Published online 8 June 2016

Nucleic Acids Research, 2016, Vol. 44, No. 17 ***8425–8432***
doi: 10.1093/nar/gkw534

Principles of start codon recognition in eukaryotic translation initiation

Christoffer Lind and Johan Åqvist*

Department of Cell and Molecular biology, Uppsala University, Biomedical Center, Box 596, SE-75124 Uppsala, Sweden

Received March 21, 2016; Revised June 02, 2016; Accepted June 03, 2016

ABSTRACT

Selection of the correct start codon during initiation of translation on the ribosome is a key event in protein synthesis. In eukaryotic initiation, several factors have to function in concert to ensure that the initiator tRNA finds the cognate AUG start codon during mRNA scanning. The two initiation factors eIF1 and eIF1A are known to provide important functions for the initiation process and codon selection. Here, we have used molecular dynamics free energy calculations to evaluate the energetics of initiator tRNA binding to different near-cognate codons on the yeast 40S ribosomal subunit, in the presence and absence of these two initiation factors. The results show that eIF1 and eIF1A together cause a relatively uniform and high discrimination against near-cognate codons. This works such that eIF1 boosts the discrimination against a first position near-cognate G-U mismatch, and also against a second position A-A base pair, while eIF1A mainly acts on third codon position. The computer simulations further reveal the structural basis of the increased discriminatory effect caused by binding of eIF1 and eIF1A to the 40S ribosomal subunit.

INTRODUCTION

Initiation of translation in eukaryotic protein synthesis is a rather complex process. In order to start the translational machinery, the correct codon for initiation must be found. This process is orchestrated by at least 12 proteins, the eukaryotic initiation factors (eIF) (1,2). The small 40S ribosomal subunit, in complex with eIF1, eIF1A and eIF3, recruits the initiator tRNA ($tRNA_i^{Met}$) with anticodon CAU to the peptidyl-site (P-site). The tRNA is joined with the GTP-bound factor eIF2 in a ternary complex (TC), resulting in the 43S pre-initiation complex (PIC). In the next step, the mRNA binds to the 43S PIC together with eIF4 creating the 48S initiation complex. The tRNA binds to the mRNA close to the 5′-cap and begins to scan the 5′-untranslated (5′-UTR) region of the mRNA until the correct start codon, AUG, has been found. Upon binding, the ribosome is in a scanning-promoting conformation induced by eIF1 and eIF1A, referred to as the open conformation (3).

The scanning process and the importance of various initiation factors that participate in the crucial initiation steps have been thoroughly studied (4–13). However, this process is far from being completely understood. In the scanning conformation, the initiator tRNA needs to scan every new triplet of nucleotides positioned in the P-site. It has been proposed that the tRNA exists in an equilibrium between a scanning-competent and scanning-incompetent conformation, depending on which codon is positioned in the P-site (14). This conformational change depends on interactions between eIF1A and eIF5. When the initiator tRNA is scanning the codon its conformation has been denoted as P_{IN}, while when not scanning, and the next mRNA nucleotide is being positioned in the P-site, the conformation is termed P_{OUT}. When in the P_{OUT} conformation, the tRNA still sits in the P-site ∼7 Å away from the codon (15). For the ribosome to achieve the most efficient initiation, it has been argued that the sequences surrounding the start codon assist in correct initiation (7,16,17). The ability to select the correct start codon, or rather the proficiency not to initiate at an incorrect codon, has been studied in yeast (11) where the rate and affinity of TC binding to a mRNA-preloaded-ribosome were measured, as well as the effects of single base changes in the codon. Further *in vivo* experiments showed that some positional changes, particularly those leading to purine–purine mismatches in the 2nd codon position, have a higher tendency to affect the efficiency of initiation than others (11). Recent three-dimensional structures have been determined of the eukaryotic initiation system (15,17–19), which contribute with structural information regarding the eukaryotic initiation machinery.

During start codon scanning, the pre-initiation complex is bound to encounter very many near-cognate codons. How the initiator tRNA manages to accurately scan all new codons positioned in the P-site at a speed of ∼8 nucleotides per second (20,21), and still maintain a low error rate, is not fully understood. To get a deeper insight into how the start codon gets distinguished by the initiator tRNA compared

*To whom correspondence should be addressed. Tel: +46 18 471 4109; Email: aqvist@xray.bmc.uu.se

to any other codon, we have employed extensive molecular dynamics (MD) simulations and free energy perturbation calculations. Simulations were started with the yeast pre-initiation complex (19) containing AUG in the P-site (mRNA sequence 5′-UCUAUGCUC-3′) and the binding energetics associated with changing the start codon to different near-cognate codons was evaluated. These computational mRNA mutations were done for each nucleotide position of the codon, and we have analyzed two different mismatches in all three positions corresponding to the GUG, CUG, ACG, AAG, AUA and AUC near-cognate triplets. During the scanning process the initiator tRNA is accompanied by eIF1 and eIF1A, which are located toward the E-site and A-site, respectively. To determine the effect of the initiation factors due to their proximity to the tRNA, the free energy calculations were done both in their presence and absence to explore the complete energetic profile of the initiator tRNA's capacity to sense the correct start codon. The results clearly show how eIF1 and eIF1A increase the discriminatory power of mRNA scanning with regard to the 1st and 3rd codon position, but that they have no influence on a central C–A mismatch. However, the initiation factors show a large discriminatory effect also against a 2nd position A–A mismatch. Upon AUG recognition, eIF1 dissociates from the complex (8). Hence, we also carried out additional free energy calculations in the presence of only eIF1A for all the three codon position mismatches. These calculations reveal that the entire discriminatory effect against a 1st position G-U mismatch originates from the binding of eIF1 and also show a comparable effect for a central A–A mismatch.

