

SCIENTIFIC REPORTS

OPEN

Reinitiation at non-canonical start codons leads to leak expression when incorporating unnatural amino acids

Received: 26 February 2015

Accepted: 05 June 2015

Published: 08 July 2015

Tanja Kalstrup & Rikard Blunck

With the rapid development of a continuously growing selection of unnatural amino acids (UAAs), UAA insertion becomes increasingly popular for investigating proteins. However, it can prove problematic to ensure the homogeneity of the expressed proteins, when homogeneity is compromised by “leak expression”. Here, we show that leak expression may be mediated by reinitiation and can result in unwanted proteins when stop codons for UAA insertion are mutated into the N-terminus of proteins. We demonstrate that up to 25% of leak expression occurs through reinitiation in the Shaker-Kv channel when stop codons are located within the first 70 amino acids. Several non-canonical start codons were identified as translation reinitiation sites, and by removing the start codons, we were able to decrease leak expression to less than 1%. Our study emphasizes the need to carefully inspect for leak expression when inserting UAAs and demonstrates how leak expression can be eliminated.

Genetic incorporation of unnatural amino acids (UAA) is a powerful tool to investigate protein structure and function. Through nonsense suppression of stop codons, UAAs are incorporated into proteins using chemically-charged tRNAs or orthogonal tRNA/tRNA synthetase (tRNA/RS) pairs¹. With this method, numerous UAAs with different chemical and biophysical properties have been used as probes in *E. coli*, yeast, plant, amphibians and mammalian cells^{1,2}. When using UAAs, it is crucial to have control of the protein translation process in order to get reliable results. Indeed, the orthogonal tRNA/RS pair provides site-specific incorporation of the UAA, with no or only a negligible fraction of unwanted proteins lacking the UAA, which may result from crosstalk with endogenous tRNAs or synthetases.

Also cellular mechanisms can generate leak expression when using UAA mutagenesis, in spite of the orthogonality of the tRNA/RS pair. Eukaryotic cellular mRNA is mainly translated via a linear scanning model which consists of initiation, elongation and termination. During initiation, the 40S ribosomal subunit is equipped with initiation factors and Met-tRNA, and binds to the 5'-terminal mRNA 7-methylguanosine cap structure (Fig. 1A). The complex then scans in the 5' to 3' direction in a base-by-base inspection until it encounters the first ATG start codon, where elongation begins with recruitment of the large 60S subunit and release of initiation factors³ (Fig. 1A). When the ribosome encounters a stop codon during elongation, translation is terminated. The ribosome disassembles from the mRNA and prepares for a second round of translation.

However, under some circumstances, translation does not strictly follow the linear scanning model. One example is stop codon *readthrough*, where a tRNA inserts its amino acid at the stop codon⁴, or the stop codon can be ignored due to ribosomal sliding⁵. The efficiency of stop codon readthrough includes complex combinations of stop codon identity, sequence context and RNA structure^{6–8}, and such recoding strategies are especially used by viruses to generate a rich and varied protein synthesis. Therefore,

Groupe d'Études des Protéines Membranaires (GÉPROM), Departments of Physics and of Physiology, Université de Montréal, Montréal, QC, Canada. Correspondence and requests for materials should be addressed to R.B. (email: rikard.blunck@umontreal.ca)

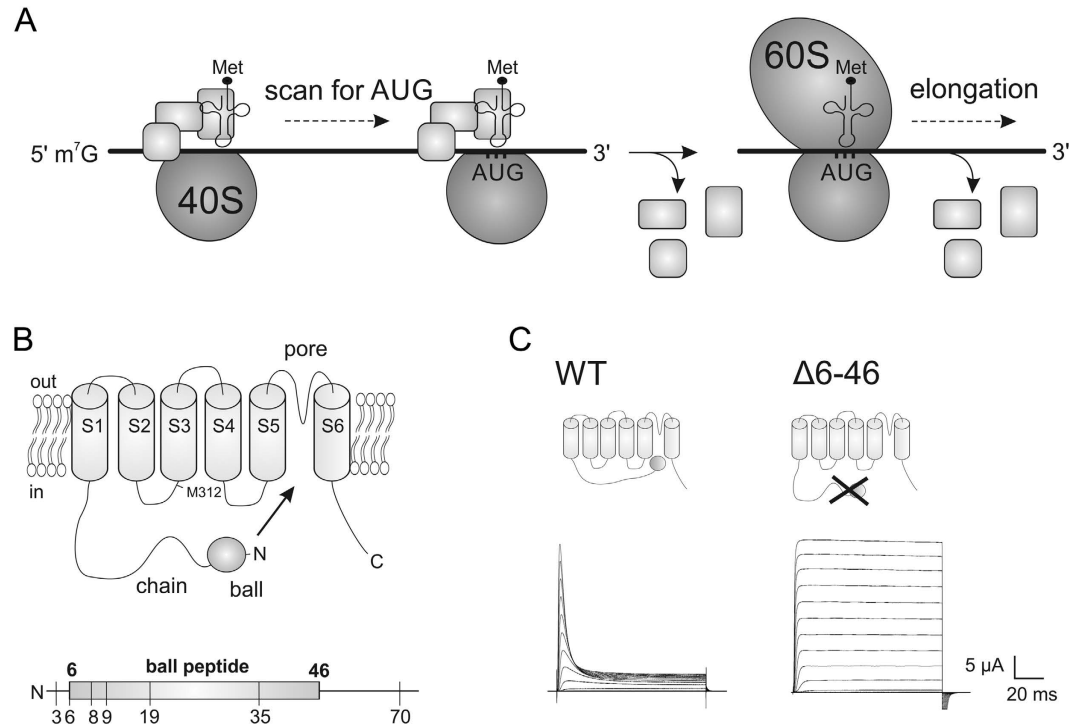


Figure 1. **A**) Simplified model for translation initiation in eukaryotes. First, the 40S small ribosomal subunit associates with initiation factors (shown as squares) and Met-tRNA^{Met} on the mRNA 5' cap structure. The assembled complex then scans the mRNA sequence until it meets the first AUG codon, after which a fraction of initiation factors are released and the 60S large ribosomal subunit is recruited. Elongation of peptide bonds can then begin and the rest of the initiation factors are released. For a more detailed view see Jackson, *et al.*¹⁵. **B**) Topology of the Shaker Kv channel (top) and the N-terminal (below) with annotated positions used in this study. **C**) Ionic currents from WT and $\Delta 6-46$ Shaker channels upon depolarization from -100 mV to $+60$ mV in steps of $+10$ mV from a holding potential of -90 mV. The corresponding structures of each constructs are shown.

