was required, the combined sequences were aligned using MUS-CLE and analyzed in IQ-TREE 2 to construct both phylogenetic and ancestral trees with 1000 bootstrap replicates [18,19]. The resulting phylogenetic trees were visualized in iTOL, using the necessary data files to enhance tree annotations and visualization [20].

2.4. Protein expression and purification

The amino acid sequences of Node660 and Node327 were derived from the ancestral tree of WT EF-G. Plasmids carrying the codon-optimized genes on the pET-20b(+) vector for expression in BL21 (DE3)pLysS competent cells were ordered from GenScript. Mid-log-phase cells (OD600 ~0.6–0.8) were induced with 0.1 mM IPTG and harvested after 3 h of growth at 37 °C. The cell pellets were collected by centrifugation at 4000×*g* for 10 min at 4 °C. The pellets were resuspended in lysis buffer (50 mM Tris-HCl, 300 mM NaCl, pH 8.0) containing a small amount of lysozyme crystals and incubated on ice for 20 min. The cells were lysed using two rounds of freeze-thaw cycles using liquid nitrogen, followed by 2 min of sonication at 10 % power (e.g., 10 s on, 20 s off). The lysate was clarified by centrifugation at 15,000×*g* for 30 min at 4 °C. The cleared lysate was loaded onto a HisTrap™ FF 5 mL column pre-equilibrated with binding buffer (50 mM Tris-HCl, 300 mM NaCl, pH 8.0). The column was washed with 10 column volumes of wash buffer (50 mM Tris-HCl, 300 mM NaCl, pH 8.0) to remove nonspecific proteins. The target proteins were eluted at approximately 200 mM imidazole using a 5–500 mM gradient on a GE ÄKTA Explorer 10 FPLC system. Eluted fractions were analyzed by SDS-PAGE to confirm purity

and subsequently buffer-exchanged into TAM10 buffer (20 mM Tris-HCl (pH 7.5), 30 mM NH_4Cl, 70 mM KCl, 10 mM $MgCl_2$, 0.5 mM EDTA, and 7 mM β-mercaptoethanol).

2.5. Preparation of the MF-pre ribosome complex

Preparation of the ribosome, the 70S ribosome from *E. coli* MRE600, and ribosome factors (His-tagged IF1, IF2, IF3, EF-G, EF-Tu, methionyl-tRNA and phenylalanyl-tRNA synthetases) has been described in detail previously, [21], as has the preparation of fMet-$tRNA^{fMet}$. A complete description is provided in the Supplementary Data.

Three mixtures were prepared: the ribosome mixture, Tu0G mixture, and Phe mixture. The ribosome mixture contained 1 μM ribosomes, 1.5 μM each initiation factors (IF1, IF2, and IF3), 2 μM mRNA, 4 μM charged fMet-$tRNA^{f_{Met}}$, and 4 mM GTP. The Tu0G mixture consisted of 4 μM EF-Tu, 4 mM GTP, 4 mM 2-phosphoenolpyruvate (PEP), and 0.02 mg/mL pyruvate kinase. The Phe mixture included 100 mM Tris-HCl (pH 7.8), 20 mM $MgCl_2$, 1 mM EDTA, 4 mM ATP, 7 mM β-mercaptoethanol (BME), 2 μM $tRNA^{ph^e}$ synthetase (PheRS), 50 A260 units/mL total tRNA, and 0.25 mM phenylalanine. Each mixture was preincubated separately at 37 °C for 15 min and then combined in a 1:2:2 vol ratio at room temperature for 2 min. The resulting MF-Pre ribosome complex was layered onto a 1.1 M sucrose cushion and purified by centrifugation at 200,000–400,000×*g* for 3 h at 4 °C using a Hitachi CS150FNX ultra-centrifuge with an S140AT rotor.

2.6. Biomolecular assays for EF-G characterization and function

We carried out five assays to experimentally characterize the EF-Gs identified by computation. These assays were: GTP binding assay with mant-GTP, GTP hydrolysis assay via thin layer chromatography (TLC), Poly(Phe) assay, minimum inhibitory concentration (MIC) assay, and super-resolution force spectroscopy (SURFS). The technical details of all the assays are provided in the Supplementary Data.

3. Results and discussion

3.1. Five high-confidence folded structures were predicted for WT EF-G

AlphaFold predictions suggest that evolutionary couplings within a MSA subset are sufficient to define a specific protein structure. This means that the key residue-residue interactions required to predict a specific form are present within a subset of the MSA. However, because the entire MSA contains all the interactions dictating all possible states, only one dominant state is typically generated while the other subpopulations are masked. This problem can be solved by a recent method that clusters the MSA using the DBSCAN algorithm [13,22]. Adopting this method, at least five distinct confidence folded structures were generated for EF-G. As illustrated in Fig. 1, the WT EF-G sequence from *E. coli* yielded an MSA of 25,939 sequences, resulting in one folded structure. Clustering with the DBSCAN algorithm generated 393 clusters and 50 % of these clusters were randomly selected for folding with AlphaFold. Among the 43 structures with pLDDT values greater than 60, some aligned well with free EF-G X-ray crystal structures (Fig. 1), indicating biological relevant predictions. The average pLDDT score reflects both low-confidence

folding regions and intrinsically flexible regions. In EF-G, the P-loop, switch I, and switch II regions are inherently flexible, which is essential for GTPase function. While a pLDDT score of 70 is the standard threshold for high-confidence structural predictions, [9], we also considered structures with pLDDT scores below this threshold to account for intrinsic flexibility, which lowered the overall folding scores. As shown in Fig. S1, structures with pLDDT <60 failed to cluster consistently with the reference structures, supporting 60 as a reasonable threshold. Among the structures with pLDDT values between 60 and 70, only two clustered with con1, while the rest did not form distinct clusters. In Fig. S1 inset 1, EX_367 (pLDDT = 67.6) clustered with con1, showing no major discrepancies in pLDDT compared with EX_360 (pLDDT = 69.6), a high-confidence structure. In Fig. S1 inset 2, EX_298 (pLDDT = 57.5) clustered near EX_250 (pLDDT = 76.0), yet its residue-wise pLDDT plot revealed significant discrepancies in the structurally stable regions of EX_250. This suggests that EX_298 contains poorly predicted segments, further supporting 60 as a reasonable pLDDT threshold.

After grouping the similar structures, four distinct structures exhibited variations in the angles between domains I-II and IV. One structure, however, showed an altered arrangement between domains IV and V, with no similar X-ray structure (Fig. S2). Two of these structures (con1 and con2) result from multiple clusters and are embedded within larger MSA sizes, whereas the other three (EX_250, con3 and EX_144) have smaller MSA sizes. Additionally, a large cluster (EX_000) with over 4600 sequences dominated the folding process to generate con1. This large cluster explains the single conformation produced by AlphaFold or ColabFold prior to MSA clustering (Fig. S3). However, analysis of the smaller clusters allowed us to identify the diverse conformations that are important for EF-G function [13,23,24]. However, the MSA clusters that yielded low-confidence structures were not considered.