MATERIALS AND METHODS

Molecular dynamics (MD) simulations

Initial coordinates of the yeast PIC complex were retrieved from the cryoEM reconstruction (PDB accession number 3J81) with initiator tRNA in the P_{IN} state and the eukaryotic initiation factors 1 and 1A (19). MD simulations were performed with the Q software (22) and the simulation scheme was adopted from earlier simulations of tRNA binding to ribosome (23,24). A simulation sphere of 25 Å radius centered on the N1 of residue A35 of the $tRNA_i^{Met}$ was solvated in a droplet of water with 36 Å radius. Water molecules close to the sphere boundary were subjected to radial and polarization restraints according to the SCAAS model (22,25). Heavy solute atoms further than 22 Å from the center were harmonically restrained to their initial positions with a 10 kcal/mol/Å^2 force constant throughout the simulations. The total charge of the system was kept neutral by scaling down RNA phosphate charges and neutralizing ionizable protein groups near there the sphere boundary, as done in earlier work (26,27), and no additional ions were added. Five residues of eIF2 were included in the simulation system, but were unionized and located in the restrained boundary region, and thus have no influence on the calculated binding energetics. The MD simulations were performed using CHARMM22 force field (28,29) starting with a 1 fs time step during initial low temperature equilibration, and subsequently increased to 2 fs for production simulations and free energy calculations. A direct 10 Å cutoff for non-bonded interactions was used, with electrostatic interactions beyond the cutoff treated by the local reaction field multipole expansion method (30). No cutoffs were used for atoms of the nucleotide bases that were transformed in the free energy calculations. The simulation systems were gradually heated during equilibration to a final temperature of 310 K.

Free energy calculations

The free energy perturbation (FEP) method (31) was used to calculate relative free energies of binding. Mismatches were introduced by mutating one nucleotide at each of the three start codon positions (AUG → GUG, AUG → CUG, AUG → AAG, AUG → ACG, AUG → AUA and AUG → AUC, respectively). The calculations were done in the presence and absence of initiator tRNA according to the thermodynamic cycles in Figure 1. Simulations were also carried out with and without eIF1 and eIF1A in order to explore the effect of the initiation factors on codon selection. The free energy calculations used 51 FEP windows and 10 independent replicas, with randomized initial velocities generated from the Maxwell–Boltzmann velocity distribution. This yielded a total data collection time of 40 ns of for each FEP calculation after equilibration and heating of the system. When transforming uracil to adenine at the 2nd codon position and guanine to cytosine in the 3rd codon position a dual topology method was used. Addition of the base requires modeling of its initial position. This was achieved by aligning the new base on top of the base present in the cryo-EM structure, so that the geometries around N9 in G and N1 in C agree with each other. In case of the U→A mutation we needed to shift the Watson–Crick edge of the adenine base to avoid the steric clash in a purine–purine mismatch. The free energy calculations for the U→A and G→C transformations were then initiated in the mixed state with 50% of each end-point potential, and then propagated toward the uracil/guanine and adenine/cytosine endpoints. This approach is preferable for avoiding conformational bias from modeling and MD equilibration of the initial structure.

RESULTS AND DISCUSSION

Energetics of start codon scanning

Extensive MD simulations and free energy perturbation calculations were used to evaluate the free energies of initiator tRNA binding to near-cognate elongation codons (GUG, CUG, ACG, AAG, AUA, AUC), relative to the standard initiation codon AUG (Figure 2). The MD simulations were based on the yeast 48S PIC structure, with initiator tRNA, eIF1, eIF1A and TC (19) bound to the small ribosomal subunit. To evaluate the role of the eukaryotic initiation factors 1 and 1A for the accuracy of start codon selection, the free energy calculations were done both in the presence and absence of the factors. Additional calculations with only the eIF1A factor present were also carried out. Relative binding free energies ($\Delta\Delta G_{bind}$) were obtained via standard thermodynamics cycles (Figure 1) which require simulations of each mutation both with and without initiator tRNA bound to the system. In practice, the tRNA needs to dynamically