translational stop codon readthrough is best understood in viruses, but has also been observed in yeast, *Drosophila* and humans^{9,10}. Another recoding strategy is *reinitiation*, where multiple initiation sites can result in mRNAs harboring more than one open reading frame (ORF)¹¹. If the eukaryotic ribosome encounters a stop codon shortly after initiation, a fraction of the initiation factors are still associated to the ribosome^{12,13}. This allows the 40S subunit to remain on the mRNA and as a result, the 40S subunit continues scanning for a downstream start codon and reinitiates translation, resulting in a downstream ORF (dORF). Reinitiation only occurs when the distance between the first start codon and the next stop codon is short^{13,14}, likely increasing the probability that initiation factors are still present. As a result, N-terminal truncated proteins can be translated from the upstream ORF (uORF). In eukaryotes, the ribosomal initiation complex involves 13 initiation factors (eIFs)¹⁵ but the identity and number of eIFs required for reinitiation is unknown. Reinitiation occurs in viruses¹⁶ and throughout the eukaryotic kingdom¹⁷⁻¹⁹. Reinitiation also occurs in prokaryotes, but the mechanism is different due to a distinct initiation machinery¹⁷.

In contrast to readthrough, reinitiation is more common in eukaryotes, and approximately 40% of mammalian mRNAs have a short uORF, although not all of them necessarily involve translation of both ORFs^{20,21}. Examples of functional reinitiation at downstream methionines include regulation of Aquaporin-4 isoforms²² and LQT2 mutations in hERG channels leading to altered gating properties²³ and defective trafficking²⁴.

Previously, we successfully incorporated a fluorescent UAA into the Shaker voltage-gated potassium (Kv) channel². The nonsense stop codons were located either central or C-terminal, and no leak expression was detected in the absence of the UAA, reflecting a strong orthogonality of the tRNA/RS pair used. Here, we show that when stop codons are introduced in the region near the N-terminus of the same Kv channel, a considerable amount of leak expression results from translation reinitiation which do not reinitiate at downstream methionines, but rather uses non-canonical (non-AUG) start codons.

The Kv channel is a homotetramer where each subunit consists of six transmembrane segments (S1-S6) containing the voltage sensor (S1-S4) and the pore (S5-S6). The N-terminal region (amino acids 1-46) of each subunit is a cytosolic soluble "ball-and-chain" region responsible for fast N-type

inactivation (Fig. 1B)^{25,26}, where the N-terminus move into the pore region and occludes the pore in response to sustained depolarization (Fig. 1C). When the N-terminus of a Shaker Kv channel is deleted ($\Delta 6-46$), the inactivation is disrupted such that the channel remains open (Fig. 1C)²⁵. The different current phenotypes of full length and N-terminal deletion channels allowed us to determine the absence or presence of the N-terminus in various Shaker mutants expressed in *Xenopus* oocytes using the cut-open voltage clamp technique²⁷.

Results

Functional channels are translated despite stop codons in the N-terminus. In contrast to our previous findings where we introduced stop codons in the central and C-terminal part of the protein², leak expression was present when stop codons were inserted within the first 70 amino acids in the Shaker Kv channel (A3stop, G6stop, Y8stop, G9stop, K19stop, E35stop and D70stop, Fig. 1B, Fig. 2A). No leak expression was measured for any stop codons tested downstream of residue D70 (L170stop, F196stop and V234stop shown in Fig. 2A, and A359stop and H486stop in Kalstrup & Blunck 2013²), which indicates that the mechanism responsible for leak expression is specific for N-terminal positions. D70stop expressed less than 1% whereas the other mutants expressed between 5 and 25% (Fig. 2C).

The leak expression presented in Fig. 2A occurs in the presence of the channel RNA only, and the mechanism leading to expression here therefore cannot be attributed to crosstalk between the modified tRNA/synthetase pair and endogenous amino acids. The observed leak expression could be due to readthrough or translational reinitiation beyond the inserted stop codon. One indication was the dependence of leak expression on the position of the stop codon. When stop codons are inserted proximal to the N-terminus, a short uORF is generated in the case of reinitiation. Due to the short length of the uORF, the likelihood is high that (some of) the initiation factors are still attached, allowing a second initiation to take place at the next start codon. The further downstream a stop codon is located, the longer the translated uORF will be and the higher will be the chance that all initiation factors separated from the ribosome, such that reinitiation cannot occur. Readthrough, on the other hand, is not dependent on the presence of initiation factors and should have no dependence on the relative position. We only found leak expression for N-terminal stop codons, pointing towards reinitiation as the prevalent mechanism for leak expression.

We were able to further corroborate this indication by investigating expression of the uORF. In the Shaker Kv channel, used in this study, we were also able to determine whether the uORF has been expressed because the N-terminus forms the ball peptide responsible for N-type inactivation and its presence is visible in the functional data. Deletion of residues 6–46 removes inactivation (Fig. 1C) while synthetic N-terminal Shaker peptides of the first 20 amino acids can restore inactivation in $\Delta 6-46$ channels²⁸.