3.2. Single point mutations at switch I loop shift the conformational landscape

Our previous biochemical and translocation studies revealed that the T49E and T49V mutations induced ribosomal “−1” frameshifting and compromised the Poly(Phe) synthesis activity [25,26]. To theoretically understand these mutant effects, clustering analysis of AlphaFold predictions was conducted. As expected, the large dominant cluster (EX_000) persisted in the MSA of both mutants, with more than 95 % overlap with that of the WT (Fig. S3). This finding suggests that the core structure “con1” remains largely conserved despite mutations. However, some new states were identified from the smaller MSA clusters. Using WT-con1 and WT-con2 as standards for screening, we found that both EF-G mutants adopted con1 and con2 conformations (Fig. 2), along with other less populated conformations. Interestingly, a new structure appeared in both mutants that exhibited an intermediate conformation between con1 and con2 (cyan oval). This intermediate state may play a role in the altered function of these mutants. Furthermore, while T49E shows a conformation similar to that of con3, the T49V mutant lacks con3 or EX_250. Lacking the conformations on the two extremes of those in WT may suggest that T49V mutation restricts the flexibility of EF-G.

The phylogenetic trees from these MSAs provided further insights on mutagenesis effects (Fig. 3A). The red and green leaves indicate con1 and con2, respectively. Compared to WT and T49V, T49E shows a much higher preference for con1, with only approximately 12 % con2, whereas the WT and T49V contain 26 % and 50 % con2, respectively. The strong preference for one conformation in T49E may imply diminished flexibility. Therefore, both mutants seem to reduce protein flexibility, albeit through different mechanisms. The T49 residue is located on the switch I loop of the GTP binding pocket, and flipping of this loop is critical for the release of Pi after GTP hydrolysis, as shown by cryo-EM structural studies [3,27]. Therefore, altering the flexibility around this region can have detrimental effects [25,26]. However, these mutations seem to affect the global structural flexibility rather than the local folding of the GTP binding pocket, as no significant alterations in conformation were observed in that region in the different conformations.

These data suggest that AlphaFold effectively captures the global rather than the local impact of single mutations through MSA. Notably, when a single mutation was introduced into the query sequence while using the WT-generated MSA, the resulting cluster-folding structures resembled those of the WT, not the mutant. This indicates that the mutagenesis effects are driven by the cooperativity of all residues via sequence alignment; folding alphabets in evolution, deciphered by AlphaFold, enables illustration of the long-range effects of single-point mutations.

These findings suggest that the functional consequences of T49E and T49V mutants stem from their impact on EF-G conformational flexibility. These changes probably interrupt the allosteric effect between the GTP binding pocket in domain I and the tRNA-EF-G interaction points in domain IV. Since ribosome translocation requires coordinated conformational changes in both the ribosome and EF-G, these disruptions may explain the observed increase in frameshifting and reduced translocation functionality, as we reported previously. Additionally, T49V showed enhanced ribosome co-sedimentation, suggesting that it stabilizes a ribosome-bound state, thus inhibiting next factor binding [25]. Notably, a functionally analogous regulatory mechanism exists in the human EF-G counterpart, where phosphorylation at the equivalent residue inhibits translocation [28]. This finding suggests that perturbations at this position may broadly influence EF-G function by modulating its interaction with the ribosome and the GTP hydrolysis cycle.

3.3. Con1 and con2 correlate with different EF-G functions

The proteins decoded in the MSAs, which are also integral components of the Last Universal Common Ancestor (LUCA), offer critical information about EF-G's evolutionary linkage to other proteins, not only in terms of sequence homology but also in terms of co-evolution of structure and function [29,30]. Approximately half of the MSA IDs are available in the Universal Protein Resource (UniProt.org). The decoded proteins include EF-G, EF2 (archaea), EF-G1, EF-G2, Ribosome-Releasing Factor 2 mitochondrial (RRF2mt), tetracycline-resistant or ribosome protection proteins (Tetra), GTP binding proteins (GTPB), P-loop containing proteins (Ploop), and Tr-type G domain-containing proteins (TrG, translational GTPase superfamily). For clarity, the last three are grouped into an "other" category, as they are nonspecific names [31,32]. A distinct protein distribution pattern

aligned with conformation: EF2, EF-G1, RRF2mt, and “other” are confined to con1; EF-G2 predominantly adopts con2; notably, EF-G and Tetra adopt both con1 and con2 (Fig. 3B, and Fig. S4B). EF2 and EF-G1 are the core translocases in archaea and mitochondria, suggesting that strong translocase function is inherently associated with con 1. This hypothesis is further supported by the observation that *E. coli*, *Saccharomyces cerevisiae*, and *Homo sapiens*, each containing a single EF-G or EF2 form, are also predicted to adopt con1 (Fig. S5A). On the other hand, *Bacteroides thetaiotaomicron* and *Thermus thermophilus* possess two EF-G forms: the experimentally confirmed translocases are predicted in con1; and the experimentally confirmed weak translocases are predicted in con2 (Fig. S5B) [15,33]. Consistent with this conclusion, EF-G2 predominantly adopts con2, suggesting its primary function is not translocase activity. However, adopting con1 does not necessarily imply strong translocase activity. For instance, RRF2mt, which is exclusively involved in ribosome recycling without GTP hydrolysis, is also found to adopt con1 [34,35]. EF-G adopts both con1 and con2, suggesting that some bacteria possess two distinct forms of EF-G. However, it remains experimentally unclear whether any species possess EF-G solely in con2 without a corresponding con1 form. Tetra is similar to EF-G because they exhibit an intertwining pattern in the phylogenetic tree, suggesting a gene duplication origin [36,37].

3.4. Protein association to con1 and con2 MSA clusters

Fig. S6 complements Fig. 3B by showing protein associations with specific MSA clusters. To ensure single count for each protein, MSAs from WT, T49E and T49V were sorted without overlap, such that WT-con1 included sequences that were exclusively present in the WT-con1 MSA, whereas WT/T49E-con2 included sequences that were exclusively present in both WT-con2 and T49E-con2, and so on (Fig. S4A). WT-con1 and WT-con2 predominantly contribute to EF2 and EF-G, respectively, with T49E-con1 and T49E-con2 following a similar pattern, while T49V-con1 and T49V-con2 are mainly associated with EF-G. The overlapping sequences across the three MSAs show contributions consistent with unique sequences. Meanwhile, T49E-con1 predominates over T49E-con2, whereas T49V exhibits the opposite trend, indicating that the T49E mutation favors con1, while T49V favors con2. Consequently, T49E is incompatible with EF-G2, and T49V is incompatible with EF2, aligning with their respective tendencies to favor con1 and con2. LOGO analysis revealed that valine predominated at position 49 in con2 in the WT and both mutants (Fig. S7).

3.5. Con1 and con2 correlate with duplicated gene analysis

Sequence logo analysis revealed the striking similarities of con1 and con2 to EF-G1 and EF-G2 from a previous computational study (Fig. 4) [14]. This prior study categorized EF-G into EF-G1 and EF-G2 based on sequence comparisons, whereas we classified EF-G on the basis of conformation. The G1-G5 motifs and the G’ domain negative patches were compared between con1 and con2 for two types of aligned sequences: all the sequences in WT (WT_all) of Fig. 3A and only the EF-G sequences (grey-colored band) in the WT of Fig. 3A. The sequence logo comparisons showed a high degree of similarity between WT_all and WT EF-G datasets and were nearly identical to the previous computational work. For example, the first conserved residue for G1 motif is “A” for con1 and “G” for con2. The G2 motif is not conserved in con2, except for a single conserved “S” at

the 4th position, whereas the G3, G4, and G5 motifs are similar in both con1 and con2. These patterns align with those found for EF-G1 and EF-G2 in computational work [14]. Furthermore, additional type I and type II conserved residues and motifs outside these regions have been identified [14], and our sequence logos match most of them (Fig. S8).