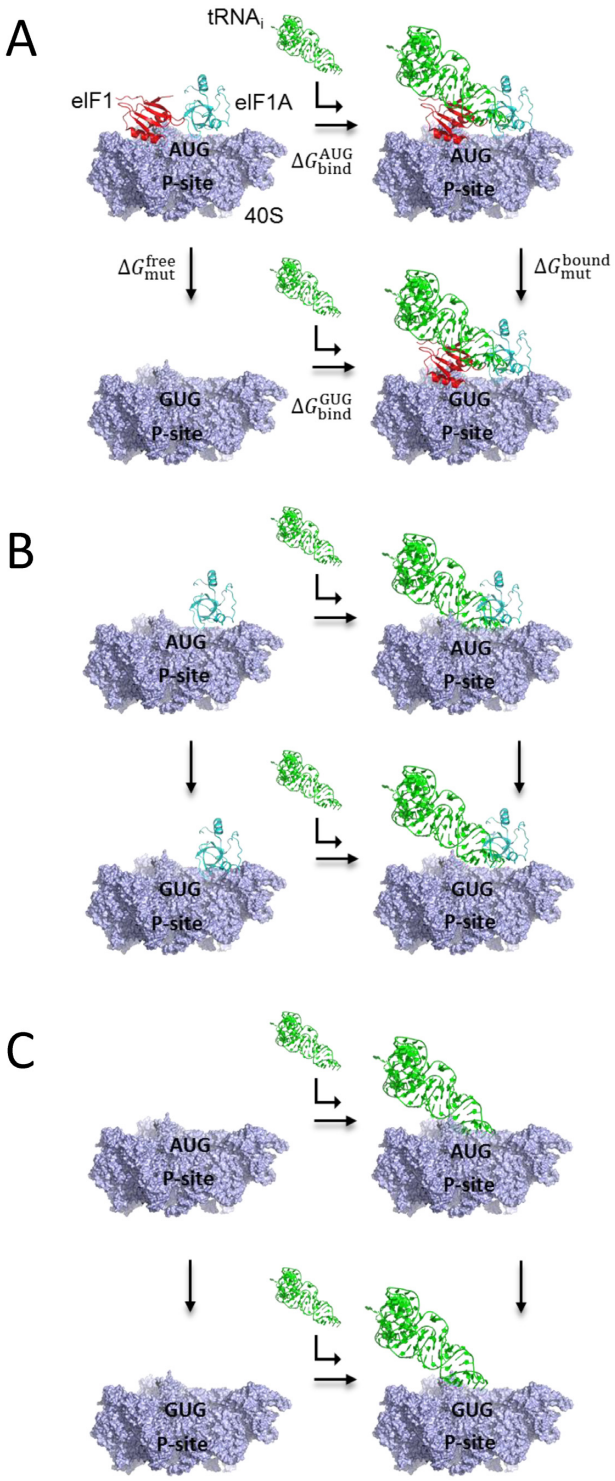

Figure 1. Schematic overview of the thermodynamic cycles used to calculate the relative free energy of tRNA (green) binding to different codons, using GUG as an example (eIF2 is omitted for clarity). The simulations start with the cognate AUG codon in the P-site, which is mutated into GUG, yielding a 1st position G-U mismatch. The three different cycles correspond to (**A**) the cases with both eIF1 (red) and eIF1A (cyan) present, (**B**) the case with only eIF1A present and (**C**) tRNA binding to the 40S subunit in absence of the two factors. The relative binding free energies are calculated as $\Delta\Delta G_{\text{bind}} = \Delta G^{GUG}_{\text{bind}} - \Delta G^{AUG}_{\text{bind}} = \Delta G^{bound}_{\text{mut}} - \Delta G^{free}_{\text{mut}}$, where molecular dynamics (MD) simulations are used to obtain the energetics of the vertical legs.

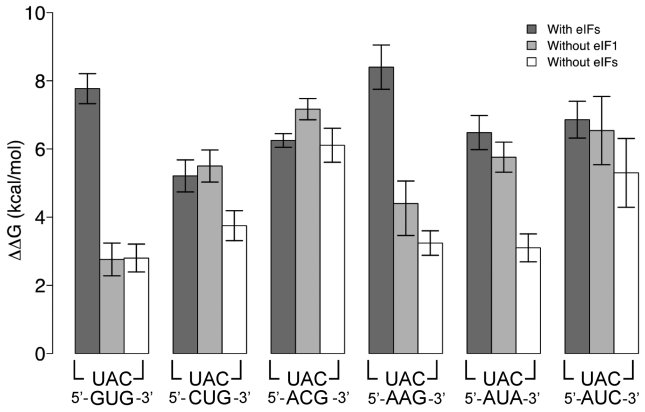

Figure 2. Calculated initiator tRNA binding free energies to different near-cognate codons, relative to AUG. Dark grey and white bars denote ribosome initiation complexes with and without eIF1 and eIF1A present, respectively. Light grey bars correspond to the state with only eIF1A bound. Error bars are standard errors of the mean from 10 independent simulations.

move in and out so that the mRNA can be translocated through the ribosome, thereby positioning new nucleotides for scanning in the P-site. Hence, the MD simulations without bound tRNA can be also seen as a model for the P_{OUT} conformation, since in the P_{OUT} state initiator tRNA anticodon is ∼7 Å away from the mRNA codon and makes no contact with it.

The calculated energetics (Figure 2 and Supplementary Table S1) shows that the initiator tRNA binding free energy penalties are large when either of the three codon positions are mutated from their cognate AUG sequence. With both eIF1 and eIF1A bound, the discrimination against near-cognate codons is 5–8 kcal/mol, which thus demonstrates a very strong specificity for the AUG initiation codon corresponding to more than a factor of 4000. A 1st position G-U or C-U mismatch yields a tRNA binding free energy penalty of 7.8 and 5.2 kcal/mol, respectively. A 2nd position C–A or A–A base pair gives 6.2 and 8.4 kcal/mol, respectively, and 3rd position A–C and C–C mismatches yield 6.5 and 6.9 kcal/mol, respectively. These results are also consistent with *in vivo* initiation efficiencies, which show a relatively uniform discrimination across the three positions, except for the AAG and AGG codons that have very low initiation efficiency (11,32). Furthermore, the CUG codon, which has the lowest energy penalty in our calculations, appears to be most efficient near-cognate initiation codon *in vivo* (11,32). However, in absence of the two initiation factors there is a marked loss of fidelity for near-cognate base pairs in the 1st and 3rd codon positions, as well as for a 2nd position A–A mismatch. In the former cases, this would also be expected from a structural viewpoint since eIF1 and eIF1A have their closest contacts with these positions of the codon-anticodon minihelix, while the effect on the central position is more surprising. Hence, with neither eIF1 nor eIF1A present the discrimination against the 1st position G-U mismatch drops to 2.8 kcal/mol, and for the 2nd position A–A and 3rd position A–C mismatches it similarly drops to 3.2 and 3.1 kcal/mol, respectively. The calculations thus predict that the intrinsic discrimination for initiator tRNA

binding to the P-site is only about a factor of 150 against these near-cognate codons, and that eIF1 and eIF1A together add a factor of the same magnitude. In contrast, the two factors are found to have essentially no effect on the 2nd position near-cognate C–A mismatch, which also has the largest intrinsic energy penalty (Figure 2 and Supplementary Table S1). Interestingly, we find that the less conservative pyrimidine–pyrimidine mismatch (C–C) in the 3rd position also has an intrinsically large penalty associated with it (5.3 kcal/mol) in absence of the two initiation factors.

