

ARTICLE

DOI: 10.1038/s41467-017-01757-0

OPEN

# Origin of the omnipotence of eukaryotic release factor 1

Christoffer Lind<sup>1</sup>, Ana Oliveira<sup>1</sup> & Johan Åqvist<sup>1</sup> 

Termination of protein synthesis on the ribosome requires that mRNA stop codons are recognized with high fidelity. This is achieved by specific release factor proteins that are very different in bacteria and eukaryotes. Hence, while there are two release factors with overlapping specificity in bacteria, the single omnipotent eRF1 release factor in eukaryotes is able to read all three stop codons. This is particularly remarkable as it is able to select three out of four combinations of purine bases in the last two codon positions. With recently determined 3D structures of eukaryotic termination complexes, it has become possible to explore the origin of eRF1 specificity by computer simulations. Here, we report molecular dynamics free energy calculations on these termination complexes, where relative eRF1 binding free energies to different cognate and near-cognate codons are evaluated. The simulations show a high and uniform discrimination against the near-cognate codons, that differ from the cognate ones by a single nucleotide, and reveal the structural mechanisms behind the precise decoding by eRF1.

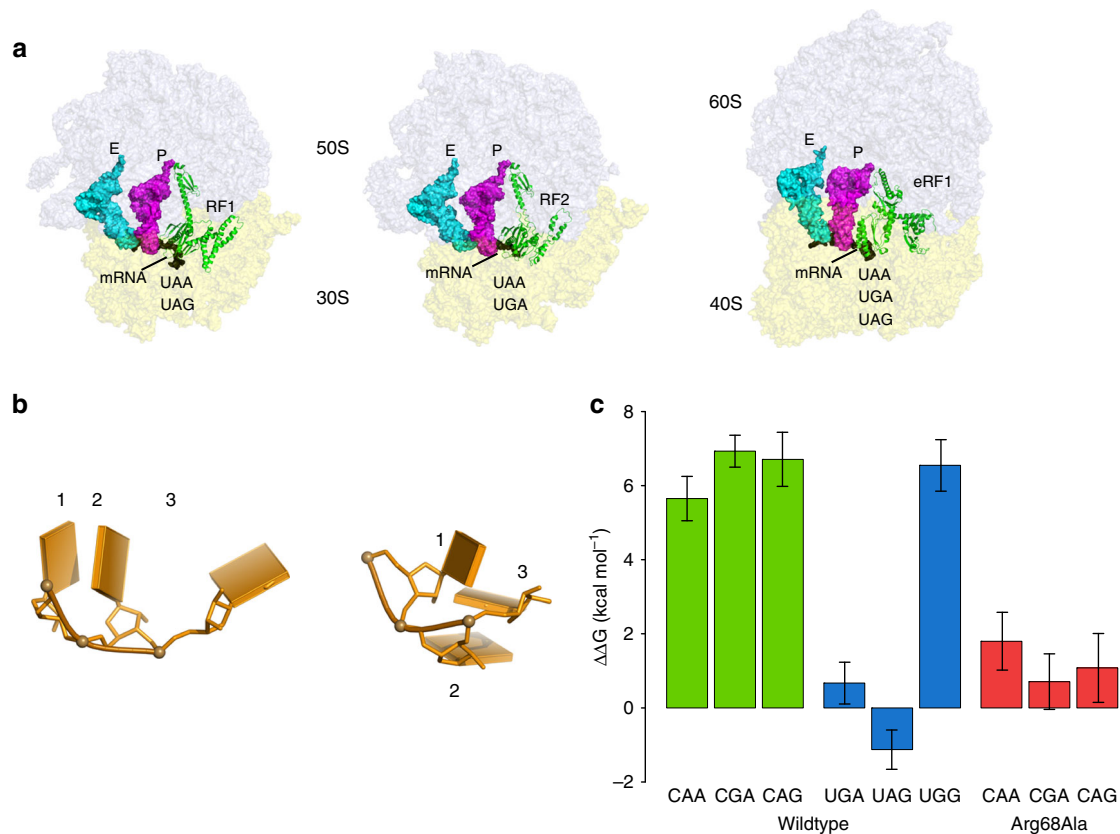
<sup>1</sup>Department of Cell & Molecular Biology, Uppsala University, Biomedical Center, Box 596, SE-751 24 Uppsala, Sweden. Correspondence and requests for materials should be addressed to J.Å. (email: [aqvist@xray.bmc.uu.se](mailto:aqvist@xray.bmc.uu.se))

A major difference in the protein synthesis machinery between bacteria and eukaryotes regard the termination of mRNA translation on the ribosome. While the three stop codons UAA, UAG, and UGA are largely conserved throughout the three kingdoms of life, albeit with notable exceptions<sup>1</sup>, the way they are decoded differs substantially. Hence, in bacteria two class-1 release factors (RFs) with overlapping specificity are generally required for stop codon recognition. Here, RF1 reads the UAA and UAG codons, while RF2 reads UAA and UGA. In eukaryotes and archaea, on the other hand, a single omnipotent RF is able to read all three stop codons<sup>2</sup>. This is quite remarkable since these RFs (eRF1 in eukaryotes and aRF1 in archaea) also apparently can discriminate against the tryptophan codon with high fidelity. That is, the omnipotent RF can recognize three combinations of purines in the second and third codon position (UAA, UAG, UGA), but avoid reading the fourth alternative (UGG) (Fig. 1). How this can be achieved from a structural viewpoint has remained a mystery, but recent cryo-EM structures<sup>3,4</sup> of eRF1 complexes with the ribosome now hint at a solution of the problem.