In the case of reinitiation, an N-terminal peptide and an N-terminal truncated channel are translated. Depending on the length of the peptide and where the second translation begins, such N-terminal truncation could lead to disrupted N-type inactivation. E35stop and D70stop exhibited weak and complete N-type inactivation, respectively, while the other mutants did not undergo any N-type inactivation and were identical to $\Delta 6-46$ channels (Figs 1C and 2A,D). It seems that the uORFs generated in E35stop and D70stop were responsible for the inactivation of truncated channels, while the uORFs in the other mutants were too short to inactivate (<19 amino acids, Fig. 1B). We elaborate on this in the next section.

Principally, inactivation in E35stop and D70stop could also be due to stop codon readthrough, where full length channels are translated, and thus undergo N-type inactivation. However, readthrough would also have led the mutants with a shorter uORF to inactivate (Fig. 2A). In addition, removal of the start codon (M1stop) did not reduce leak expression (Fig. 2B–C), indicating that ribosomes are indeed capable of initiating translation downstream in the sequence. Taken together, these results indicate that the most likely mechanism causing leak expression here is reinitiation. The question remained at what site translation was reinitiated.

One possibility would be initiation at a downstream methionine; it was for instance previously reported that Kv1.4 channels with a stop codon at residue 19 exhibited leak currents from truncated channels which were eliminated by removal of a methionine at residue 108²⁹. However, here, removal of the next downstream methionine (M312L) in Shaker did not significantly influence leak expression (Fig. 2B). We ignored later methionines since the resulting translated protein would not yield functional channels, lacking most of or the entire voltage sensor. Reinitiation thus has to occur at sites other than the standard start codon AUG.

Non-canonical start codons are used as translation reinitiation sites. Increasing evidence from the literature has underscored the use of non-canonical start codons which deviate from the canonical AUG codon by one nucleotide, except AAG and AGG^{30–32}. We therefore screened the Shaker mRNA sequence for such non-canonical codons and found 9 in-frame codons within the first 121 amino acids at position I40, L45, L47, L63, L66, I89, I101, T114 and T121 (highlighted in bold in Fig. 3A). To verify that these codons were used as start codons in reinitiation, silent mutations were introduced into the E35stop background. First, we mutated L47 (#1 in Fig. 3A), then both L47 and L45 (#2) and then L47, L45 and I40 (#3). Each mutation led to a decreased expression level in a stepwise manner (Fig. 3B). Therefore, all ribosomes do not necessarily reinitiate at the first codon they encounter (I40). For the subsequent mutations of L63 (#4), L66 (#5) and I89 (#6), only #5 led to a decreased expression. When the last three