3.6. Evolutionary divergence of con1 and con2

To investigate evolutionary divergence suggested by the distinct con1 and con2 branches, we reconstructed ancestral phylogenies and analyzed the internal nodes within a branch distance of 1.8 from the midpoint-rooted tree (Fig. 5A). Using maximum likelihood estimation in IQ-TREE2, we inferred the sequences and assessed their variation across 706 residues via Jensen-Shannon divergence (JSD) [38,39]. Clustering by the first off-diagonal (upper) values (threshold >0.4) identified 26 sequence groups, which are color-coded on the tree (Fig. 5B). Con2 showed greater sequence divergence. Among 461 internal nodes, 250 belong to con1, forming 8 clusters, whereas 211 belong to con2, forming 19 clusters, indicating greater structural variability in con2. When considering only EF-G and EF2, the con1 nodes exhibited even stronger sequence conservation, with distinct con1 clusters serving as precursors to non-EF-G proteins. Notably, the EF2 proteins formed a single, highly conserved cluster, indicating strong functional constraints on sequence variation. In contrast, most con2 clusters were precursors to EF-G, showing the greater divergence in EF-G sequence over evolutionary time.

Given the strong translocase activity of con1 in modern proteins, its higher internal node correlation raises the question of whether it represents the ancestral EF-G form. This suggests that con1 may have been the dominant structural framework before functional specialization, whereas con2 followed a more dynamic trajectory to evolve into multiple functional variants. To determine the conformation of the common ancestor of con1 and con2, we predicted 11 internal nodes near the midpoint root, assuming that the lower-level nodes inherit their ancestral conformations. Fig. 5C shows the predicted conformations, all of which exhibited high confidence (pLDDT >80). The individual con1 and con2 branches consistently show bootstrap values of 100, confirming their grouping. The merged nodes have values of 78, 84, and 100 for the WT, T49E, and T49V phylogenies, respectively, supporting a moderate to strong evolutionary relationship (Fig. S9). However, the uncertainty in rooting prevents a precise conclusion on ancestral conformation. When all three MSAs were combined, the bootstrap values at the convergence of con1 and con2 remained uniformly high (>92) (Fig. S10).

3.7. Two ancestral EF-Gs contain minimal GTPase and low translocase activity

The ancestral sequences relevant to tetracycline resistant proteins were pruned from the WT phylogenetic tree (Fig. 6A). Node660 and Node681 contain con1 tetracycline resistance protein, EF-G and GTP binding proteins, while Node323 contains con2 tetracycline resistance protein and EF-G. Logo sequence analysis of all the sequences derived from the same WT tree in Fig. 3A revealed that sequences of Node660 and Node327 closely resemble those of con1 and con2 of EF-G, respectively (Fig. 6B). However, Node660 exhibits intermediate features between con1 and con2 in motif 2, suggesting a transitional state for evolutionary adaptation or functional versatility. Furthermore, both nodes display

nuanced variations in the G5 motif, which could signify subtle differences in their functional roles or binding affinities.

To experimentally evaluate these ancestors, the protein sequences of Node660 and Node327 (a subbranch of Node323) were converted and codon-optimized for *E. coli* expression, yielding purified proteins (Fig. S11). The inferred sequences of Node660 and Node327 have average confidences of 91 % and 93 %, respectively. Fig. S12 presents the probability values for each residue and highlights residues with <50 % probability on the predicted structures. Most low-confidence residues are located in secondary structural motifs, where residue-specific conservation is less critical than preservation of the overall motif structure.

Four functional assays were performed: GTP hydrolysis, GTP binding, poly(Phe) synthesis, and tetracycline minimum inhibitory concentration (MIC) assays. The TLC separation of fluorescent-labeled mant-GTP and mant-GDP following GTPase activity is presented in Fig. 6C, with the GTPase activity results in Fig. 6D. GTP hydrolysis by WT EF-G increased from 1.9 % in the absence of the ribosome to 64 % in the presence of the ribosome. In contrast, the yields for Node327 and Node660 were not significantly different between without and with ribosome (Fig. 6D, top), ranging from 12 % to 16 % in the absence of the ribosome and remained constant at 5 %, in the presence of the ribosome, respectively. These findings indicate that ancestral nodes exhibited low GTPase activity and were not stimulated by ribosome binding, suggesting an evolutionary adaptation of GTPase activity over time. The GTP binding affinities of these proteins were somewhat similar, ranging from 1 to 12 μM (Fig. 6D, bottom). Node327 exhibited the highest binding affinity. WT EF-G displayed a binding affinity of 4.22 ± 1.72 μM, which aligns with reported literature values [40]. Additionally, in vitro poly(U)-based poly-phenylalanine synthesis (poly(Phe)) assays revealed a weaker functionality of Node327 than WT EF-G (Fig. 6E). Node660's translocase activity was almost the same as no EF-G. This reduced activity is likely attributable to their ancestral status and their divergency from the modern ribosome. Node327 and Node660 represent earlier evolutionary intermediates, and their reduced activity suggests that translocase activity had not yet fully evolved. Since these nodes are ancestors of modern tetracycline resistance proteins, their ability to confer antibiotic resistance was assessed by MIC assay to tetracycline. As shown in Fig. 6F, the overexpression of these proteins in *E. coli* BL21 (DE3) pLysS cells did not result in obvious tetracycline resistance. The absence of detectable resistance in these ancestral proteins suggests that the evolutionary trajectory from generalized translocation activity to antibiotic-specific ribosome protection involved significant adaptive pressures.

Despite their structural differences, both ancestral EF-Gs exhibited no detectable GTPase activity and low catalytic efficiency in contrast to their modern counterparts. This result suggests that early EF-Gs diverged into two forms before con1 acquired robust GTPase activity that is observed in modern EF-Gs. Further exploration of the intermediate nodes is necessary to trace the gradual emergence of this functionality in the evolutionary trajectory.

3.8. The ancestor-catalyzed translocation preserves 3-nucleotide step

The ribosome's position on the mRNA was precisely mapped by super-resolution force spectroscopy (SURFS), a method invented by our groups. This technique enables the

direct measurement of translocation fidelity in a single-turnover reaction, which cannot be resolved by the bulk poly(Phe) assay. As shown in Fig. 7, the pre-translocation (Pre) and post-translocation (Post) ribosome complexes were probed using the P14 DNA probe, which allowed the formation of 14 basepair (bp) and 11 bp duplexes, respectively, between the exposed mRNA and DNA. Duplex formation was revealed by the dissociation forces, which were 48.3 ± 0.8 pN for 14 bp and 19.4 ± 0.3 pN for 11 bp [26,41]. Notably, no intermediate steps or deviations were detected, which would result in a signal decrease in between the two force values. This suggests that the ribosome transitions directly between the Pre and Post states without frameshifting under the catalysis of all three proteins. The 3-nucleotide translocation step was further confirmed by repeating the same experiments with the P15 probe, as shown in Fig. S13. These observations indicated high-fidelity movement along the mRNA, even with Node660 and Node327 that exhibit significantly lower GTPase activities.

These findings imply that translocation fidelity is a highly conserved feature that has been preserved from the ancestral origin of these proteins. In contrast, the increased speed and efficiency observed in the modern EF-G con1 likely evolved later to meet the growing demands of more complex cellular conditions. Alternatively, it is also possible that translocation fidelity is an intrinsic property of the ribosome itself, with EF-G's primary role being to enhance the kinetics of translocation. Further investigations are needed to clarify these possibilities. Even though the absence of a truly primitive ribosome presents a fundamental experimental limit, the combination of new theoretical frameworks and innovative experimental techniques, as shown in this work, provides progressive information.