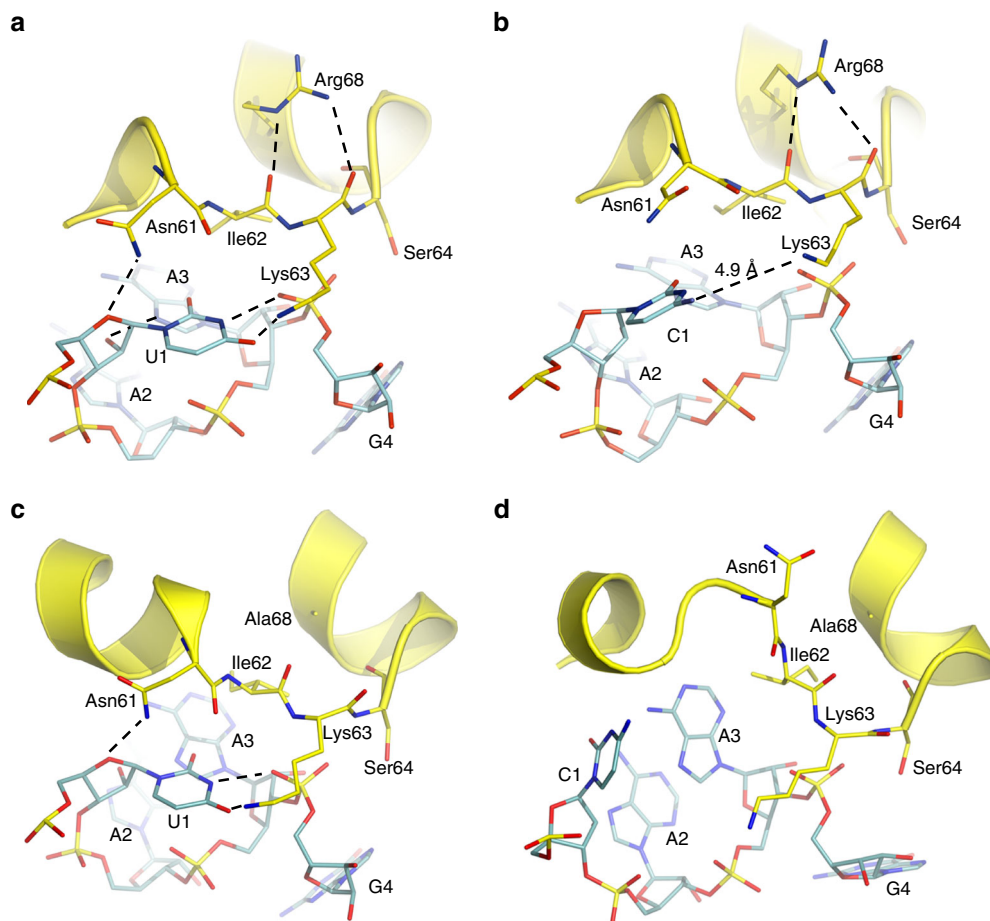
When a stop codon is presented at the decoding site, eRF1 binds to the ribosome in ternary complex with the class-2 release factor eRF3 and GTP. The latest cryo-EM structures<sup>5</sup> reveal that eRF1 first binds in its closed conformation, similar to what has been suggested also for bacterial RFs<sup>6,7</sup>. This is followed by GTP hydrolysis and dissociation of eRF3, whereupon eRF1 accommodates on the ribosome and inserts the universally conserved GGQ motif of its M domain into the A-site of the peptidyl-transferase center (PTC)<sup>5,8</sup>. It is the glutamine sidechain of this

motif that coordinates a catalytic water molecule in the PTC and promotes hydrolysis of the ester bond between the P-site transfer RNA and the nascent peptide<sup>9,10</sup>. It should also be noted here that the PTC is located some 75–80 Å away from the decoding site. The structures further show that the mRNA stop codon and the N-domain of eRF1, which is responsible for codon recognition, maintain the same overall conformation in the decoding site throughout the termination process. That is, the detailed structure of the eRF1–codon interaction is virtually identical in the pre-accommodated state and the final state where the M domain has accommodated into the ribosomal A-site<sup>5</sup>.

Although eRF1 and its bacterial counterparts RF1 and RF2 have the same biological function, they are structurally different and have little sequence homology, apart from the universally conserved GGQ motif which is common to all class-1 RFs<sup>11</sup>. Moreover, crystal structures of bacterial termination complexes revealed that the first two bases of the stop codon are stacked, while a histidine sidechain from the RF intercalates between these and the third base of the stop codon<sup>12–14</sup>. This histidine residue is also highly conserved in both RF1 and RF2 (His193 and His202, respectively; *Thermus thermophilus* numbering)<sup>12–14</sup>. However, the stop codon conformation adopted in eukaryotic ribosome complexes with eRF1 is distinctly different. Here, the stacked conformation of the first and second bases is lost and the codon instead presents an unexpected U-turn conformation<sup>3,4</sup> (Fig. 1). This conformation has a more compact shape and is clearly different from the standard stacked shape of sense codons. Before the emergence of 3D structural data, the codon specificity of eRF1 was biochemically mapped to highly conserved amino-acid motifs



**Fig. 1** Stop codon recognition in bacteria and eukaryotes. **a** View of bacterial (RF1 and RF2) and eukaryotic release factors (eRF1) bound to their respective ribosomes. The latter is omnipotent and reads all three stop codons. **b** The mRNA stop codon conformation is different in bacteria (left) and eukaryotes (right), where the latter is characterized by an mRNA U-turn. **c** Calculated binding free energy changes associated with single mutations from cognate to near-cognate stop codons (green and red bars for wildtype and mutant eRF1, respectively) and the corresponding binding free energies for the UGA, UAG, and UGG codon relative to UAA (blue bars). Error bars, 1 s.e.m. from at least 15 independent simulations