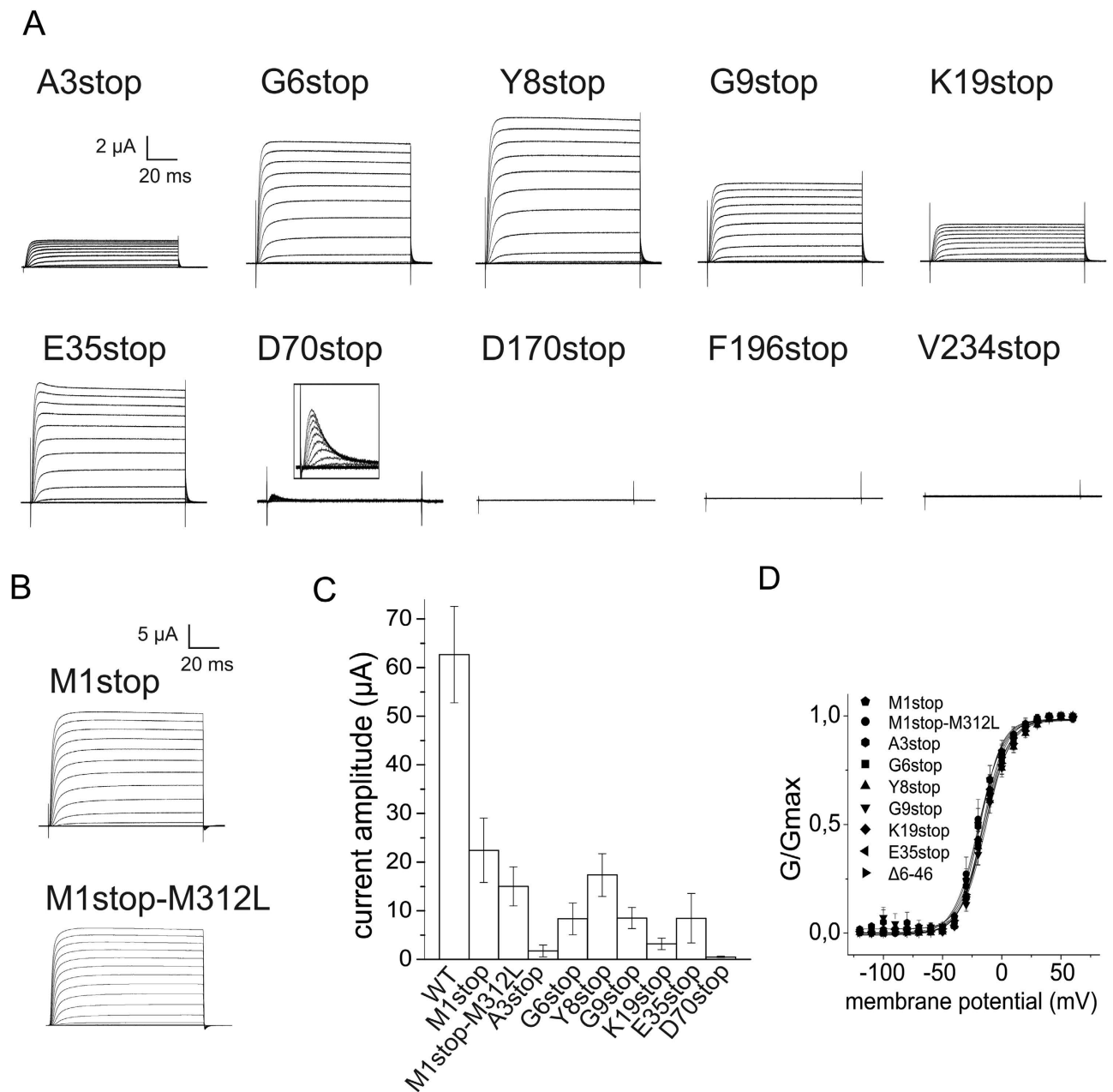


Figure 2. A–B) Ionic currents from Shaker mutants obtained with the same protocol as in 1C. C) Comparison of peak current amplitudes at +60 mV. Error bars indicate SEM with $n=10$ –20 oocytes. D) Conductance-voltage (GV) relationships were fitted to a Boltzmann equation ($V_{1/2, M1stop} = -19.1$ mV, $dV_{M1stop} = 9.9$ mV, $V_{1/2, M1stop-M312L} = -23.0$ mV, $dV_{M1stop-M312L} = 11.3$ mV, $V_{1/2, A3stop} = -18.1$ mV, $dV_{A3stop} = 9.1$ mV, $V_{1/2, G6stop} = -15.5$ mV, $dV_{G6stop} = 9.7$ mV, $V_{1/2, Y8stop} = -14.6$ mV, $dV_{Y8stop} = 10.8$ mV, $V_{1/2, G9stop} = -13.0$ mV, $dV_{G9stop} = 10.0$ mV, $V_{1/2, K19stop} = -16.3$ mV, $dV_{K19stop} = 9.3$ mV, $V_{1/2, E35stop} = -14.5$ mV, $dV_{E35stop} = 10.1$ mV, $V_{1/2, \Delta 6-46} = -15.8$ mV, $dV_{\Delta 6-46} = 10.6$ mV).

codons were removed in addition (#9), expression levels further decreased. Expression could partially be restored upon reinsertion of I40 into the #9 background (#9-1) confirming that reinitiation is not limited to the first start codon. As a control, the unnatural amino acid Anap was incorporated in response to the stop codon using the orthogonal tRNA/Anap-RS pair (pAnap)^{2,33,34}. With Anap, wildtype expression levels and N-type inactivation were fully recovered, reflecting full length channels (Fig. 3B–C). The expression levels with Anap did not decrease with the removal of start codons, excluding the possibility that RNA stability might play a role in the decreased expression. If the RNA had become unstable by silently removing the non-canonical start codons, also expression in presence of Anap would have been