To further quantify the effect of the two initiation factors, we carried out free energy calculations for the case with only eIF1A present. This state would be representative of a successful start codon recognition event, which induces P_i release and is followed by dissociation of eIF1 (8). These calculations show that it is eIF1 that is responsible for protection against a near-cognate 1st position G-U mismatch. Hence, in the absence of eIF1 the energetic penalty drops to 2.8 kcal/mol, which is the same as without any of the two factors bound. That is, eIF1 contributes 5 kcal/mol to the G-U discrimination, or a factor of ∼4000. Conversely, the tRNA binding affinity change for the 3rd position near-cognate A–C mismatch is essentially unaffected by removal of eIF1, which shows that in this case it is eIF1A that predominately causes enhanced specificity for the start codon position. The fact that two factors mainly exert their effect on 1st and 3rd position codons with near-cognate mismatches, associated with intrinsically low discrimination, suggests that this may be their prime role in initiation. However, it can also be seen (Figure 2) that eIF1 has a large effect on the penalty against a 2nd position A–A mismatch. This corresponds to a factor of ∼800 (4 kcal/mol) and indicates that eIF1 also has a role in discriminating against purines at the middle codon position.

Structural origin of discrimination against near-cognate mismatches

In order to interpret the structural basis of the calculated energetics average MD structures were generated for all simulation replicas. Examination of the structures with a cognate start codon shows that the pre-initiation complex and its tRNA are very stable at AUG. During the MD simulations the backbone structure of the eIF1 β-hairpin loop-1 is positioned slightly closer to the anticodon (Figure 3) than seen in the cryo-EM reconstruction (19). The simulations further show the β-hairpin loop does not directly interact with the start codon, instead it is packed closer to the third nucleotide of the anticodon (U36), thereby providing stabilizing interactions to the tRNA and the anticodon via backbone contacts. Whereas the sidechains do not directly participate in start codon recognition, Lys37 of the β-hairpin loop-1 is positioned below the sugar moiety of the codon between the 1st and 2nd position (Figure 3). Arg33 and Arg36 of eIF1, located on opposite sides of the β-hairpin loop, are pointing toward the E-site and A-site, respectively. The N-terminal tail (NTT) of eIF1A is, as expected, quite flexible but becomes more ordered around Gly9. eIF1A is located toward the A-site and is close to the 3rd codon position where the flexible NTT points away from the codon. However, Gly8 interacts with the amine of the 3rd position guanine (G3) which gives rise to an extra hydrogen bond in addition to those of the G-C base pair (Figure 4). Further, C34 of the tRNA anticodon interacts directly with the backbone of Lys10 via hydrogen bonding to its amide NH group. These new interactions are attained by a shift of the NTT backbone ∼2 Å closer to the G-C pair, compared to the cryo-EM structure (19). As a result of this, the Lys10 sidechain gets deeper into the cleft between G3 and the following C(+4) residue (Figure 4).

When introducing a mismatch in the codon, the MD structures diverge considerably from the initial structure with cognate AUG (Figure 5). It is noteworthy that when introducing a 1st position mismatch, the eIF1 β-hairpin loop becomes significantly more disordered than for mismatches in the 2nd and 3rd codon positions. Again, the NTT of eIF1A is observed to be more flexible regardless of codon-anticodon pairing. These results are well in line with the conclusion that eIF1 is mainly involved in assuring a correct 1st position nucleotide rather than a correct start codon as a whole. Compared to the simulations without the initiation factors (Supplementary Figure S1) the codon-anticodon interactions are generally kept more stable in the presence of eIF1 and eIF1A.

Introducing a G-U mismatch into the 1st codon position is predicted to push the eIF1 β-hairpin loop away from the tRNA, thus creating more space for the anticodon to accommodate the unfavorable hydrogen bonding in a G-U base pair (Figure 5A). Hence, the loss of guanine hydrogen bonds and the structural shift of eIF1 are likely to be the cause of the large energetic penalty of ∼8 kcal/mol against GUG. Yet, even though the β-hairpin loop undergoes a rather large conformational change Lys37 remains positioned at approximately the same location. This suggests that Lys37 has an important role in sensing the mRNA and holding the codon in place, and it is also seen to exclude water molecules from the Watson–Crick edge of the codon.

The 2nd position base is sequestered between the two purines in the start codon, and located between the loops of eIF1 and eIF1A without the possibility of any direct interaction with either of them. The calculated binding free energy changes associated with mutation of the cognate uracil suggests that the strong discrimination against a C in the middle position predominantly originates from the strictly sandwiched conformation of the nucleotide base. Reading of the 2nd position is essentially entirely determined by the conformation of the middle position bases and an A–C mismatch inevitably has unsatisfied hydrogen bonds in the absence of compensating solvent interactions (Figure 5C). In case of the near-cognate A–C pair, the free energy penalty is not affected by removal of the initiation factors and the middle position of the AUG start codon thus only appears to be sensed by the initiator tRNA. Such an intrinsically larger discrimination against second position A–C mismatches (compared to the 1st and 3rd codon position) has also been seen both from computation and experiment in standard peptide elongation decoding (24,33). In contrast, removal of either eIF1 or both initiation factors has a pronounced effect on an A–A mismatch in the same position. In this case the MD simulations predict a clear distortion of the anticodon in order to accommodate the