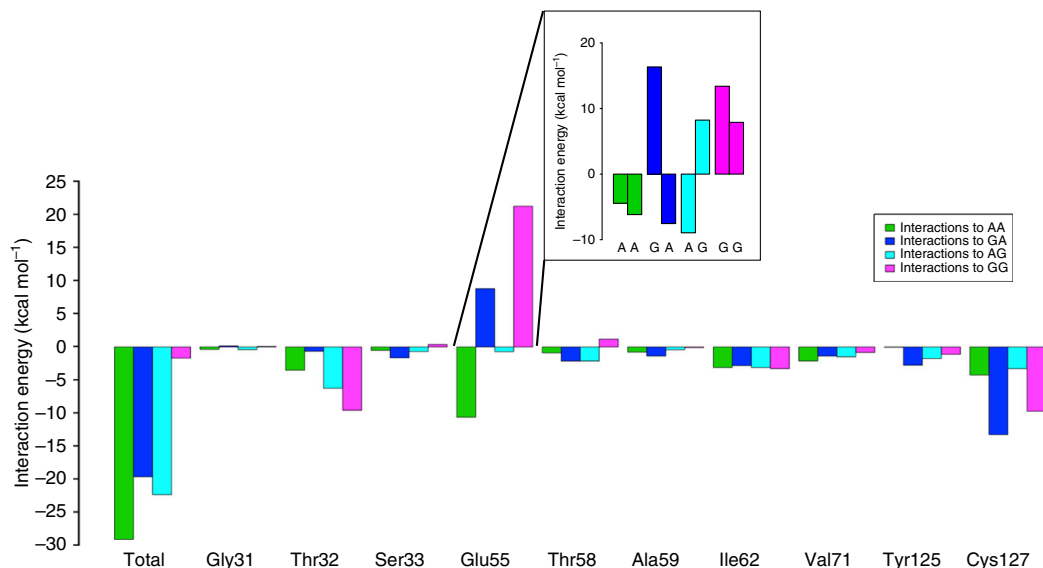
**Fig. 2** Reading of the invariant first position uracil. **a** Average MD structure showing the favorable interactions of the first position uracil (UAA codon) with Lys63 of the NIKS loop and an additional hydrogen bond to the phosphate backbone of the +4 nucleotide. **b** With a first position cytosine, the repulsion with Lys63 leads to loss of hydrogen bonding and an increased distance between the base and this eRF1 sidechain. **c** Average MD structure of the Arg68Ala mutation with the UAA codon bound, which is similar to the corresponding wild-type structure. **d** The non-cognate CAA codon is predicted to induce a sizeable conformational change of the NIKS loop in the Arg68Ala mutant, where the stabilizing arginine interaction with the NIKS backbone is missing

in its N-domain. Photochemical cross-linking experiments suggested that the TASNIKS motif (residues 58–64) is crucial to recognize the invariant first position uracil<sup>15</sup>, while the GTS loop (residues 31–33) and the YxCxxxF motif (residues 125–131) could be associated with reading of the second and third stop codon positions<sup>16–19</sup>. The highly conserved Glu55 has also been shown to be prominent for correct stop codon recognition<sup>17</sup>. In addition to the biochemical and mutagenesis studies that have attempted to determine the key residues for recognition at the different stop codon positions<sup>16,17,20–25</sup>, the ribosome-bound structures of eRF1<sup>3,4</sup> now clearly provide a breakthrough for understanding the eRF1-stop codon recognition.

The two structural studies of eRF1 complexes<sup>3,4</sup> both point to the importance of Lys63 for recognizing the first position uracil and also suggest that Glu55 is the key to discriminating against a double guanine pair in the last two codon positions. It was thus hypothesized that Glu55 could introduce unfavorable repulsions with a putative GG pair, thereby disturbing the hydrogen bonding network required for proper stop codon recognition<sup>3,4</sup>. However, while Beckmann and co-workers<sup>4</sup> emphasize a direct sidechain contact between Cys127 of the YxCxxxF motif and the second codon base, the Ramakrishnan group structure<sup>3</sup> rather suggests that the Cys127 main chain interaction with the rRNA base A1825 is important for stacking of the two last bases in the codon U-turn. The two structures further agree with each other regarding the interaction between the GTS loop and third

position nucleotide. This loop is described as flexible with the ability to adopt different conformations depending on which purine nucleotide is to be sensed<sup>3,19</sup>. In this context, it is also interesting to look at exceptions from the standard coding of termination. For example, in two ciliate species *Paramecium* and *Stylonychia* the release factor only exhibits UGA decoding specificity<sup>26</sup>, while UAA and UAG are sense codons<sup>27</sup>. *Paramecium tetraurelia* is of particular interest here since the highly conserved Glu55 and Thr58 residues are mutated to Asn and Glu, respectively, although Cys127 remains invariant.

While 3D structures of ribosome complexes yield invaluable information for formulating reasonable mechanistic hypotheses regarding function, it is often not so straightforward to actually validate or invalidate such proposals by experimental means. This also holds true for the above hypotheses for how eRF1 is able to achieve its peculiar codon specificity. Due to the low resolution of electron density for the codon binding site in the N-domain, a clear atomistic understanding of the structural basis of stop codon recognition is still lacking. For instance, the density for the hypothesized crucial Glu55 sidechain is completely missing<sup>3,4</sup>. In this work, we have computationally investigated the principles by which eRF1 recognizes each individual stop codon while effectively discriminating against the near-cognate sense codons CAA, CGA, CAG, and UGG. Here, computer simulations can play a very important role in bridging between structure and biochemistry, since the detailed energetics of molecular recognition



**Fig. 3** Interaction energies between the codon bases and their surroundings. Average non-bonded interaction energies (electrostatic and van der Waals) between the last two codon bases and key eRF1 residues around the codon are shown. Green, blue, cyan, and pink bars denote the UAA, UGA, UAG, and UGG codons, respectively. The critical interactions with Glu55 are further divided into their second and third base components in the inset, where the two bars of equal color denote interactions with these two bases, respectively, for each codon

processes can nowadays be evaluated with sufficiently high fidelity. That is, with a typical free energy calculation accuracy of about  $\pm 1$  kcal mol<sup>-1</sup>, cognate vs. non-cognate codon binding can be reliably quantified. Hence, we use molecular dynamics (MD)-based free energy calculations to explore the energetics and structural origin of stop codon recognition by eRF1. The relative eRF1 binding free energies for the different stop codons show that there is little variation in binding strength between these. In contrast, there are large energetic penalties associated with binding to different sense codons, including the tryptophan codon UGG. The simulations further allow identification of specific amino-acid interaction patterns depending on which codon is positioned in the decoding site.