4. Conclusion

Our work established an AlphaFold-assisted analysis of the conformational landscape of EF-G, linking its sequence, conformation, and function. While our findings strongly suggest that con1 drives translocation and that con2 functions without GTPase activity, direct biochemical validation is necessary to confirm these functional distinctions. Similarly, ancestral sequence reconstruction indicated that essential GTPase activity emerged after con1-con2 divergence, but experimental assays are needed to determine whether this transition occurred gradually and under what evolutionary pressures. Additionally, our structural predictions relied on AlphaFold, which does not capture dynamic interactions with the ribosome. Future biochemical and structural studies, such as ribosome binding assays and kinetic measurements, will be essential for determining how conformational flexibility affects EF-G's interaction with the ribosome and its role in translation fidelity.

The computational pipeline used here can be applied to other translation factors, providing broader insights into translational GTPases and the evolution of the translational machinery. The AlphaFold predictions played a key role in linking structure, function, and evolution, demonstrating its ability to uncover mechanistic details and evolutionary relationships. This approach paves the way for future studies on the origins and adaptations of life's molecular machinery and other biological processes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This work was supported by US National Institutes of Health (5R01GM111452), and US National Science Foundation (BIO-2130427). J.H.S. was supported by US National Institutes of Health (3R01GM111452-07S1). Y. W. was supported by the Welch Foundation E-1721.

References

[1]. Rodnina MV, Savelsbergh A, Katunin VI, Wintermeyer W, Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome, Nature 385 (1997) 37–41, 10.1038/385037a0. [PubMed: 8985244]

[2]. Nierhaus KH, Decoding errors and the involvement of the E-site, Biochimie 88 (2006) 1013–1019, 10.1016/j.biochi.2006.02.009. [PubMed: 16644089]

[3]. Petrychenko V, Peng BZ, de APSAC, Peske F, Rodnina MV, Fischer N, Structural mechanism of GTPase-powered ribosome-tRNA movement, Nat. Commun 12 (2021) 5933, 10.1038/s41467-021-26133-x. [PubMed: 34635670]

[4]. Carbone CE, Loveland AB, Gamper HB Jr., Hou YM, Demo G, Korostelev AA, Time-resolved cryo-EM visualizes ribosomal translocation with EF-G and GTP, Nat. Commun 12 (2021) 7236, 10.1038/s41467-021-27415-0. [PubMed: 34903725]

[5]. Rundlet EJ, Holm M, Schacherl M, Natchiar SK, Altman RB, Spahn CMT, Myasnikov AG, Blanchard SC, Structural basis of early translocation events on the ribosome, Nature 595 (2021) 741–745, 10.1038/s41586-021-03713-x. [PubMed: 34234344]

[6]. Stark H, Rodnina MV, Wieden HJ, van Heel M, Wintermeyer W, Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation, Cell 100 (2000) 301–309, 10.1016/s0092-8674(00)80666-2. [PubMed: 10676812]

[7]. Lin J, Gagnon MG, Bulkley D, Steitz TA, Conformational changes of elongation factor G on the ribosome during tRNA translocation, Cell 160 (2015) 219–227, 10.1016/j.cell.2014.11.049. [PubMed: 25594181]

[8]. Rexroad G, Donohue JP, Lancaster L, Noller HF, The role of GTP hydrolysis by EF-G in ribosomal translocation, Proc. Natl. Acad. Sci. U. S. A 119 (2022) e2212502119, 10.1073/pnas.2212502119. [PubMed: 36282914]

[9]. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, et al. , Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, 10.1038/s41586-021-03819-2. [PubMed: 34265844]

[10]. Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M, ColabFold: making protein folding accessible to all, Nat. Methods 19 (2022) 679–682, 10.1038/s41592-022-01488-1. [PubMed: 35637307]

[11]. Yang Z, Zeng X, Zhao Y, Chen R, AlphaFold2 and its applications in the fields of biology and medicine, Signal Transduct. Targeted Ther 8 (2023) 115, 10.1038/s41392-023-01381-z.

[12]. Evseev P, Gutnik D, Shneider M, Miroshnikov K, Use of an integrated approach involving AlphaFold predictions for the evolutionary taxonomy of duplodnaviria viruses, Biomolecules 13 (2023), 10.3390/biom13010110.

[13]. Wayment-Steele HK, Ojoawo A, Otten R, Apitz JM, Pitsawong W, Homberger M, Ovchinnikov S, Colwell L, Kern D, Predicting multiple conformations via sequence clustering and AlphaFold2, Nature 625 (2024) 832–839, 10.1038/s41586-023-06832-9. [PubMed: 37956700]

[14]. Margus T, Remm M, Tenson T, A computational study of elongation factor G (EFG) duplicated genes: diverged nature underlying the innovation on the same structural template, PLoS One 6 (2011) e22789, 10.1371/journal.pone.0022789. [PubMed: 21829651]

[15]. Han W, Peng BZ, Wang C, Townsend GE, Barry NA, Peske F, Goodman AL, Liu J, Rodnina MV, Groisman EA, Gut colonization by Bacteroides requires translation by an EF-G paralog

lacking GTPase activity, EMBO J. 42 (2023) e112372, 10.15252/embj.2022112372. [PubMed: 36472247]

[16]. Steinegger M, Söding J, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets, Nat. Biotechnol 35 (2017) 1026–1028, 10.1038/nbt.3988. [PubMed: 29035372]

[17]. Sander J, Ester M, Kriegel HP, Xu XW, Density-based clustering in spatial databases: the algorithm GDBSCAN and its applications, Data Min. Knowl. Discov 2 (1998) 169–194, 10.1023/A:1009745219419.

[18]. Edgar RC, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, 10.1093/nar/gkh340. [PubMed: 15034147]

[19]. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, Mol. Biol. Evol 37 (2020) 1530–1534, 10.1093/molbev/msaa015. [PubMed: 32011700]

[20]. Letunic I, Bork P, Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool, Nucleic Acids Res. 52 (2024) W78–W82, 10.1093/nar/gkae268. [PubMed: 38613393]

[21]. Altuntop ME, Ly CT, Wang YH, Single-molecule study of ribosome hierarchic dynamics at the peptidyl transferase center, Biophys. J 99 (2010) 3002–3009, 10.1016/j.bpj.2010.08.037. [PubMed: 21044598]

[22]. Ester M, Kriegel H-P, Sander J, Xu X, A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, AAAI Press, 1996.

[23]. Monteiro da Silva G, Cui JY, Dalgarno DC, Lisi GP, Rubenstein BM, High-throughput prediction of protein conformational distributions with subsampled AlphaFold2, Nat. Commun 15 (2024) 2464, 10.1038/s41467-024-46715-9. [PubMed: 38538622]

[24]. Del Alamo D, Sala D, McHaourab HS, Meiler J, Sampling alternative conformational states of transporters and receptors with AlphaFold2, Elife 11 (2022), 10.7554/eLife.75751.

[25]. Johnson JL, Steele JH, Lin R, Stepanov VG, Gavriliuc MN, Wang Y, Multi-Channel smFRET study reveals a Compact conformation of EF-G on the Ribosome, bioRxiv (2024), 10.1101/2024.01.27.577133.