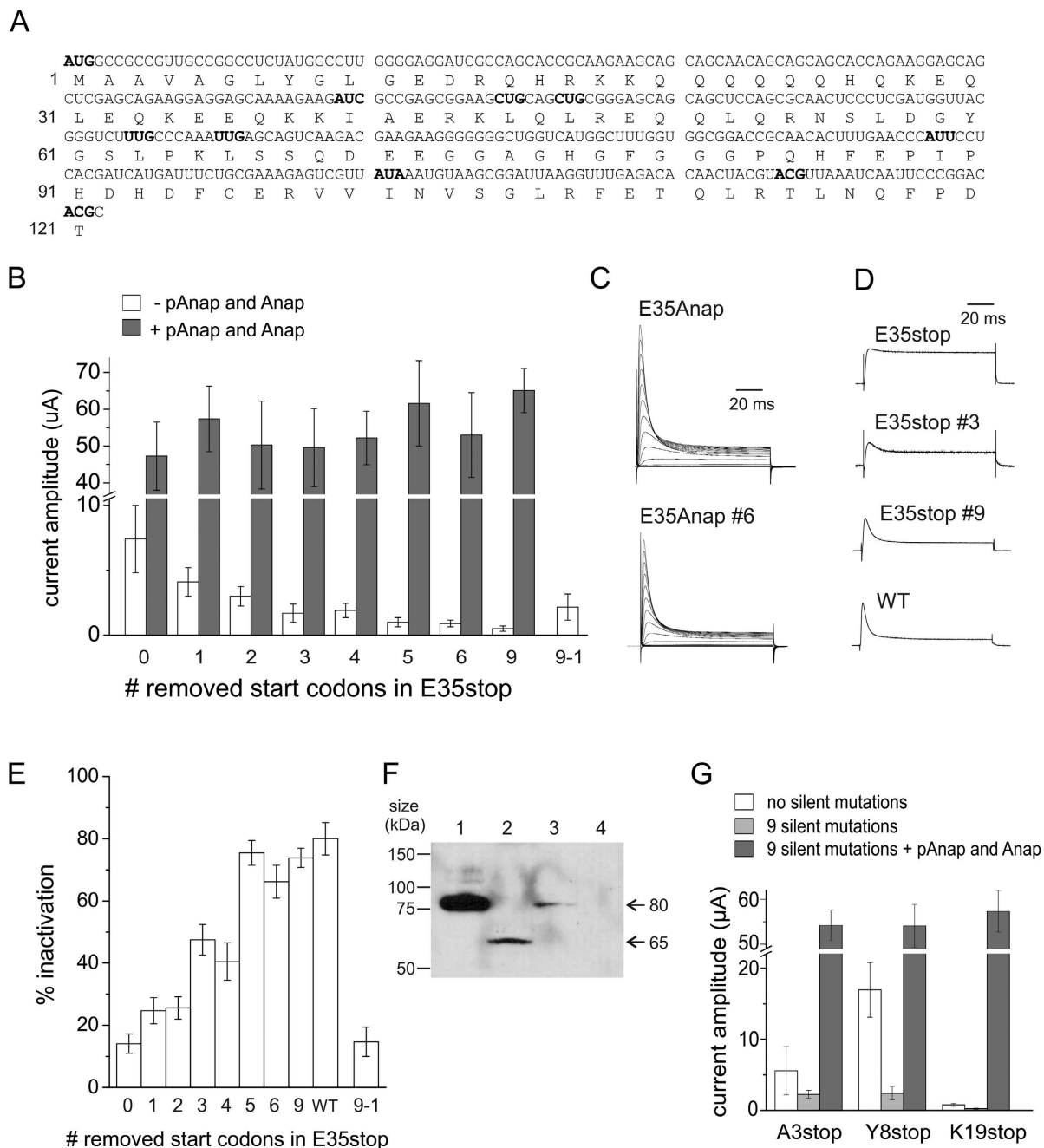


Figure 3. **A**) Sequence of amino acids 1–121 and the corresponding mRNA nucleotides in Shaker. The nucleotides of the canonical start codon (M1) and the following 9 downstream non-canonical start codons (I40, L45, L47, I89, L36, L66, I101, T114, and T121) are shown in bold. AAG and AGG codons were ignored. Numbers refer to the amino acid sequence. **B**) Comparison of peak current amplitudes at +60 mV of E35stop with silent mutations, without and with Anap incorporation. Numbers refer to the number of mutated start codons such that 1 is E35stop-L45L, 2 is E35stop-L45L-L47L, 3 is E35stop-I40L-L45L-L47L, 4 is E35stop-I40L-L45L-L47L-L63L, 5 is E35stop-I40L-L45L-L47L-L63L-L66L, 6 is E35stop-I40L-L45L-L47L-L63L-L66L-I89L and 9 is E35stop-I40L-L45L-L47L-L63L-L66L-I89L-I101L-T114T-T121T. Error bars indicate SEM with $n = 20$ –30 oocytes. **C**) Inactivation is recovered with Anap incorporation. Shown are examples of ionic currents from E35Anap and E35Anap #6 (six mutated start codons). Voltage protocols are the same as in Fig. 1. **D**) Comparison of inactivation measured as the ratio between the current amplitude at the end of the +60 mV pulse, and the maximum current amplitude. Error bars indicate SEM with $n = 10$ –20 oocytes. **E**) Depolarization from -90 mV to +60 mV is shown for E35stop mutants and WT showing how inactivation increases when channel expression decreases. **F**) Western blot of isolated *Xenopus* oocytes membranes expressing Shaker WT (lane 1, 30 oocytes), E35stop (lane 2, 70 oocytes), E35Anap (lane 3, 30 oocytes) and non injected (lane 4, 70 oocytes). **G**) Comparison of peak current amplitudes at +60 mV for A3stop, Y8stop and K19stop, and removal of 9 start codons without and with Anap incorporation. Error bars indicate SEM with $n = 5$ –10 oocytes.

reduced. Therefore, our experiments demonstrate that several non-canonical start codons in the region of amino acids 40–121 are recognized by the ribosome to reinitiate translation in E35stop channels.

The sequence context of start codons has significant influence on the initiation efficiency of both canonical and non-canonical start codons^{35,36}. The Kozak consensus sequence gccRccAUGG is the optimal context in eukaryotes, with the nucleotides in capital letters at position -3 and $+4$ (relative to AUG) being most important. The purine (R = A or G) at -3 is more important than a G at $+4$, and the other nucleotides only play a role if there is neither a purine at -3 or a G at $+4$ ³⁷. For the 9 codons tested here, L47, L63 and I89 have the weakest Kozak sequence (Fig. 3A), which could explain why L63 and I89 are not used as initiation sites. The codon for L47 (CTG) is one of the most probable non-AUG codons³⁰ and may explain why it is recognized by the ribosome despite a weak Kozak sequence.

When we reduce the translation efficiency of reinitiated channels, only the dORF is affected, while the short uORF is not. Therefore, the ratio of free N-terminal peptides to truncated channels will become larger when expression of the truncated channels decreases. This would explain why the slight inactivation of E35stop channels increases and almost reaches WT levels when the alternative start codons are removed (Fig. 3D–E). These findings are supported by the fact that the efficiency of N-type inactivation indeed depends on both the length and the concentration of the N-terminal peptide²⁸. We estimated the intracellular concentration of the uORF (amino acids 1–34) to be at least $2\mu\text{M}$ (see Methods), which is in agreement with previous peptide concentrations, where application of free peptides (amino acids 1–20) to $\Delta 6-46$ Shaker channels resulted in partial and full inactivation at concentrations of $10\mu\text{M}$ and $100\mu\text{M}$, respectively²⁸. Alternatively, the residual expression could be caused by readthrough, which would also show complete inactivation.

To confirm protein truncation in E35stop, a Western blot was performed for direct visualization of protein lengths with C-terminal myc-tagged Shaker constructs (Fig. 3F). Full-length wildtype Shaker exists in core-glycosylated ($\sim 80\text{kDa}$) and mature ($\sim 110\text{kDa}$) form^{38,39}. For wildtype and E35Anap, we mainly see the core-glycosylated band at 80kDa since the mature form only appears after longer expression times. E35stop in contrast, shows one band at $\sim 65\text{kDa}$. At 65kDa , we can exclude the possibility that E35stop leads to inefficient glycosylation as unglycosylated Shaker channels migrate at $\sim 75\text{kDa}$ ³⁸.

The expected truncated proteins can be ordered into two groups: Those which reinitiate at I40, L45, L47 or L66 would theoretically be $\sim 5-7\text{kDa}$ shorter than wildtype. Those which reinitiate at I101, T114 or T121 would be $\sim 11-14\text{kDa}$ shorter than wildtype. We only observed a single band at 65kDa , despite loading 10 times more of E35stop on the SDS-PAGE gel compared to wildtype (not shown). The band at 65kDa agrees with core-glycosylated truncated channels, reinitiating at I101, T114 and T121. From reinserting only the first non-canonical start codon, we know, however, that also the first start codons are used. It is possible that low expression combined with different constructs resulted in protein concentrations too low to detect on a western blot. Alternatively, the small differences in length may result in truncated versions to migrate collectively.