## Results

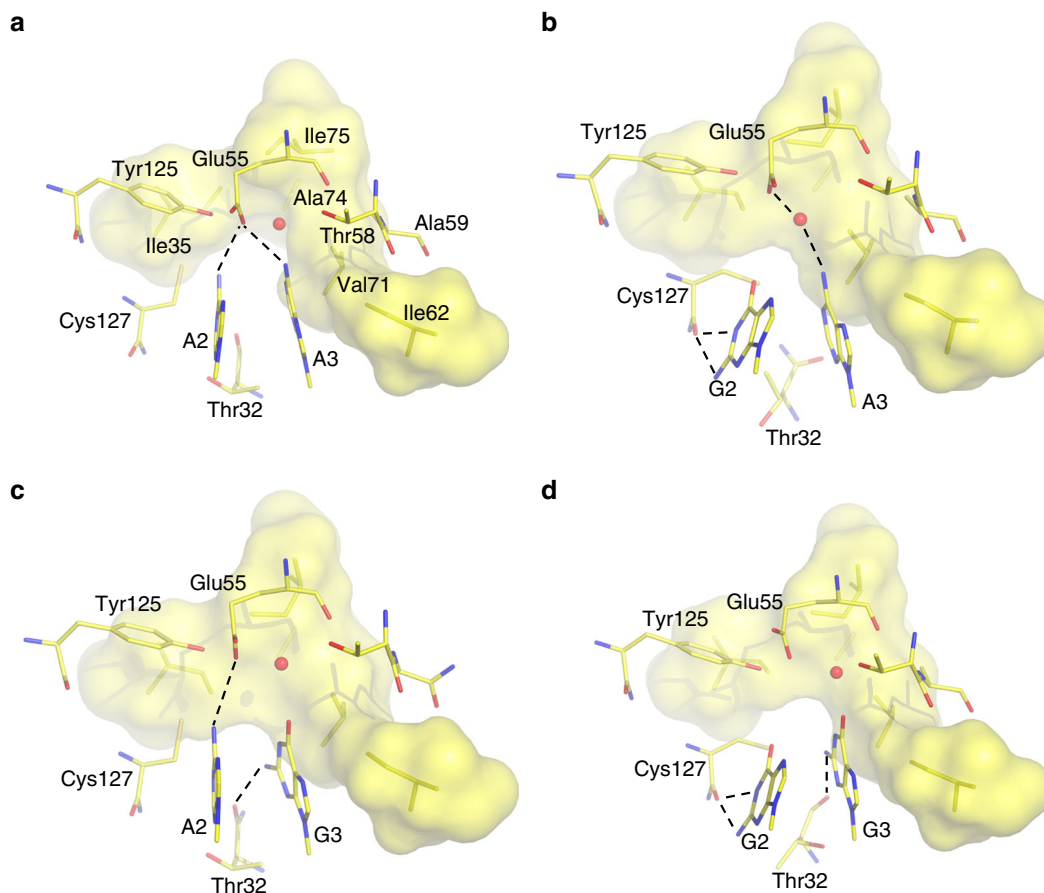
### Molecular dynamics simulations and free energy perturbation.

Free energy perturbation method (FEP) calculations<sup>28</sup> were performed utilizing the cryo-EM structures of the mammalian release factor (eRF1) complexes with the three stop codons UAA, UAG, and UGA (PDB accession numbers 3JAG, 3JAH, and 3JAI, respectively)<sup>3</sup>. To evaluate the relative free energies of binding to different codons single-nucleotide mutations were performed, starting from each of the different stop codons. These calculations were carried out both with and without eRF1 bound to the ribosome, which allows the relative binding free energies to be evaluated by a standard thermodynamic cycle<sup>29</sup>. All calculations were repeated 15 times with different initial velocities in the MD simulations.

**Reading of the first codon position.** Uracil occupies the first nucleotide position in all standard stop codons. The specificity for this nucleotide should thus be a requirement for initiating the termination process. To evaluate the energetic penalty against the seemingly near-cognate CAA, CAG and CGA (glutamine and arginine) codons, we mutated the first position in all three stop codons in the free energy simulations. The results clearly show a large and uniform discrimination against these three codons. The binding free energies of eRF1 to the three sense codons CAA, CAG and CGA, relative to the stop codons differing by a single

mutation, are thus all predicted to be about 6 kcal mol<sup>-1</sup> (Fig. 1c). This corresponds to more than a factor of  $10^4$  in terms of binding affinity, demonstrating the high specificity of eRF1 for the first position uracil. The high bias toward uracil reading (Fig. 2a) is indeed found to be largely caused by the repulsive interactions between the Lys63 sidechain and the amine of the cytosine base in the sense codons (Fig. 2b), as indicated by the cryo-EM structures<sup>3,4</sup>. Consequently, the cytosine becomes dislocated from the compact U-turn structure observed with cognate stop codons. The cryo-EM structures differ with regard to the rotamer of the cognate uracil base and in all our MD simulations the nucleobase initially rotates to the conformation observed in ref. <sup>4</sup>, indicating that this is the preferred rotamer.