[26]. Chen Y, Gavriliuc M, Zeng Y, Xu S, Wang Y, Allosteric effects of EF-G domain I mutations inducing ribosome frameshifting revealed by multiplexed force spectroscopy, Chembiochem (2024) e202400130, 10.1002/cbic.202400130. [PubMed: 38923096]

[27]. Majumdar S, Emmerich A, Krakovka S, Mandava CS, Svard SG, Sanyal S, Insights into translocation mechanism and ribosome evolution from cryo-EM structures of translocation intermediates of Giardia intestinalis, Nucleic Acids Res. 51 (2023) 3436–3451, 10.1093/nar/gkad176. [PubMed: 36912103]

[28]. Zhang B, Zou J, Zhang Q, Wang Z, Wang N, He S, Zhao Y, Naman CB, Progress in the development of eukaryotic elongation factor 2 kinase (eEF2K) natural product and synthetic small molecule inhibitors for cancer chemotherapy, Int. J. Mol. Sci 22 (2021), 10.3390/ijms22052408.

[29]. Woese C, The universal ancestor, Proc. Natl. Acad. Sci. U. S. A 95 (1998) 6854–6859, 10.1073/pnas.95.12.6854. [PubMed: 9618502]

[30]. Woese CR, Fox GE, The concept of cellular evolution, J. Mol. Evol 10 (1977) 1–6, 10.1007/BF01796132. [PubMed: 903983]

[31]. Pandit SB, Srinivasan N, Survey for g-proteins in the prokaryotic genomes: prediction of functional roles based on classification, Proteins 52 (2003) 585–597, 10.1002/prot.10420. [PubMed: 12910458]

[32]. Leipe DD, Wolf YI, Koonin EV, Aravind L, Classification and evolution of P-loop GTPases and related ATPases, J. Mol. Biol 317 (2002) 41–72, 10.1006/jmbi.2001.5378. [PubMed: 11916378]

[33]. Connell SR, Takemoto C, Wilson DN, Wang H, Murayama K, Terada T, Shirouzu M, Rost M, Schuler M, Giesebrecht J, et al. , Structural basis for interaction of the ribosome with the switch regions of GTP-bound elongation factors, Mol. Cell 25 (2007) 751–764, 10.1016/j.molcel.2007.01.027. [PubMed: 17349960]

[34]. Suematsu T, Yokobori S, Morita H, Yoshinari S, Ueda T, Kita K, Takeuchi N, Watanabe Y, A bacterial elongation factor G homologue exclusively functions in ribosome recycling in the spirochaete Borrelia burgdorferi, Mol. Microbiol 75 (2010) 1445–1454, 10.1111/j.1365-2958.2010.07067.x. [PubMed: 20132446]

[35]. Tsuboi M, Morita H, Nozaki Y, Akama K, Ueda T, Ito K, Nierhaus KH, Takeuchi N, EF-G2mt is an exclusive recycling factor in mammalian mitochondrial protein synthesis, Mol. Cell 35 (2009) 502–510, 10.1016/j.molcel.2009.06.028. [PubMed: 19716793]

[36]. Taylor DE, Chau A, Tetracycline resistance mediated by ribosomal protection, Antimicrob. Agents Chemother 40 (1996) 1–5, 10.1128/AAC.40.1.1. [PubMed: 8787868]

[37]. Kobayashi T, Nonaka L, Maruyama F, Suzuki S, Molecular evidence for the ancient origin of the ribosomal protection protein that mediates tetracycline resistance in bacteria, J. Mol. Evol 65 (2007) 228–235, 10.1007/s00239-007-9006-z. [PubMed: 17676364]

[38]. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R, IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era, Mol. Biol. Evol 37 (2020) 1530–1534, 10.1093/molbev/msaa015. [PubMed: 32011700]

[39]. Lin J, Divergence measures based on the Shannon entropy, IEEE Trans. Inf. Theor 37 (1991) 145–151, 10.1109/18.61115.

[40]. Wilden B, Savelsbergh A, Rodnina MV, Wintermeyer W, Role and timing of GTP binding and hydrolysis during EF-G-dependent tRNA translocation on the ribosome, Proc. Natl. Acad. Sci. U. S. A 103 (2006) 13670–13675, 10.1073/pnas.0606099103. [PubMed: 16940356]

[41]. Jia H, Wang Y, Xu S, Super-resolution force spectroscopy reveals ribosomal motion at sub-nucleotide steps, Chem. Commun 54 (2018) 5883–5886, 10.1039/c8cc02658k. CRediT authorship contribution statement.

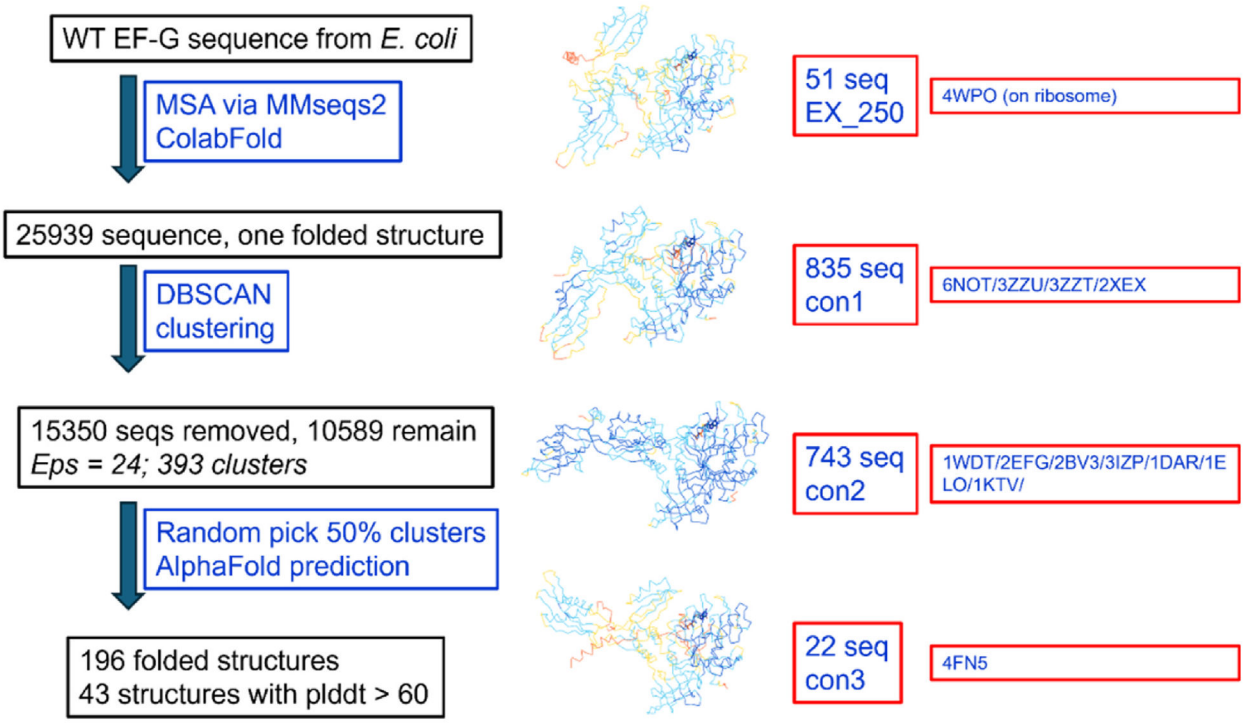

Fig. 1.

AFclustering analysis predicting multiple high-confidence *E. coli* EF-G structures. The four predicted structures align with the experimental data.