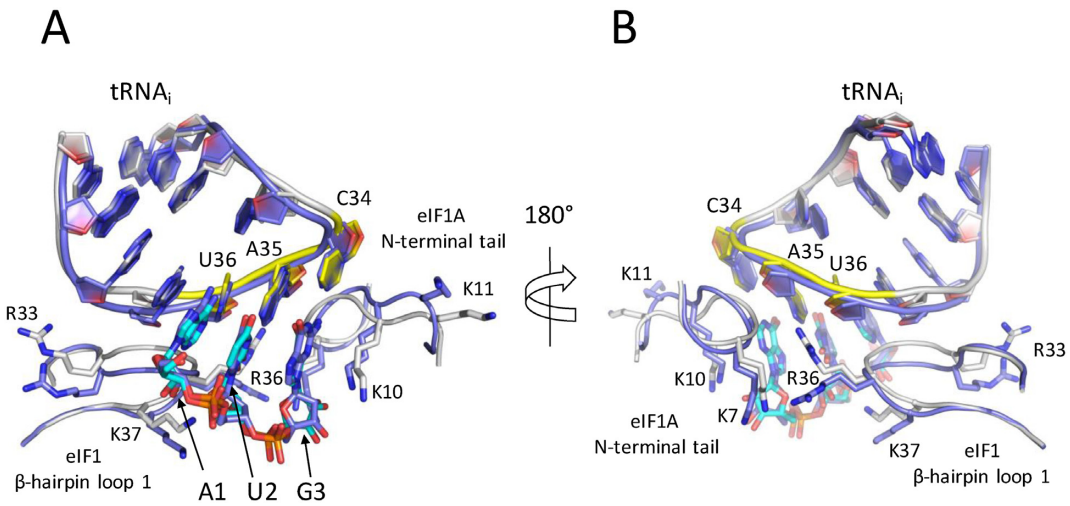

Figure 3. Comparison of the initial structure (blue) and average MD structure for the complex with the cognate AUG start codon. Initiation factors and the initiator tRNA stem loop are shown in white, the CAU anticodon in yellow and the mRNA codon in cyan. The right and left images are related by a 180° rotation as indicated.

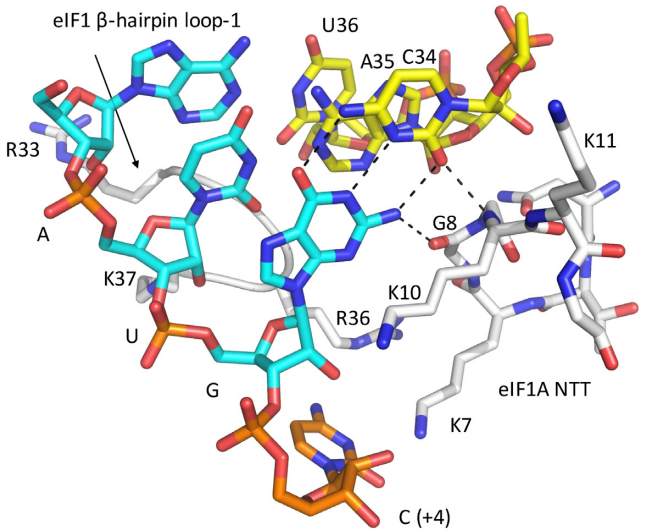

Figure 4. Average MD structure of the AUG complex with eIF1 and eIF1A present. The backbone of eIF1A provides additional hydrogen bonds contributing to the specificity for reading a guanine at the 3rd position. The mRNA codon is shown in cyan, the CAU anticodon in yellow and eIF1A and eIF1 in white. The first mRNA nucleotide of the A-site (+4 nucleotide) is shown in orange.

much larger adenine (Figure 5D), even with the two factors present.

Reading of the 3rd position depends both on favorable codon-anticodon interactions and on communication with eIF1A, as discussed above. With an adenine in the 3rd codon position, the described interaction pattern will be lost and an A–C mismatch is structurally more distortive than near-cognate mismatches in either the 1st or 2nd position (Figure 5E). The anticodon nucleotide is shifted away from the codon and mobility of the eIF1A NTT is dramatically increased. Nonetheless, the position of Lys10, wedged between the 3rd codon position and the +4 nucleotide, is still maintained. Aside from the codon-anticodon and eIF1A interactions, the anticodon is stacking with the C1637 rRNA base, which also interacts with Lys11 of the NTT loop. If a perfect base paring occurs, then the NTT appears to communicate with the eIF1 β-hairpin loop, which has Arg36 pointing toward eIF1A. In the case of reading AUC with a resulting C–C mismatch (Figure 5F), the cavity accommodating the base pair is too large to be compensated by other interactions and the intrinsic energy penalty without eIFs is considerably higher than for an A–C mismatch (Figure 2). With the initiation factors present the discrimination is approximately the same (∼7 kcal/mol) as for the A–C mismatch. Furthermore, the NTT is considerably more flexible when a cytosine is present as the 3rd nucleotide.

The surroundings of the start codon have been shown to be important for start codon recognition (16,34,35). In the yeast pre-initiation complex reconstruction (19), the immediate upstream 5′-UTR sequence of the start codon (Kozak sequence) consists of U(-1), C(-2) and U(-3), while the downstream A-site codon is C(+4), U(+5) and C(+6). Note, that while the vertebrate initiation sequence has a strong preference for G(+4) this is not the case for yeast (36). The shape of the mRNA is rather peculiar, with the mRNA nucleotides at positions −1 and +4 flipped with about 180° with respect to each other (Figure 6). The cytosine at +4 was proposed to interact with Trp70 of eIF1A (19). However, the present simulations suggest that it does not, instead the flipped nucleotide at the +4 position is dependent on the position of eIF1A and the described location of Lys10. Simultaneously, Trp70 stacks with Lys7 rather than the +4 nucleotide. This line-up of both rRNA and amino acid residues surrounding the 3rd position appears to protect the 3rd codon position more than the 1st and 2nd. On the other hand, the nucleotide at −1 is more flexible than +4, and predominantly stacks with the G1150 rRNA base, but the conformation at the −1 nucleotide does not seem to contribute to the start codon selection (Figure 6).