As had been suggested by cross-linking experiments<sup>15</sup>, and now supported by the structural data<sup>3,4</sup>, the NIKS loop is in close proximity to the first codon position. Here, the O4 atom interacts with the amine group of Lys63 of the NIKS motif and N3 makes an additional hydrogen bond to the mRNA phosphate group at position +4 (the first position U is denoted +1). This conformation allows the first position U to satisfy two hydrogen bonds as in a U–A Watson–Crick pair with donor–acceptor distances of 3.0 and 2.9 Å, respectively (Fig. 2a). Similar interactions involving N3 and O4 hydrogen bonds were also found in earlier work on both bacterial and mitochondrial termination complexes<sup>29,30</sup>. The second O2 oxygen atom, which is present in both uracil and the cytosine, shows no interaction with eRF1 in the MD simulations. The Asn61 sidechain of eRF1 further stabilizes the stop codon conformation by interacting with the sugar base of the first nucleotide. Moreover, the 2'-oxygen of U + 1 hydrogen bonds to N7 of the third codon base, which again stabilizes the U-turn, but neither of these two interactions contribute to discrimination since they do not involve the first position nucleobase itself. In fact, the NIKS loop appears very stable and rigid during the MD simulations, both with sense and stop codons bound, which presumably contributes to its role in codon discrimination. This is evidenced by the evolution of its backbone root mean square deviation (RMSD) from the corresponding experimental starting structures during the free energy simulations. The NIKS loop RMSD thus remains below 0.6 Å during all of the FEP mutations



**Fig. 4** Structural basis of second and third position stop codon reading. Average MD structures of eRF1 bound to the **a** UAA, **b** UGA, **c** UAG, and **d** UGG codons. The hydrophobic cluster of residues (Ile35, Ile62, Val71, Ala74, Ile75, and Cys127) is shown as a transparent surface and key water molecules are depicted as red spheres. Key hydrogen bonds involved in base recognition are indicated as dashed lines

and the average over all replicate simulations of the eRF1 complexes (135 in total) is only 0.38 Å, indicating that the loop is not able to remodel its conformation in response to these codon changes. Furthermore, the highly conserved Arg68 residue<sup>31</sup> appears crucial for maintaining the NIKS loop conformation and two hydrogen bonds between its sidechain and the backbone of Ile62 and Lys63 could be identified (Fig. 2). This finding agrees well with the loss of termination efficiency observed experimentally for the Arg68Ala mutation<sup>25</sup>. To further investigate the NIKS loop stability and the role of Arg68 in stabilizing the NIKS loop, we mutated Arg68 to alanine and performed free energy calculations on the ability of this eRF1 mutant to discriminate against a cytosine in the first codon position. These calculations predict a remarkable loss of fidelity, where the penalty for reading a cytosine now drops to 1–2 kcal mol<sup>-1</sup> (Fig. 1c). Examination of the average structure from these simulations further reveals that the NIKS loop undergoes large conformational changes with non-cognate codon bound, as a response to the missing backbone interactions with Arg68 (Fig. 2c, d).

**Codon recognition at the second and third positions.** To evaluate the binding specificity at the last two codon positions, we first carried out free energy perturbation calculations of the synonymous stop codon mutations from UAA to UGA and UAG and vice versa. These mutations show that eRF1 has little or no preference for any of the three stop codons, in agreement with kinetic data for human eRF1<sup>27</sup> (Fig. 1). The calculated binding free energies of eRF1 to the UGA and UAG stop codons relative to UAA are thus both near-zero,  $\Delta\Delta G_{\text{bind}} = 0.7$  and

–1.1 kcal mol<sup>-1</sup> for UGA and UAG, respectively (Fig. 1). In contrast, the important and non-trivial feature of a high discrimination against the tryptophan codon UGG is clearly seen from the simulations. Thus, the energetic penalty against reading UGG is predicted by the calculations to be very high,  $\Delta\Delta G_{\text{bind}} = 6.6$  kcal mol<sup>-1</sup> relative to UAA. Hence, the picture emerging from these calculations is a high and uniform penalty of about 6 kcal mol<sup>-1</sup> for all the four non-cognate codons examined. This is in agreement with kinetic measurements in a yeast termination assay, where more than a 10<sup>3</sup>-fold difference in  $k_{\text{cat}}/K_{\text{M}}$  for peptide release on the UAA and UGG codons is observed (Prof. M. Ehrenberg, personal communication).

To investigate the proposal<sup>3,4</sup> that the negatively charged Glu55 sidechain is responsible for discriminating solely against UGG among the four UNN codons, we examined the behavior of Glu55 during the MD simulations. This residue turns out to be remarkably stable and essentially maintains the same conformation regardless of which codon is present (Supplementary Fig. 1). However, the interaction energetics between eRF1 and the last two codon positions turns out to be very informative. We evaluated the average pairwise non-bonded interaction energies between these codon positions and key eRF1 residues that have been implicated from mutagenesis experiments to be crucial for correct codon reading<sup>16,19,22,25</sup> (Fig. 3). While the total non-polar contribution to these interaction energies is essentially constant among all four calculated codons, it is the polar interactions that give distinct differences, which explain the origin of the eRF1 omnipotence in stop codon reading. Importantly, it can be seen that the three cognate codons all have large favorable (negative)

total interaction energies with the 10 eRF1 residues surrounding the second and third codon bases. The arrangement of these residues surrounding the codon is shown in Fig. 4. However, when the UGG codon is bound the overall interaction is considerably less favorable (Fig. 3). This effect is indeed entirely due to electrostatic repulsion between the Glu55 sidechain and the double guanine pair in UGG. In the case of a UAA codon, the AA interaction with Glu55 is about 10 kcal mol<sup>-1</sup> negative, with about equal attractive contribution from the two bases. In the case of UAG, the Glu55 attraction with the second position is more or less balanced by its repulsion with the third. However, the Glu55 repulsion is about twice as strong with a G in the second position compared to the third, making its overall interaction with UGA unfavorable. The situation with UGA is, however, saved by the fact that Cys127 has a distinctly favorable backbone interaction with a second position G, which more than compensates for the unfavorable Glu55 interaction. Since Glu55 strongly prefers A's in both of the last codon positions, the guanine penalties add up to a large repulsion with UGG and, in this case, the Cys127 interaction is not enough to compensate the energetic penalty. Hence, the distinct discrimination against UGG can, at least qualitatively, be explained in terms of a fine-tuned energetic interplay predominantly between Glu55 and Cys127. We say qualitatively, since the energy components discussed are only enthalpic contributions and entropic effects associated with these interactions are not included in Fig. 3.