To verify that reinitiation resulted in leak expression in the other mutants, we removed the same 9 alternative start codons in A3stop, Y8stop and K19stop. The expression levels decreased for all three mutants when removing the alternative start codons in the same manner as for E35stop (Fig. 3F). In addition, Anap incorporation rescued successfully the full length channels with WT expression levels in presence of the 9 silent mutations, confirming that reinitiation is a general mechanism causing leak expression.

Discussion

In this study, we have identified translation events which require attention when UAAs are incorporated specifically in the N-terminal region of a protein. The efficient *Xenopus* oocyte expression system and the high sensitivity of electrophysiological recordings allowed us to measure low expression levels ($<5\%$) which typically are difficult to resolve without direct access to protein function. Our system also allowed to efficiently discriminate between truncated and full-length proteins.

Based on the expression levels in Fig. 2C, we estimate that reinitiated translations can generate 5–25% leak expression when the length of the uORF is between 3 and 35 codons. Since *cis*-acting mRNA sequences and RNA structure can affect reinitiation efficiency^{13,40}, the lengths mentioned here cannot be exactly extrapolated to other proteins and remain approximations. Our results are consistent with studies of the length and time dependence of reinitiation at downstream methionines, where expression levels decreased to 30% in the presence of a 13-codon uORF¹³, or to 50% in the presence of a 28-codon uORF¹⁴. The same studies showed that reinitiation was optimal for a certain uORF length, and reduced gradually when the uORF was lengthened further, most likely due to loss of initiation factors. Examples include reduction in reinitiation when a 13-codon uORF was lengthened to 33 codons¹³, or a complete inhibition when a 24-codon uORF was lengthened to 40 codons¹⁴. These reported magnitudes of reinitiation inhibition by uORF lengths also correlate with our results, where reinitiation decreased to less than 1% when the uORF was lengthened to 70 codons (Fig. 2C). Our results thus indicate that the nature of the reinitiation site (canonical versus non-canonical) does not play a major role.

Leak expression was reduced from 14% to less than 1% in E35stop channels when 9 non-canonical start codons between codon 40 and 121 were removed. Although the remaining leak expression was very subtle, it means nevertheless that the 40S ribosomal subunit is capable of scanning at least 258 nt (the length in nucleotides between residue 35 and 121) for a downstream non-canonical start codon.

Comparison with scanning lengths for reinitiation at methionines shows similar results with 63 nt and 129 nt in hERG^{23,24} and 270 nt in Kv1.4 channels²⁹.

Kozak showed that reinitiation does not happen when the distance between the stop codon and the next start codon is short (ie. 8 nt) and that reinitiation is not limited to the closest downstream ATG codon¹¹, because the 40S subunit only gradually becomes capable of reinitiation. This is confirmed by our results for non-canonical start codons, which show that the first start codon (I40) is only used by a fraction of the ribosomes, while others continue scanning to reinitiate at later start codons.

Our results let us conclude that the established conditions for reinitiation at methionines also apply to non-canonical start codons. Taken together, the likeliness for reinitiation is balanced between the length of the uORF (optimal length is 2–50 codons), the scanning length between the stop and the next start codon (between 8 and 300 nt) and the Kozak sequence context. During the first initiation process, the 40S subunit scans base-by-base for a start codon. It is therefore likely that during scanning for reinitiation, the 40S subunit also scans base-by-base and thus has no memory of the codon-frame of the uORF. As a result, it is not unlikely that out-of-frame start codons can serve as reinitiation sites. In that case, the translated dORF would be non-functional and thus be of no concern for UAA experiments. In the quadruplet codon technique for UAA insertion⁴¹, a frame-shift occurs during termination, but since the 40S subunit does not scan in codons, but in a base-by-base fashion, reinitiation is likely equally prone to occur.

In cases where full length and N-terminal truncated proteins are functionally identical, it can be difficult to differentiate between stop codon readthrough and reinitiation when leak expression exists in the absence of both UAA and the tRNA/RS pair. With the present study, we emphasize the importance of translation reinitiation when using UAA mutagenesis. Since reinitiation only occurs when the uORF is short, the issue is particularly relevant when UAAs are incorporated near the N-terminus of the protein of interest. Furthermore, we underline that attention should not be limited to downstream methionines, but that non-canonical start codons also serve as reinitiation sites.

Our results provide evidence that when modifying the genetic code for incorporation of UAAs, other cellular mechanisms may be affected as well. The high level of regulation at the translation initiation step is sensitive to minimal changes at the 5' end of the mRNA sequence in terms of codons, sequence context and secondary structure. As a result, all these factors must be considered when inserting UAAs into the first 100 amino acids of any protein and especially in proteins where the function is not as readily visible as in ion channels.

It is possible that the efficiency of reinitiation-mediated leak expression is lower in the presence of UAAs than in their absence, and as a consequence low enough to be neglected. Nevertheless, in order to avoid N-terminal truncated proteins and proteins lacking the UAA to co-express with the UAA containing protein – in particular in multimeric proteins, – caution should be exercised against reinitiation, and removal of downstream start codons may decrease or eliminate such leak expression.

Methods

Molecular Biology and Oocyte Injection. All mutations were introduced into the *Drosophila* Shaker H4 gene in the pBSTA vector⁴². The tRNA/AnapRS synthetase pair were encoded on a cDNA (pAnap) with CMV promoter for the synthetase²; kind gift of P.G. Schultz, Scripps Research Institute. *Xenopus laevis* oocytes were injected with 35 ng of *in vitro* transcribed RNA and incubated for 2 days at 18°C in Barth solution. For incorporation with Anap, 9.2 nl of 0.1 ng/nL cDNA encoding the tRNA/AnapRS pair² was injected into the oocyte nucleus 6–24 hours prior to coinjection of 23 nL 1 mM Anap and 35 ng RNA. For leucine and threonine silencing, CTC and ACT codons were used, respectively. Isoleucines were mutated to leucines using CTC, since all isoleucine codons are potential start codons. The codon used for the M312L mutation was CTC.