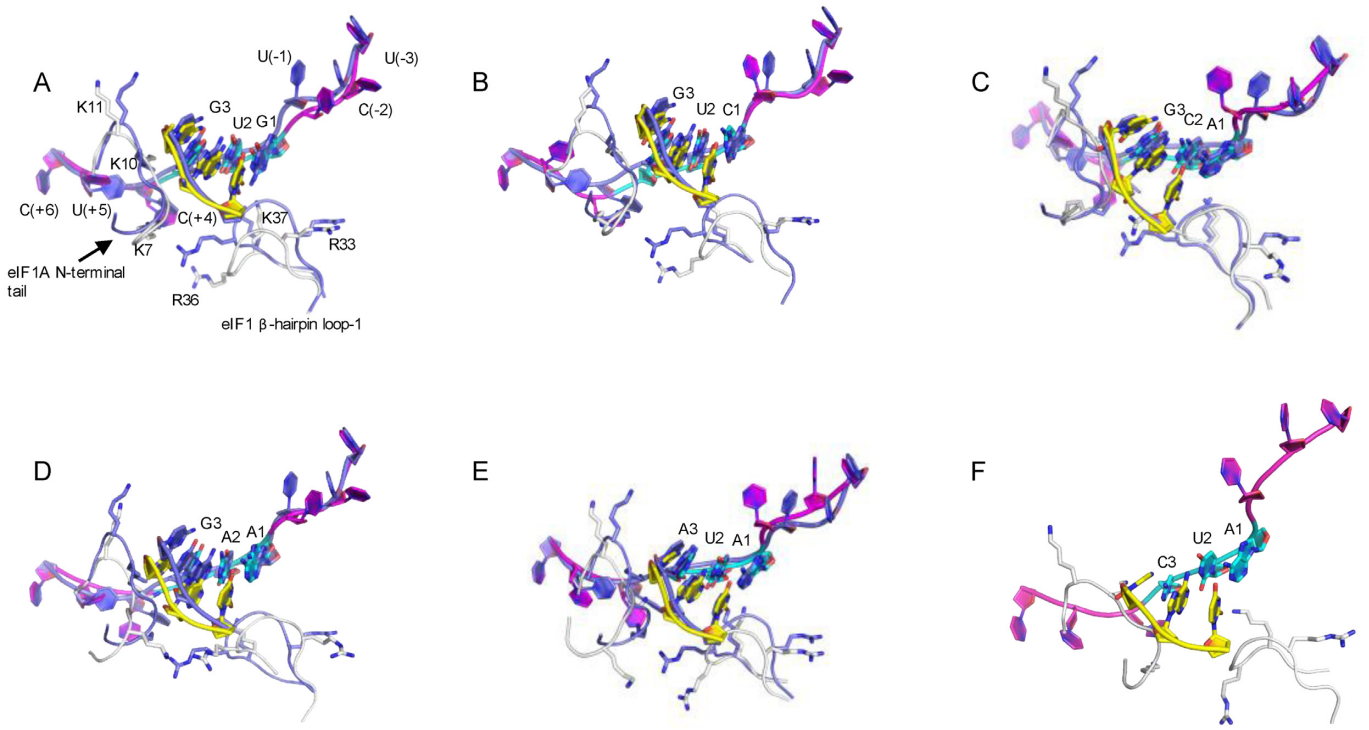

Figure 5. Structural view of the near-cognate mutations with bound tRNA, showing the anticodon of the initiator tRNA, eIF1, eIF1A and the mRNA passing through the mRNA channel. Panel (**A**) – GUG codon, (**B**) – CUG, (**C**) – ACG, (**D**) – AAG, (**E**) – AUA and (**F**) – AUC. Initiation factors are shown in white, codon and anticodon in cyan and yellow, respectively, and the mRNA in magenta. The initial structure with AUG is shown in blue.

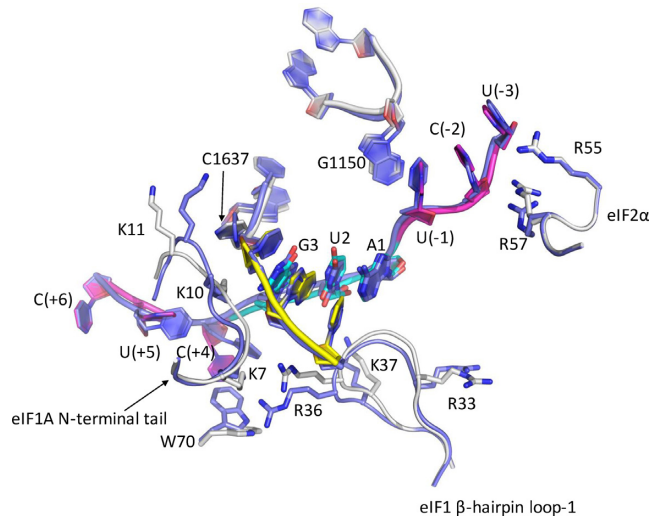

Figure 6. Average MD structure of the AUG complex, with the initial structure (blue carbons) overlaid, showing the 5′-UTR Kozak element consisting of a weak sequence UCU in the P_{IN} state. rRNA residues G1150 and C1637 and part of eIF2α are also indicated. The MD structure is shown in white, with the start codon in cyan, anticodon in yellow and the up- and downstream mRNA in magenta.

Structural properties for scanning a correct start codon in the P_{OUT} state

Assuming that the tRNA cannot sense which codon is positioned in the P-site when the anticodon is in the P_{OUT} conformation, our simulations in the absence of the initiator tRNA serve as a model for this specific state. In the tRNA-free simulations with the AUG codon, both eIF1 and eIF1A behave similarly as in the simulations with initiator tRNA present. Notably, the eIF1 β-hairpin loop-1 bends away from the codon, thereby creating more space for the mRNA to move (Supplementary Figure S2). The loop adopts a similar conformation to that previously seen in the mammalian ribosome complex (17). Hence, this further indicates that the β-hairpin loop directly interacts with the tRNA rather than with the codon in the initial selection. Additionally, the NTT of eIF1A maintains its overall structure, as seen in the cryo-EM reconstruction with AUG, with Lys10 positioned properly to keep the +4 nucleotide flipped out from the codon. As this interaction is seemingly needed, mutation of a nucleotide at any of the three codon positions causes disorder of the NTT and the mRNA (e.g. Figure 5A). Nonetheless, eIF1 does not seem to be as affected by the movement of eIF1A with an incorrect codon, suggesting that in the P-site eIF1 and eIF1A do not communicate in order to guide the tRNA in selecting the correct over incorrect start codon. The communication between the factors rather occur upon a perfect codon-anticodon interaction as described above. The calculated energetics thus suggest that the initiation factors boost the discriminatory effect of the tRNA to prevent misreading, rather than actively participate in start codon selection.