The residue Cys127 is also highly conserved in eRF1 and aRF1 and has been suggested from experiments to be pivotal for correct codon recognition<sup>22,25</sup>. Our MD simulations show that this residue, or rather its backbone, is particularly selective for a second position guanine, with which it makes double hydrogen bonds (Fig. 4b, d). The strong conservation of Cys127 is likely due to its bulky sulfur atom being sequestered between the codon and the also highly conserved residues Ile35 and Tyr125. This appears to enforce a backbone conformation at the tip of  $\beta$ -strand 4, with the Cys127 carbonyl group in appropriate orientation for sensing the second nucleotide position. Its carbonyl has the capability of hydrogen bonding with both the N1 and N2 atoms of guanine and these interactions are lost with an adenine in the second codon position, which is also reflected by the interaction energetics in Fig. 3. Besides its direct interaction with second codon position, the Cys127 sidechain forms a hydrophobic cluster together with Ile35, Ile62 (of the NIKS motif), Val71, Ala74, and Ile75 around the third codon base (Fig. 4). This cluster is dispersed along the eRF1 amino-acid sequence and has therefore not been identified as a unique motif, but it is clear that it forms a distinct structural motif that is highly conserved among the eRF1 sequences<sup>31</sup>. Hence, both Ile62 and Cys127 would perhaps be better described as part of this hydrophobic motif than of the NIKS and YxCxxxF motifs to which they are usually assigned. Due to the packing of this sidechain cluster around the edge and (downstream) face of the third base, it prevents entry of excessive solvent that could otherwise compensate for missing hydrogen bonds. Thus, the role of the hydrophobic cluster clearly appears to be partial desolvation of the third codon position. It is therefore likely that all of the cluster members (Ile35, Ile62, Val71, Ala74, Ile75, and Cys127) make small hydrophobic contributions to the codon selectivity. In order to test this hypothesis, we mutated Cys127 to glycine and repeated the FEP calculations of the UAA  $\rightarrow$  UGG codon mutation. These calculations predict the Cys127Gly mutation to reduce the penalty against UGG from 6.6 to 3.9 kcal mol<sup>-1</sup>, with accompanying water penetration, which shows that the hydrophobic cluster indeed is critical. As far as the YxCxxxF motif is concerned, the hydroxyl group of Tyr125 interacts with Glu55, which assists in stabilizing the hydrogen bond with the amine

group in a second position adenine. The conserved GTS loop does not provide any large contributions to codon binding, with the exception of a hydrogen bond between the Thr32 backbone carbonyl and a third position G. This favorable interaction is also seen from the energetics in Fig. 3.

## Discussion

While the common stop codons UAA, UGA, and UAG are used in the genetic codes of most species, the mechanisms for recognizing them are diverse. In bacteria, two RFs with overlapping specificities are required to terminate mRNA translation and yield the correct protein product. In eukaryotes, on the other hand, the single eRF1 release factor is able to read all three stop codons. Numerous studies of eRF1 mutants have investigated the principles behind its omnipotence, or triple codon reading ability, but it is only with the emergence of 3D structures of termination complexes that quantitative structure–function relationships can be established.

Our molecular dynamics free energy perturbation calculations show that eRF1 has a common pattern of binding to all three stop codons and that this reflects a very specific interplay of amino acids within the codon reading domain. The simulations show that the first position uracil consistently has its O4 and N3 hydrogen bonds satisfied, just as in a standard A–U base pair, and that it is the rotated base conformation<sup>4</sup> that permits this. Hence, in addition to Lys63 it is the +4 phosphate group of the mRNA that holds the uracil in place in its special U-turn conformation (Fig. 2). The presence of a charged residue at position 63 indeed appears to be crucial for efficient recognition of the first stop codon position, as the Lys63Arg mutation is the only one observed experimentally at this position to retain termination efficiency<sup>25</sup>. It should, however, be noted that Lys63 becomes post-translationally hydroxylated at the delta carbon (C4). This modification has been shown to increase the efficiency of eRF1 termination, albeit by less than a factor of two<sup>32</sup>. The modification was thus not modeled here as it is not present in either of the cryo-EM structures used, but it seems likely that the role of it is to further stabilize the Lys63 conformation for interaction with the mRNA. Furthermore, we predict the conformation and stability of the NIKS loop to be affected by the highly conserved Arg68 residue that interacts with the loop backbone. Removal of these interactions by the Arg68Ala mutation is predicted to render the NIKS loop more flexible and to yield a substantial drop in discrimination against a non-cognate first position cytosine.

The first position uracil can thus be said to be read separately from the last two nucleotides and the principal challenge for eRF1 is to efficiently read only the correct combinations of these nucleotides from the four different possibilities. There is thus a notable difference from how bacterial release factors read the stop codons, in which case each nucleotide is recognized individually<sup>29</sup>. That is, the exclusion of the tryptophan UGG codon arises from the strict specificity of RF1 and RF2 for an adenine in the second and third positions, respectively. By forcing the mRNA U-turn conformation, eRF1 will essentially read the last two purine nucleotides as a block unit rather than individually. The hypothesis from the cryo-EM structures<sup>3,4</sup> is that the negatively charged Glu55 would prevent reading of a double guanine pair due to electrostatic repulsion with the two O6 oxygens, although electron density for its sidechain is missing. This indeed turns out to be true as judged from our simulation energetics, but the fine-tuning behind this phenomenon is quite remarkable. That is, while the favorable Glu55 interactions with second and third position adenines are similar, the repulsion with guanine is considerably stronger in the second position as the guanine dipole moment in that case is oriented antiparallel to the Glu55 electric

field, while that of a third position guanine is more perpendicular (Fig. 4d). The favorable interaction between Cys127 and a second position guanine partly compensates for this repulsion but, since the interactions are more or less additive, it is only the GG pair that results in a large net discrimination (Fig. 3). The story is, however, not that simple since solvent molecules could potentially easily compensate for unsatisfied hydrogen bonds and here the conserved hydrophobic sidechain cluster (Ile35, Ile62, Val71, Ala74, Ile75, and Cys127) apparently protects, particularly the second base, against such interactions. Hence, just as in the case of normal mRNA decoding, where the monitoring bases (A1492 and A1493 in bacteria) have been shown to boost discrimination against non-cognate codons by desolvation<sup>33</sup>, we see the same effect here exerted by the hydrophobic cluster.