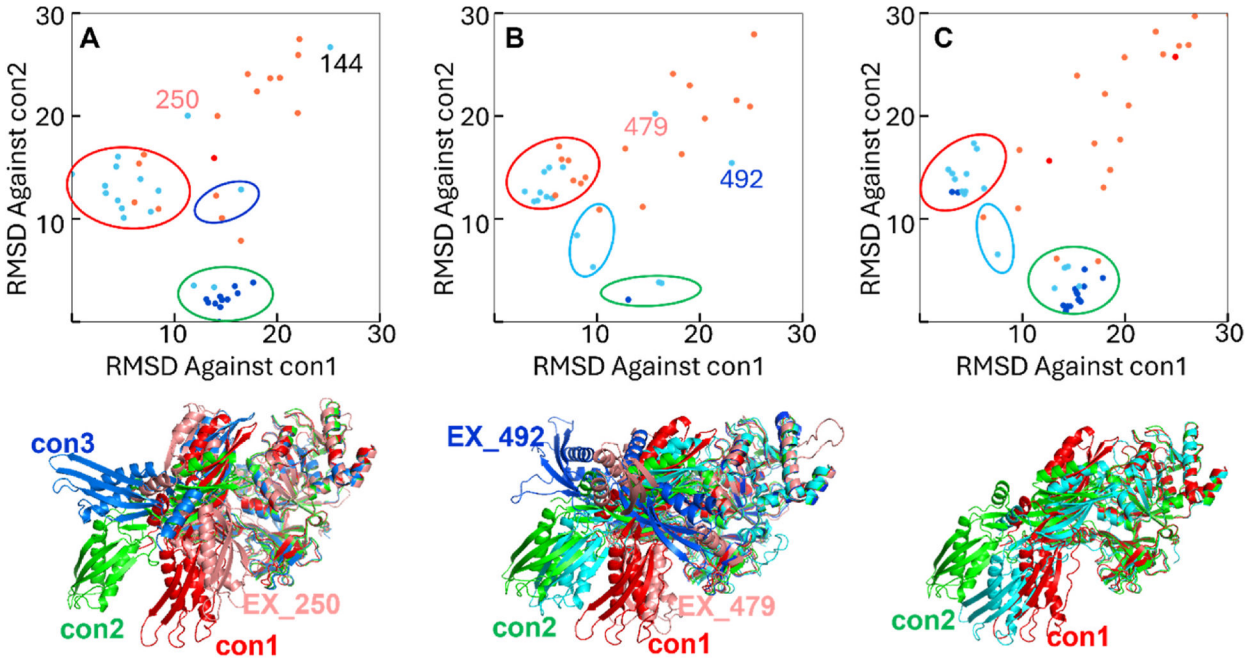

Fig. 2.

Clustering of the AFClustering-predicted structures by RMSD. Panels (A), (B) and (C) illustrate clustering for WT, T49E and T49V, respectively. The red, green, and blue ovals represent con1, con2 and con3 clusters, respectively. The cyan oval (mutants only) lies between con1 and con2. Dot colors indicate the pLDDT values: dark blue (>80), light blue (>70), orange (>60), red (>50). The numbered dots represent single-structure clusters. A representative structure is shown below each RMSD plot.

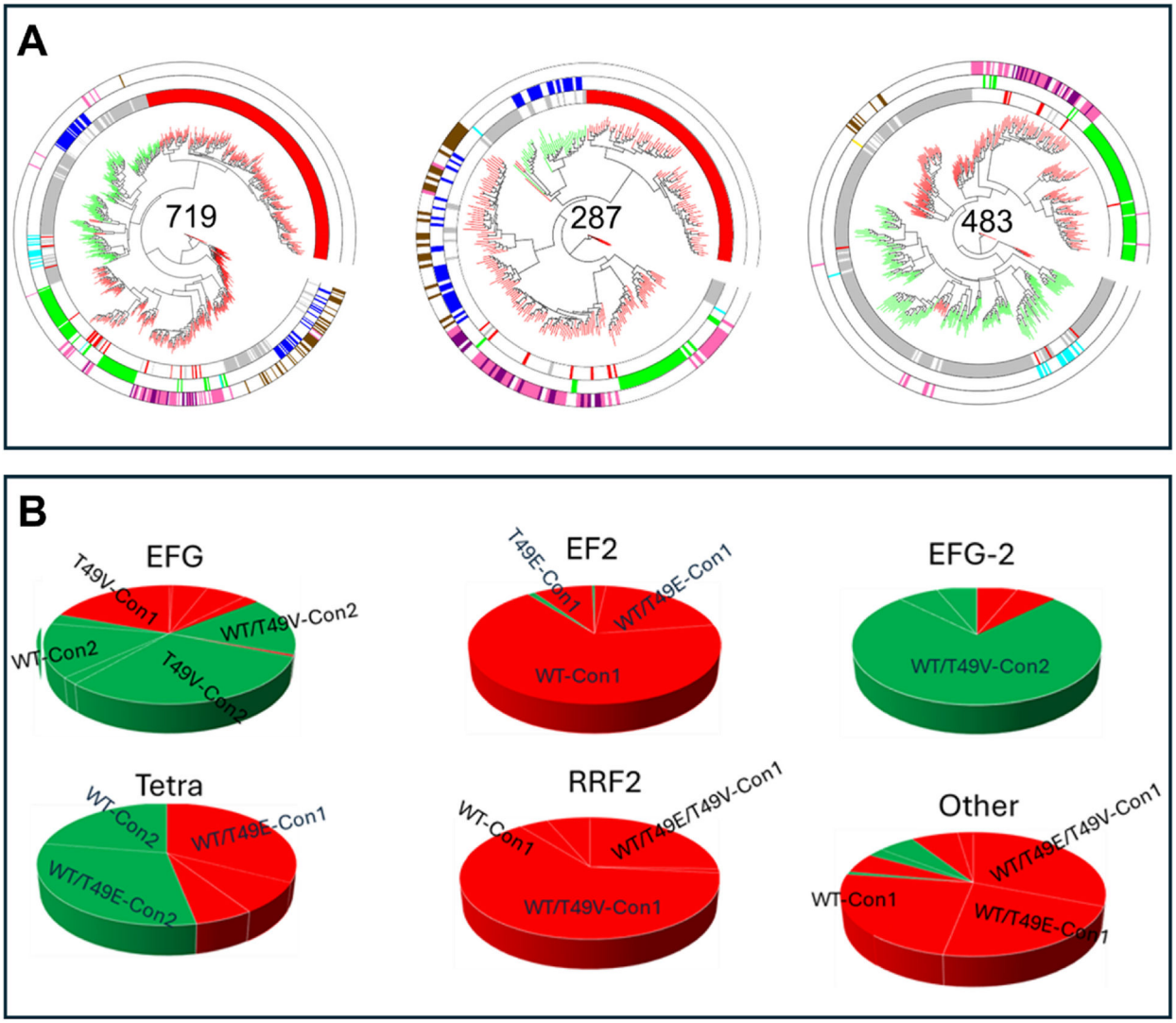

Fig. 3.

Phylogenetic analysis of the MSAs that folded into con1 and con2. (A) Phylogenetic trees for WT, T49E, and T49V, with red and green branches indicating con1 and con2, respectively. The numbers indicate total sequences. Circumference rings show protein IDs (legend on the side). (B) Pie charts illustrating con1/con2 distributions in the WT and mutants.