The eIF1A factor has been described to effectively bind RNA (12,37). Examining the average structures generated from the MD simulations, we have identified interaction

patterns of the eIF1A NTT. The NTT contains several positively charged amino acids close to the ribosomal P- and A-sites. The discussed conformation of Lys10 is not only identified for AUG but is also seen for mismatches at all mutated codon positions, both in presence and absence of the tRNA, as well as in the simulations without eIF1. This suggests that the NTT interactions may be of importance for the scanning mechanism. The NTT has been shown to promote leaky scanning when alanine mutations are being introduced at different locations in the tail (38), where mutations of residues 7–11 are lethal. We have observed these residues of eIF1A both to interact with the anticodon and to provide additional hydrogen bonds to the amine of the 3rd codon position guanine (Figure 4). The interaction network described here could thus potentially be crucial in preventing leaky scanning, by promoting the flipped-out conformation of the +4 nucleotide. In this way, the NTT of eIF1A could play a role in the rate-limiting step of tRNA switching between the P_{OUT} and the P_{IN} state, by letting through one amino acid to the P-site at a time in a stepwise manner. It is evident from the tRNA-free simulations that Lys10 and C1637 (Supplementary Figure S2) might act like a conformational gate, preventing the mRNA to be protracted by more than one nucleotide at the time. In this way, the mRNA has to be pulled in from a bent conformation rather than straight, which in turn might have an effect on the rate with which the mRNA is being scanned.

CONCLUSIONS

The present free energy calculations clearly show how the two initiation factors eIF1 and eIF1A together play an important role in the eukaryotic translation initiation process. The two factors significantly increase the energetic penalty for incorrect base pairing in start codon recognition, thereby providing additional discrimination in the mRNA scanning process. The effect is similar to what has been seen for tRNA selection in the ribosomal A-site during decoding, where both the so-called monitoring bases and tRNA modifications boost the selectivity (24). Somewhat surprisingly, the largest effect of eIF1 and eIF1A is predicted for reading of the near-cognate GUG and AUA codons, with 1st and 3rd position mismatches, respectively, and for the AAG codon with a central A–A base pair. In these cases, the discrimination is increased by a factor of 300–5000 against incorrect codons due to the presence of the two factors. In contrast, the occurrence of an A–C mismatch in the 2nd position is found here to be unaffected by the initiation factors. Moreover, their effect on discrimination against the CUG codon is also rather small and CUG is predicted to have the smallest free energy penalty among the examined codons. These findings agree relatively well with *in vivo* and *in vitro* experiments on initiation efficiency and ternary complex binding to the mRNA programmed 40S subunit (11,32). The experiments show a rather uniform discrimination across the three codon positions, except for AAG and AGG where the *in vivo* measurements indicate very low initiation efficiency. The CUG codon is also the one with highest initiation efficiency, besides the cognate AUG. Some data also indicate that it may be $tRNA^{Leu}$ that is responsible for initiation at CUG (39). At any rate, the measured *in vivo* efficiency of CUG initiation of up to 10% of that of AUG (11,32) seems unusually high, as it would imply a very high initiation error frequency.

The mRNA scanning rate has been determined to ∼8 nucleotides per second and shows a linear dependence on the length of the 5′-UTR (21). Furthermore, these data indicate that the initiator tRNA scans non-cognate triplets with the same rate regardless of which codon is present in the P-site, which is consistent with the uniform relative binding free energies obtained from our calculations. Obviously, the probability of incorrect initiation increases the more codons the initiator tRNA has to scan. So, for long 5′-UTRs a large energetic penalty for each and every codon, except AUG, would be needed for a normal initiation to occur.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Knut and Alice Wallenberg Foundation; The Swedish Research Council (VR); The Swedish National Infrastructure for Computing (SNIC). Funding for open access charge: The Knut and Alice Wallenberg Foundation [KAW 2011.0081].
Conflict of interest statement. None declared.

REFERENCES

1. Hinnebusch,A.G. (2014) The scanning mechanism of eukaryotic translation initiation. *Annu. Rev. Biochem.*, **83**, 779–812.
2. Kapp,L.D. and Lorsch,J.R. (2004) The molecular mechanics of eukaryotic translation. *Annu. Rev. Biochem.*, **73**, 657–704.
3. Passmore,L.A., Schmeing,T.M., Maag,D., Applefield,D.J., Acker,M.G., Algire,M.A., Lorsch,J.R. and Ramakrishnan,V. (2007) The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell*, **26**, 41–50.
4. Saini,A.K., Nanda,J.S., Lorsch,J.R. and Hinnebusch,A.G. (2010) Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNA(i)(Met) binding to the ribosome. *Genes Dev.*, **24**, 97–110.
5. Pestova,T.V. and Kolupaeva,V.G. (2002) The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.*, **16**, 2906–2922.
6. Zhang,F., Saini,A.K., Shin,B.-S., Nanda,J. and Hinnebusch,A.G. (2015) Conformational changes in the P site and mRNA entry channel evoked by AUG recognition in yeast translation preinitiation complexes. *Nucleic Acids Res.*, **43**, 2293–2312.
7. Kozak,M. (1989) The scanning model for translation: an update. *J. Cell Biol.*, **108**, 229–241.
8. Maag,D., Fekete,C.A., Gryczynski,Z. and Lorsch,J.R. (2005) A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol. Cell*, **17**, 265–275.
9. Lomakin,I.B., Kolupaeva,V.G., Marintchev,A., Wagner,G. and Pestova,T.V. (2003) Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev.*, **17**, 2786–2797.
10. Cigan,A.M., Feng,L. and Donahue,T.F. (1988) tRNAi(met) functions in directing the scanning ribosome to the start site of translation. *Science*, **242**, 93–97.
11. Kolitz,S.E., Takacs,J.E. and Lorsch,J.R. (2009) Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA*, **15**, 138–152.
12. Pestova,T.V., Borukhov,S.I. and Hellen,C.U. (1998) Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature*, **394**, 854–859.