Electrophysiology. Voltage clamp was performed with a CA-1B amplifier (Dagan). Currents were recorded in the cut-open oocyte voltage-clamp configuration as described⁴³ and analyzed by using GPatch (Department of Anesthesiology, University of California, Los Angeles). Linear capacitance currents were subtracted online using the P/4 protocol⁴⁴. The external solution contained in mM: 5 KOH, 110 NMDG, 10 Hepes, and 2 Ca(OH)₂, and the internal solution contained in mM: 115 KOH, 10 Hepes, and 2 EDTA. Both solutions were adjusted to pH 7.1 with MES. Conductance (G) was calculated from the steady state currents (I) via $G = I/(V - V_{rev})$, where V_{rev} is the reversal potential. Conductance-voltage relationships (GV) were fitted to a Boltzmann relation of the form $G/G_{max} = (1 + \exp(-(V - V_{1/2})/dV))^{-1}$.

Estimation of the intracellular inactivation peptide concentration. To estimate a lower limit for the concentration of the intracellular inactivation peptides, we determined how many of the original constructs (Shaker-W434F) express under identical conditions. The N-terminal peptide should, as a short cytosolic peptide, reach at least the same concentration. The number of intracellular uORFs (the number of N-termini) per channel equals the number of subunits. The number of subunits n , expressed at the surface, can be calculated using 3.3 elementary charges per subunit^{45–47} ($n = \frac{Q}{3.3e}$), where Q is the total gating charge (26 nC) obtained with cut-open voltage clamp after 2 days of RNA incubation with Shaker-W434F. Since only part of the oocyte is clamped in cut-open voltage clamp, the ratio of whole

oocyte surface to clamped spherical cap surface was calculated using an oocyte radius of 0.5 mm and an angle of 30° from the center of the oocyte to the spherical cap edge. The final concentration of N-termini is then given by equation 1, where V is the oocyte volume (0.52 μl) as calculated from the radius.

$$[\text{peptide}] = n/N_A \cdot \frac{\text{area}_{\text{oocyte}}}{\text{area}_{\text{cap}}} \cdot \frac{1}{V} \quad (1)$$

Western blot. For detection of Shaker, we generated myc-tagged Shaker constructs with Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu inserted in the COOH terminus (after Val650). Oocytes were injected with myc-Shaker RNA and incubated at 18°C for 4 days. Membrane isolations were obtained as previously described with minor modifications⁴⁸ 30–70 oocytes were homogenized in 1 ml HEDP buffer (100 mM HEPES, 1 mM EDTA, pH 7.6 with NaOH) supplemented with 0.5 mM phenylmethanesulfonyl fluoride, 0.5 μg/ml aprotinin, 0.5 μg/ml leupeptin and 0.7 μg/ml pepstatin, and centrifuged at 3,000 g for 10 min. All steps were performed at 4°C. The lipid layer on the surface was removed, and the supernatant centrifuged at 18,000 g for 30 min. The pellet was suspended in external recording solution and 4xSDS sample buffer with 50 mM DTT, and incubated for 10 min at 70°C. The samples were loaded on an 8% SDS gel. Running buffer contained 190 mM glycine, 25 mM Tris base and 0.1% SDS for electrophoresis. Precision Plus Protein Standard (Bio-Rad) was used as protein weight marker. Wet transfer onto a nitrocellulose membrane was performed with a transfer buffer containing 190 mM glycine, 25 mM Tris base and 20% methanol. The membrane was then incubated with the anti-myc antibody (Invitrogen) and protein was detected using SuperSignal West Dura Substrate (Invitrogen).

References

- Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu Rev Biochem* **79**, 413–444, doi: 10.1146/annurev.biochem.052308.105824 (2010).
- Kalstrup, T. & Blunck, R. Dynamics of internal pore opening in K(V) channels probed by a fluorescent unnatural amino acid. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 8272–8277, doi: 10.1073/pnas.1220398110 (2013).
- Kozak, M. The scanning model for translation: an update. *J Cell Biol* **108**, 229–241 (1989).
- von der Haar, T. & Tuite, M. F. Regulated translational bypass of stop codons in yeast. *Trends in microbiology* **15**, 78–86, doi: 10.1016/j.tim.2006.12.002 (2007).
- Herr, A. J., Nelson, C. C., Wills, N. M., Gesteland, R. F. & Atkins, J. F. Analysis of the roles of tRNA structure, ribosomal protein L9, and the bacteriophage T4 gene 60 bypassing signals during ribosome slippage on mRNA. *Journal of molecular biology* **309**, 1029–1048, doi: 10.1006/jmbi.2001.4717 (2001).
- Napthine, S., Yek, C., Powell, M. L., Brown, T. D. & Brierley, I. Characterization of the stop codon readthrough signal of Colorado tick fever virus segment 9 RNA. *Rna* **18**, 241–252, doi: 10.1261/rna.030338.111 (2012).
- Bonetti, B., Fu, L., Moon, J. & Bedwell, D. M. The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. *Journal of molecular biology* **251**, 334–345, doi: 10.1006/jmbi.1995.0438 (1995).
- Cassan, M. & Rousset, J. P. UAG readthrough in mammalian cells: effect of upstream and downstream stop codon contexts reveal different signals. *BMC molecular biology* **2**, 3 (2001).
- Dunn, J. G., Foo, C. K., Belletier, N. G., Gavis, E. R. & Weissman, J. S. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* **2**, e01179, doi: 10.7554/eLife.01179 (2013).
- Namy, O. *et al.* Identification of stop codon readthrough genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **31**, 2289–2296 (2003).
- Kozak, M. Effects of intercistronic length on the efficiency of reinitiation by eucaryotic ribosomes. *Mol Cell Biol* **7**, 3438–3445 (1987).
- Poyry, T. A., Kaminski, A. & Jackson, R. J. What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame? *Genes Dev* **18**, 62–75, doi: 10.1101/gad.276504 (2004).
- Kozak, M. Constraints on reinitiation of translation in mammals. *Nucleic Acids Res* **29**, 5226–5232 (2001).
- Luukkonen, B. G., Tan, W. & Schwartz, S. Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. *J Virol* **69**, 4086–4094 (1995).
- Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat Rev Mol Cell Biol* **11**, 113–127, doi: 10.1038/nrm2838 (2010).
- Powell, M. L. Translational termination-reinitiation in RNA viruses. *Biochem Soc Trans* **38**(6), 1558–1564 (2010).
- Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **234**, 187–208 (1999).
- von Arnim, A. G., Jia, Q. & Vaughn, J. N. Regulation of plant translation by upstream open reading frames. *Plant science : an international journal of experimental plant biology* **214**, 1–12, doi: 10.1016/j.plantsci.2013.09.006 (2014).
- Hinnebusch, A. G., Jackson, B. M. & Mueller, P. P. Evidence for regulation of reinitiation in translational control of GCN4 mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 7279–7283 (1988).
- Kochetov, A. V., Sarai, A., Rogozin, I. B., Shumny, V. K. & Kolchanov, N. A. The role of alternative translation start sites in the generation of human protein diversity. *Mol Genet Genomics* **273**, 491–496, doi: 10.1007/s00438-005-1152-7 (2005).
- Rogozin, I. B., Kochetov, A. V., Kondrashov, F. A., Koonin, E. V. & Milanese, L. Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon. *Bioinformatics* **17**, 890–900 (2001).
- Pisani, F., Rossi, A., Nicchia, G. P., Svelto, M. & Frigeri, A. Translational regulation mechanisms of aquaporin-4 supramolecular organization in astrocytes. *Glia* **59**, 1923–1932, doi: 10.1002/glia.21234 (2011).
- Stump, M. R., Gong, Q., Packer, J. D. & Zhou, Z. Early LQT2 nonsense mutation generates N-terminally truncated hERG channels with altered gating properties by the reinitiation of translation. *J Mol Cell Cardiol* **53**, 725–733, doi: 10.1016/j.yjmcc.2012.08.021 (2012).
- Stump, M. R., Gong, Q. & Zhou, Z. LQT2 nonsense mutations generate trafficking defective NH2-terminally truncated channels by the reinitiation of translation. *Am J Physiol Heart Circ Physiol* **305**, H1397–1404, doi: 10.1152/ajpheart.00304.2013 (2013).