Taken together, our MD free energy calculations show that the eRF1 discrimination against the near-cognate sense codons CAA, CGA, CAG, and UGG is large ( $\sim 6$  kcal mol<sup>-1</sup>) and uniform. The simulations further show that eRF1 binding to the three stop codons is also characterized by a uniform binding affinity with little energetic preference for any particular codon. The interplay between different amino acids of eRF1 is clearly important for correct codon reading and is what underlies its ability to read three out of four purine–purine combinations. Hence, the binding is largely monitored by two key residues that are universally conserved, Glu55 providing repulsion and Cys127 providing attraction to the critical stop codon guanines. Interestingly, the calculations also predict that eRF1 has a larger discrimination against sense codons than the bacterial counterparts<sup>29</sup>, which may reflect an overall higher degree of evolutionary optimization of its structure.

## Methods

**Molecular dynamics simulations.** MD simulations and FEP<sup>28</sup> were performed using the cryo-EM structures of the mammalian release complex programmed with the three stop codons UAA, UAG, and UGA (PDB accession numbers 3JAG, 3JAH, and 3JAI, respectively)<sup>3</sup>. The MD simulations utilized spherical boundary conditions with a 30 Å radius sphere centered on the N6 atom of the third position nucleotide of the UAA-bound structure, which is the geometrical center of the stop codon in its U-turn conformation. The system was solvated with TIP3P water molecules and solvent molecules close to the sphere boundary were subjected to radial and polarization restraints according to the SCAAS model<sup>34</sup>. Atoms located further than 22 Å away from the sphere center were harmonically restrained to their initial coordinates with a 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> force constant. By scaling down distant RNA phosphate charges and neutralizing charged amino acids close to the boundary, the system was kept overall neutral, as described elsewhere<sup>35</sup>. The system was stepwise heated up to the final temperature of 310 K, initially using a 1 fs time step, which was increased to 2 fs during the end of the relaxation and subsequently used for the data collection phase of the free energy calculations. A direct cutoff of 10 Å was employed for all non-bonded interactions except for the atoms directly involved FEP transformations. Electrostatic interactions beyond the 10 Å cutoff were treated with the local reaction field multipole expansion method<sup>36</sup>.

**Free energy calculations.** To calculate the relative free energy of binding of eRF1 to different codons, we used the FEP method to introduce mutations from the cognate stop codons to either near-cognate sense codons or to a synonymous stop codon. From the 3JAG structure with UAA-bound transformations to the CAA, UGA and UAG codons were made. Similarly, the 3JAH complex with UAG was used for codon transformation to CAG, UAA and UGG and the 3JAI structure with UGA bound for codon mutations to CGA, UAA and UGG. To obtain the relative eRF1 binding free energies to different codons, the simulations were carried out with and without the bound release factor in accordance with a standard thermodynamic cycle<sup>37</sup>. All FEP calculations used 51 different sampling windows and the calculations were repeated 15 times, yielding a minimum of 30 ns of data collection per transformation. Each repeat was initiated using randomized initial conditions generated from the Maxwell–Boltzmann velocity distribution. All MD simulations and free energy calculations were performed using the CHARMM36<sup>38,39</sup> force field implemented to the Q molecular dynamics software<sup>40</sup>. Since experimental structures were available for all three cognate stop codons, transformations between these were carried out in both the forward and reverse directions, starting from the relevant cryo-EM structure. Standard errors of the mean are very similar for mutations carried out only in the forward direction compared to those for which the reverse path could also be initiated from an experimental structure (Fig. 1c). Mutations to the UGG codon were started from

both UAG and UGA by transforming the second or third adenine, respectively, to guanine and the binding free energy relative to UAA was calculated by averaging the results from the independent two thermodynamic cycles.

**Code availability.** Source code for the Q program<sup>40</sup> is available at <https://github.com/qusers/Q6>, or upon request from the corresponding author.

**Data availability.** The data that support the findings of this study are available from the corresponding author upon request.