13. Fekete,C.A., Applefield,D.J., Blakely,S.A., Shirokikh,N., Pestova,T., Lorsch,J.R. and Hinnebusch,A.G. (2005) The eIF1A C-terminal domain promotes initiation complex assembly, scanning and AUG selection in vivo. *EMBO J.*, **24**, 3588–3601.
14. Maag,D., Algire,M.A. and Lorsch,J.R. (2006) Communication between eukaryotic translation initiation factors 5 and 1A within the ribosomal pre-initiation complex plays a role in start site selection. *J. Mol. Biol.*, **356**, 724–737.
15. Hashem,Y., des Georges,A., Dhote,V., Langlois,R., Liao,H.Y., Grassucci,R.A., Hellen,C.U.T., Pestova,T.V. and Frank,J. (2013) Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell*, **153**, 1108–1119.
16. Kozak,M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44**, 283–292.
17. Lomakin,I.B. and Steitz,T.A. (2013) The initiation of mammalian protein synthesis and mRNA scanning mechanism. *Nature*, **500**, 307–311.
18. Erzberger,J.P., Stengel,F., Pellarin,R., Zhang,S., Schaefer,T., Aylett,C.H.S., Cimermančič,P., Boehringer,D., Sali,A., Aebersold,R. *et al.* (2014) Molecular architecture of the 40S·eIF1·eIF3 translation initiation complex. *Cell*, **158**, 1123–1135.
19. Hussain,T., Llácer,J.L., Fernández,I.S., Munoz,A., Martin-Marcos,P., Savva,C.G., Lorsch,J.R., Hinnebusch,A.G. and Ramakrishnan,V. (2014) Structural changes enable start codon recognition by the eukaryotic translation initiation complex. *Cell*, **159**, 597–607.
20. Berthelot,K., Muldoon,M., Rajkowitsch,L., Hughes,J. and McCarthy,J.E.G. (2004) Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol. Microbiol.*, **51**, 987–1001.
21. Vassilenko,K.S., Alekhina,O.M., Dmitriev,S.E., Shatsky,I.N. and Spirin,A.S. (2011) Unidirectional constant rate motion of the ribosomal scanning particle during eukaryotic translation initiation. *Nucleic Acids Res.*, **39**, 5555–5567.
22. Marelius,J., Kolmodin,K., Feierberg,I. and Åqvist,J. (1998) Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graph. Model.*, **16**, 213–261.
23. Satpati,P., Bauer,P. and Åqvist,J. (2014) Energetic tuning by tRNA modifications ensures correct decoding of isoleucine and methionine on the ribosome. *Chem. Eur. J.*, **20**, 10271–10275.
24. Satpati,P., Sund,J. and Åqvist,J. (2014) Structure-based energetics of mRNA decoding on the ribosome. *Biochemistry*, **53**, 1714–1722.
25. King,G. and Warshel,A. (1989) A surface constrained all-atom solvent model for effective simulations of polar solutions. *J. Chem. Phys.*, **91**, 3647–3661.
26. Trobro,S. and Åqvist,J. (2007) A model for how ribosomal release factors induce peptidyl-tRNA cleavage in termination of protein synthesis. *Mol. Cell*, **27**, 758–766.
27. Lind,C., Sund,J. and Åqvist,J. (2013) Codon-reading specificities of mitochondrial release factors and translation termination at non-standard stop codons. *Nat. Commun.*, **4**, 2940.
28. Mackerell,A.D. Jr, Wiorkiewicz-Kuczera,J. and Karplus,M. (1995) An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.*, **117**, 11946–11975.
29. MacKerell,A.D., Bashford,D., Bellott,M., Dunbrack,R.L., Evanseck,J.D., Field,M.J., Fischer,S., Gao,J., Guo,H., Ha,S. *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
30. Lee,F.S. and Warshel,A. (1992) A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.*, **97**, 3100.
31. Brandsdal,B.O., Osterberg,F., Almlöf,M., Feierberg,I., Luzhkov,V.B. and Åqvist,J. (2003) Free energy calculations and ligand binding. *Adv. Protein Chem.*, **66**, 123–158.
32. Wei,J., Zhang,Y., Ivanov,I.P. and Sachs,M.S. (2013) The stringency of start codon selection in the filamentous fungus Neurospora crassa. *J. Biol. Chem.*, **288**, 9549–9562.
33. Johansson,M., Zhang,J. and Ehrenberg,M. (2012) Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 131–136.
34. Kozak,M. (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene*, **299**, 1–34.
35. Dvir,S., Velten,L., Sharon,E., Zeevi,D., Carey,L.B., Weinberger,A. and Segal,E. (2013) Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, E2792–E2801.
36. Pop,C., Rouskin,S., Ingolia,N.T., Han,L., Phizicky,E.M., Weissman,J.S. and Koller,D. (2014) Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Mol. Syst. Biol.*, **10**, 770–770.
37. Wei,C.L., MacMillan,S.E. and Hershey,J.W. (1995) Protein synthesis initiation factor eIF-1A is a moderately abundant RNA-binding protein. *J. Biol. Chem.*, **270**, 5764–5771.
38. Fekete,C.A., Mitchell,S.F., Cherkasova,V.A., Applefield,D., Algire,M.A., Maag,D., Saini,A.K., Lorsch,J.R. and Hinnebusch,A.G. (2007) N- and C-terminal residues of eIF1A have opposing effects on the fidelity of start codon selection. *EMBO J.*, **26**, 1602–1614.
39. Starck,S.R., Jiang,V., Pavon-Eternod,M., Prasad,S., McCarthy,B., Pan,T. and Shastri,N. (2012) Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science*, **336**, 1719–1723.