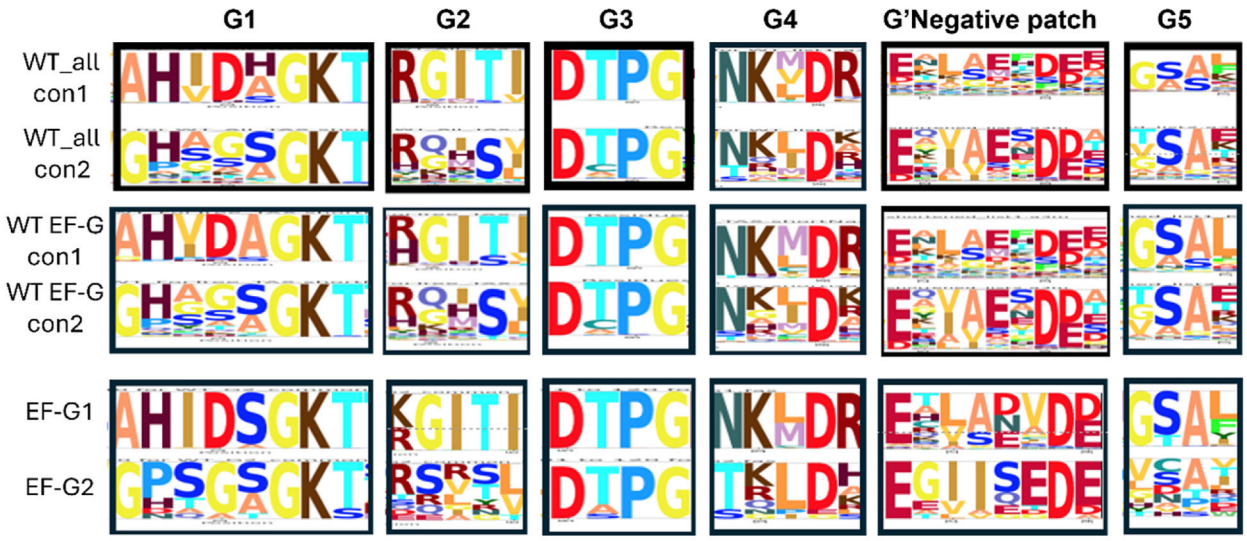

Fig. 4.

Signature motifs in con1 and con2. These motifs closely resemble the sequence-based classification of EF-G1 and EF-G2.

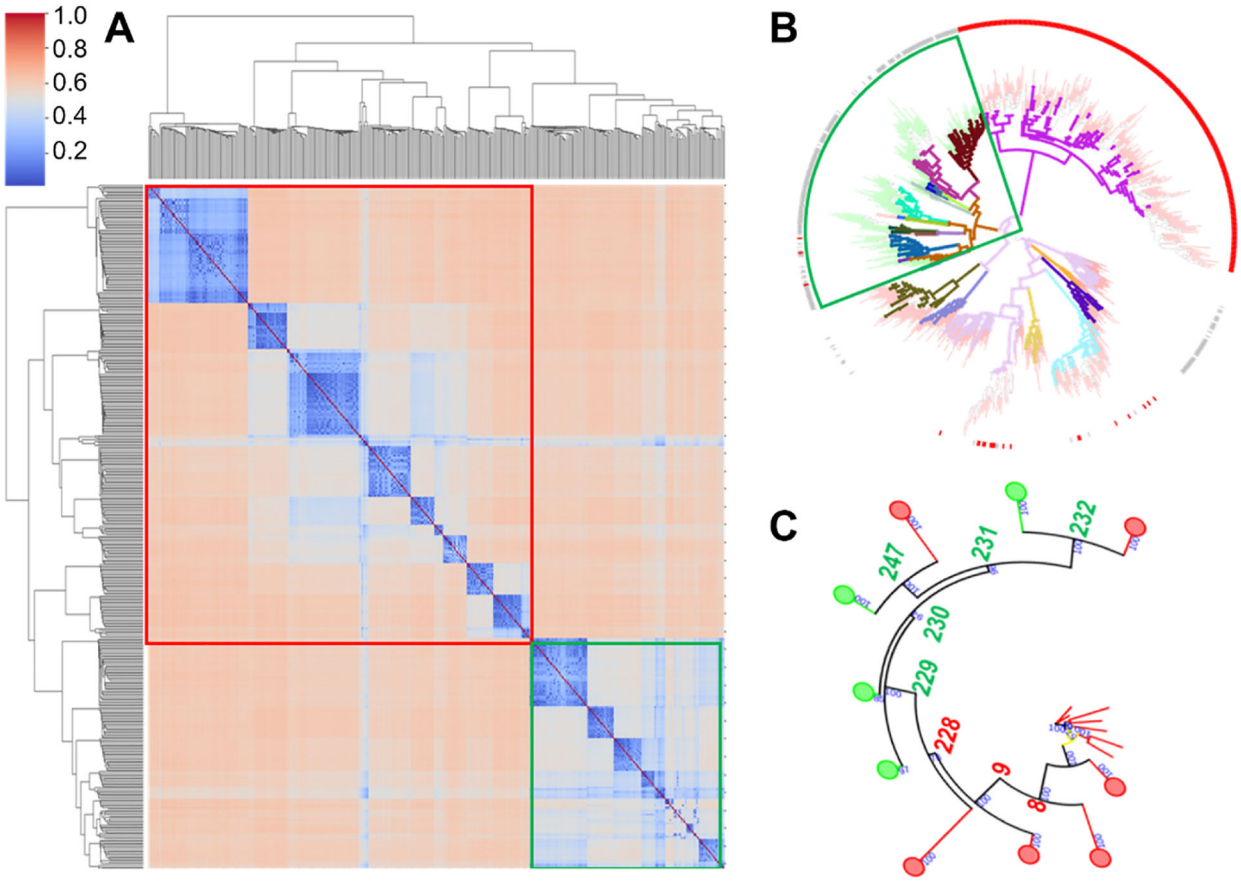

Fig. 5.

Sequence correlation of the internal nodes. (A) Nodes clustered by sequence similarity, with the heatmap scale shown in the top left corner. The red and green boxes highlight the clusters of nodes for the con1 and con2 leaves. (B) Clustered nodes mapped onto the phylogenetic tree showing the EF2 and EF-G rings. The distinct node clusters are colored on the branches. The protein rings for EF2 and EF-G are shown. The leaves are colored red and green for con1 and con2, respectively. (C) Predicted conformations of the selected nodes close to the center. Red: con1; green, con2.