Received: 23 May 2017 Accepted: 11 October 2017

Published online: 10 November 2017

## References

- Swart, E. C., Serra, V., Petroni, G. & Nowacki, M. Genetic codes with no dedicated stop codon: context-dependent translation Termination. *Cell* **166**, 691–702 (2016).
- Frolova, L. et al. A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. *Nature* **372**, 701–703 (1994).
- Brown, A., Shao, S., Murray, J., Hegde, R. S. & Ramakrishnan, V. Structural basis for stop codon recognition in eukaryotes. *Nature* **524**, 493–496 (2015).
- Matheis, S., Berninghausen, O., Becker, T. & Beckmann, R. Structure of a human translation termination complex. *Nucleic Acids Res.* **43**, 8615–8626 (2015).
- Shao, S. et al. Decoding mammalian ribosome-mRNA states by translational GTPase complexes. *Cell* **167**, 1229–1240.e15 (2016).
- Demo, G. et al. Mechanism of ribosome rescue by ArfA and RF2. *Elife* **6**, e14874 (2017).
- Trapp, K. & Joseph, S. Ribosome induces a closed to open conformational change in release factor 1. *J. Mol. Biol.* **428**, 1333–1344 (2016).
- Salas-Marco, J. & Bedwell, D. M. GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Mol. Cell Biol.* **24**, 7769–7778 (2004).
- Trobro, S. & Åqvist, J. A model for how ribosomal release factors induce peptidyl-tRNA cleavage in termination of protein synthesis. *Mol. Cell* **27**, 758–766 (2007).
- Kazemi, M., Himo, F. & Åqvist, J. Peptide release on the ribosome involves substrate-assisted base catalysis. *ACS Catal.* **6**, 8432–8439 (2016).
- Kisselev, L. Polypeptide release factors in prokaryotes and eukaryotes: same function, different structure. *Structure* **10**, 8–9 (2002).
- Laurberg, M. et al. Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852–857 (2008).
- Weixlbaumer, A. et al. Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**, 953–956 (2008).
- Korostelev, A. et al. Crystal structure of a translation termination complex formed with release factor RF2. *Proc. Natl Acad. Sci. USA* **105**, 19684–19689 (2008).
- Chavatte, L., Seit-Nebi, A., Dubovaya, V. & Favre, A. The invariant uridine of stop codons contacts the conserved NIKSR loop of human eRF1 in the ribosome. *EMBO J.* **21**, 5302–5311 (2002).
- Conard, S. E. et al. Identification of eRF1 residues that play critical and complementary roles in stop codon recognition. *RNA* **18**, 1210–1221 (2012).
- Kolosov, P. et al. Invariant amino acids essential for decoding function of polypeptide release factor eRF1. *Nucleic Acids Res.* **33**, 6418–6425 (2005).
- Bulygin, K. N. et al. Adenine and guanine recognition of stop codon is mediated by different N domain conformations of translation termination factor eRF1. *Nucleic Acids Res.* **39**, 7134–7146 (2011).
- Wong, L. E., Li, Y., Pillay, S., Frolova, L. & Pervushin, K. Selectivity of stop codon recognition in translation termination is modulated by multiple conformations of GTS loop in eRF1. *Nucleic Acids Res.* **40**, 5751–5765 (2012).
- Bertram, G., Bell, H. A., Ritchie, D. W., Fullerton, G. & Stansfield, I. Terminating eukaryote translation: domain 1 of release factor eRF1 functions in stop codon recognition. *RNA* **6**, 1236–1247 (2000).
- Frolova, L., Seit-Nebi, A. & Kisselev, L. Highly conserved NIKS tetrapeptide is functionally essential in eukaryotic translation termination factor eRF1. *RNA* **8**, 129–136 (2002).
- Seit-Nebi, A., Frolova, L. & Kisselev, L. Conversion of omnipotent translation termination factor eRF1 into ciliate-like UGA-only unipotent eRF1. *EMBO Rep.* **3**, 881–886 (2002).
- Fan-Minogue, H. et al. Distinct eRF3 requirements suggest alternate eRF1 conformations mediate peptide release during eukaryotic translation termination. *Mol. Cell* **30**, 599–609 (2008).
- Cheng, Z. et al. Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev.* **23**, 1106–1118 (2009).

25. Kryuchkova, P. et al. Two-step model of stop codon recognition by eukaryotic release factor eRF1. *Nucleic Acids Res.* **41**, 4573–4586 (2013).
26. Lekomtsev, S. et al. Different modes of stop codon restriction by the *Styloynchia* and *Paramecium* eRF1 translation termination factors. *Proc. Natl Acad. Sci. USA* **104**, 10824–10829 (2007).
27. Eliseev, B., Kryuchkova, P., Alkalaeva, E. & Frolova, L. A single amino acid change of translation termination factor eRF1 switches between bipotent and omnipotent stop-codon specificity. *Nucleic Acids Res.* **39**, 599–608 (2011).
28. Brandsdal, B. O. et al. Free energy calculations and ligand binding. *Adv. Protein Chem.* **66**, 123–158 (2003).
29. Sund, J., And er, M. &  qvist, J. Principles of stop-codon reading on the ribosome. *Nature* **465**, 947–950 (2010).
30. Lind, C., Sund, J. &  qvist, J. Codon-reading specificities of mitochondrial release factors and translation termination at non-standard stop codons. *Nat. Commun.* **4**, 2940 (2013).
31. Liang, H., Wong, J. Y., Bao, Q., Cavalcanti, A. R. O. & Landweber, L. F. Decoding the decoding region: analysis of eukaryotic release factor (eRF1) stop codon-binding residues. *J. Mol. Evol.* **60**, 337–344 (2005).
32. Feng, T. et al. Optimal translational termination requires C4 lysyl hydroxylation of eRF1. *Mol. Cell* **53**, 645–654 (2014).
33. Satpati, P. &  qvist, J. Why base tautomerization does not cause errors in mRNA decoding on the ribosome. *Nucleic Acids Res.* **42**, 12876–12884 (2014).
34. King, G. & Warshel, A. A surface constrained all-atom solvent model for effective simulations of polar solutions. *J. Chem. Phys.* **91**, 3647–3661 (1989).
35. Trobro, S. &  qvist, J. Mechanism of peptide bond synthesis on the ribosome. *Proc. Natl Acad. Sci. USA* **102**, 12395–12400 (2005).
36. Lee, F. S. & Warshel, A. A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations. *J. Chem. Phys.* **97**, 3100 (1992).
37. Lind, C. &  qvist, J. Principles of start codon recognition in eukaryotic translation initiation. *Nucleic Acids Res.* **44**, 8425–8432 (2016).
38. Denning, E. J., Priyakumar, U. D., Nilsson, L. & Mackerell, A. D. Jr. Impact of 2'-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **32**, 1929–1943 (2011).
39. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles. *J. Chem. Theory Comput.* <https://doi.org/10.1021/ct300400x> (2012).
40. Marelus, J., Kolmodin, K., Feierberg, I. &  qvist, J. Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graph. Model.* **16**, 213–225, 261 (1998).

## Acknowledgements

The support from the Swedish Research Council (VR), the Knut and Alice Wallenberg Foundation, and the Swedish National Infrastructure for Computing (SNIC) is greatly acknowledged.

## Author contributions

J.  designed the experiments. C.L. performed the experiments. C.L., A.O. and J.  analyzed the data and wrote the paper.

## Additional information

**Supplementary Information** accompanies this paper at doi:10.1038/s41467-017-01757-0.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

  The Author(s) 2017