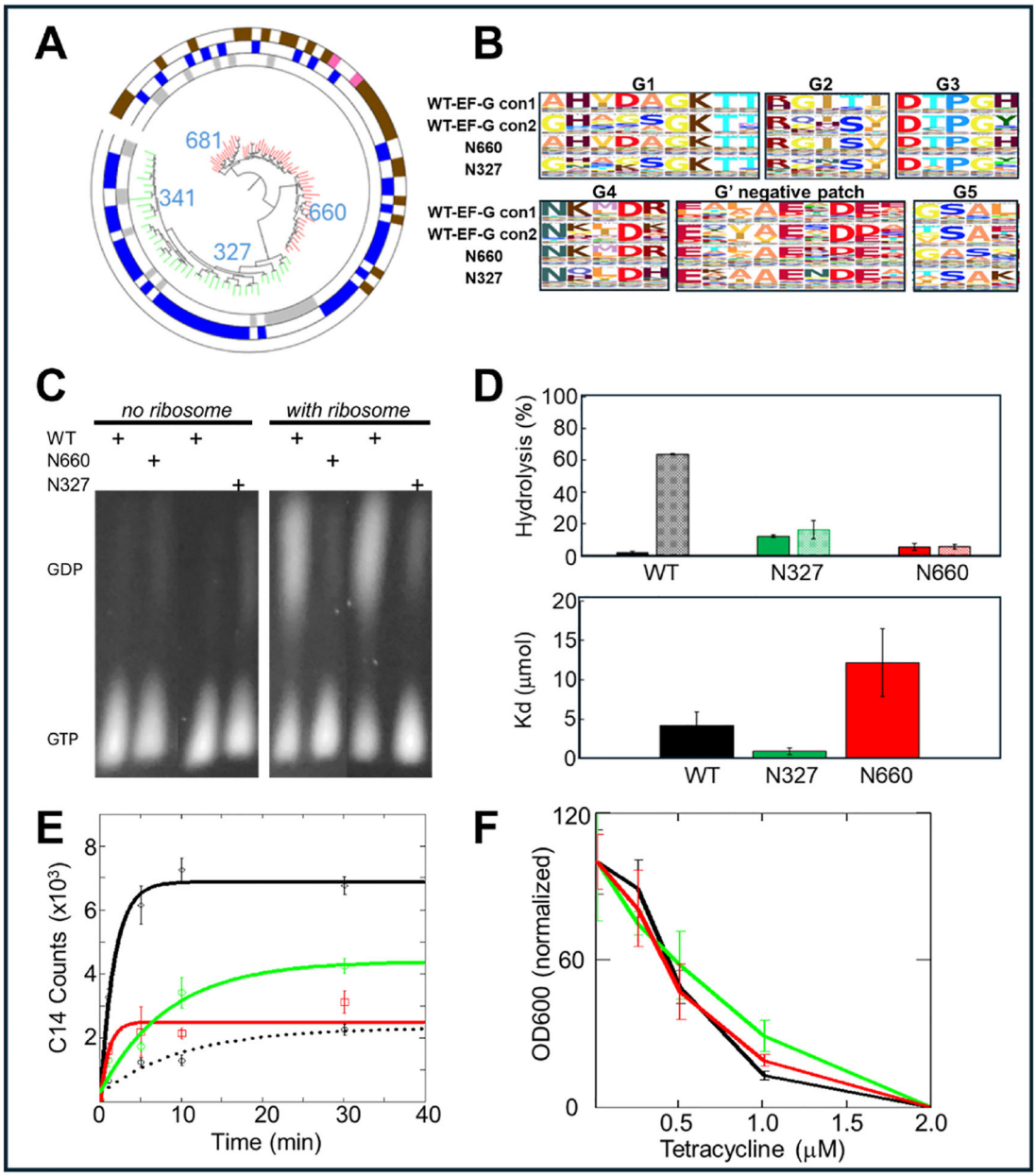

Fig. 6.

Ancestral reconstruction and functional assays. (A) Pruned tree highlighting tetracycline resistance proteins with red (con1) and green (con2) branches. (B) Logo sequence analysis of the key motifs of WT EF-G and ancestral nodes 327 and 660. (C) TLC images of GTP hydrolysis with and without ribosome. (D) GTP hydrolysis percentages (top, solid columns: no ribosome; patterned columns: with ribosome) and binding affinity (bottom). (E) Poly(Phe) assay results for WT EF-G (black), Node327 (green), and Node660 (red). Dashed black line: no EF-G control. (F) MIC assay results for WT EF-G (black), Node327 (green), and Node660 (red).

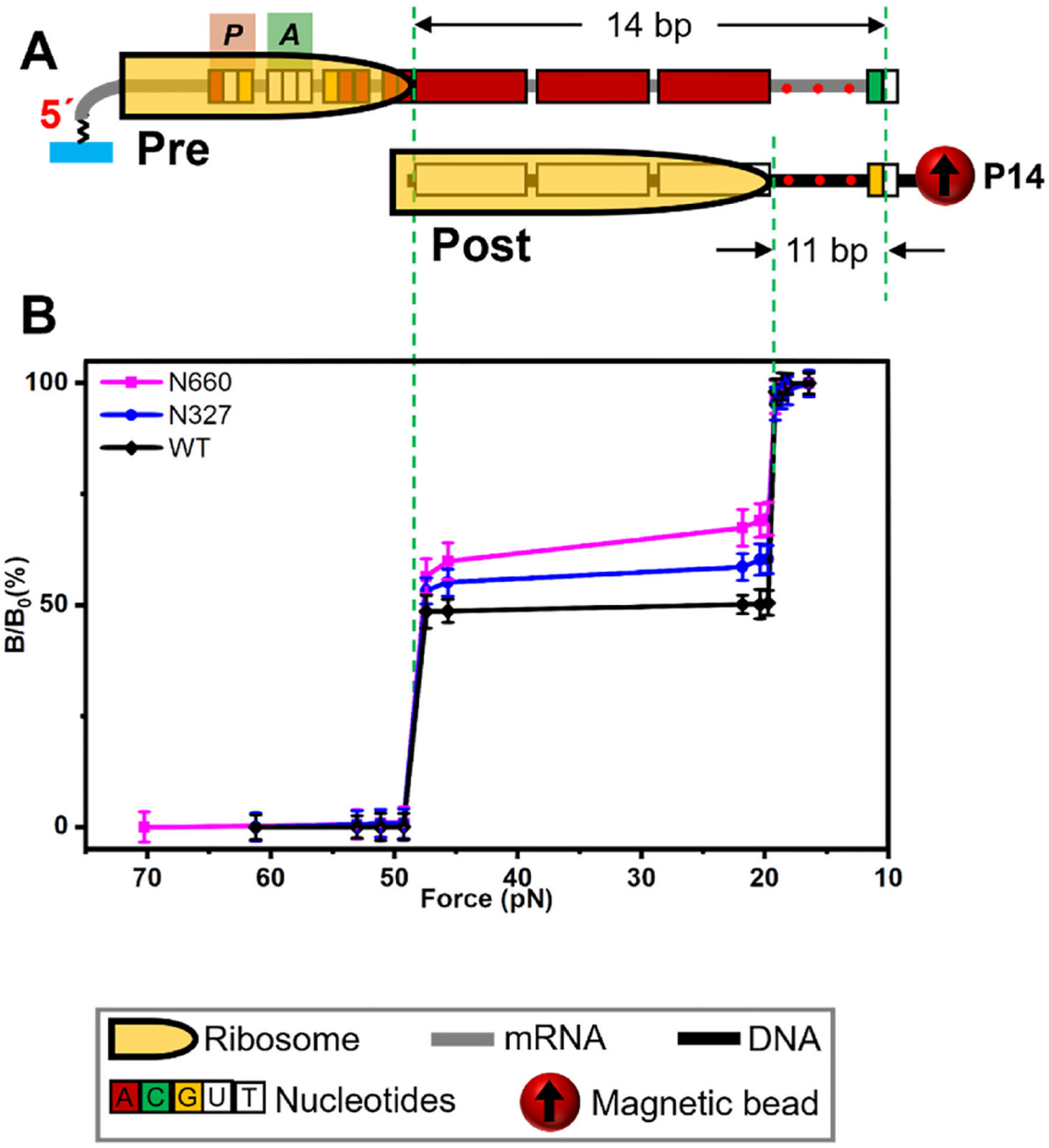

Fig. 7.

Translocation steps revealed by SURFS. (A) Probing scheme with the DNA probe P14 labeled with magnetic beads; the enlarged view highlights the three nucleotides in translocation. (B) Force spectra showing that all three EF-Gs induce the same 3-nucleotide translocation.