25. Hoshi, T., Zagotta, W. N. & Aldrich, R. W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533–538 (1990).
26. Aldrich, R. W. Fifty years of inactivation. *Nature* **411**, 643–644, doi: 10.1038/35079705 (2001).
27. Taghialatela, M., Toro, L. & Stefani, E. Novel voltage clamp to record small, fast currents from ion channels expressed in *Xenopus* oocytes. *Biophys J* **61**, 78–82, doi: 10.1016/S0006-3495(92)81817-9 (1992).
28. Zagotta, W. N., Hoshi, T. & Aldrich, R. W. Restoration of inactivation in mutants of Shaker potassium channels by a peptide derived from ShB. *Science* **250**, 568–571 (1990).
29. Wang, W. *et al.* Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat Neurosci* **10**, 1063–1072, doi: 10.1038/nn1932 (2007).
30. Peabody, D. S. Translation initiation at non-AUG triplets in mammalian cells. *J Biol Chem* **264**, 5031–5035 (1989).
31. Starck, S. R. *et al.* Leucine-tRNA initiates at CUG start codons for protein synthesis and presentation by MHC class I. *Science* **336**, 1719–1723, doi: 10.1126/science.1220270 (2012).
32. Ivanov, I. P., Firth, A. E., Michel, A. M., Atkins, J. F. & Baranov, P. V. Identification of evolutionarily conserved non-AUG-initiated N-terminal extensions in human coding sequences. *Nucleic Acids Res* **39**, 4220–4234, doi: 10.1093/nar/gkr007 (2011).
33. Lee, H. S., Guo, J., Lemke, E. A., Dimla, R. D. & Schultz, P. G. Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. *Journal of the American Chemical Society* **131**, 12921–12923, doi: 10.1021/ja904896s (2009).
34. Chatterjee, A., Guo, J., Lee, H. S. & Schultz, P. G. A genetically encoded fluorescent probe in mammalian cells. *Journal of the American Chemical Society* **135**, 12540–12543, doi: 10.1021/ja4059553 (2013).
35. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**, 8125–8148 (1987).
36. Kozak, M. Context effects and inefficient initiation at non-AUG codons in eucaryotic cell-free translation systems. *Mol Cell Biol* **9**, 5073–5080 (1989).
37. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
38. Schulteis, C. T., John, S. A., Huang, Y., Tang, C. Y. & Papazian, D. M. Conserved cysteine residues in the shaker K⁺ channel are not linked by a disulfide bond. *Biochemistry* **34**, 1725–1733 (1995).
39. Santacruz-Tolosa, L., Huang, Y., John, S. A. & Papazian, D. M. Glycosylation of shaker potassium channel protein in insect cell culture and in *Xenopus* oocytes. *Biochemistry* **33**, 5607–5613 (1994).
40. Munzarova, V. *et al.* Translation reinitiation relies on the interaction between eIF3a/TIF32 and progressively folded cis-acting mRNA elements preceding short uORFs. *PLoS Genet* **7**, e1002137, doi: 10.1371/journal.pgen.1002137 (2011).
41. Wang, K., Schmied, W. H. & Chin, J. W. Reprogramming the genetic code: from triplet to quadruplet codes. *Angewandte Chemie* **51**, 2288–2297, doi: 10.1002/anie.201105016 (2012).
42. Batulan, Z., Haddad, G. A. & Blunck, R. An intersubunit interaction between S4-S5 linker and S6 is responsible for the slow off-gating component in Shaker K⁺ channels. *J Biol Chem* **285**, 14005–14019, doi: 10.1074/jbc.M109.097717 (2010).
43. Haddad, G. A. & Blunck, R. Mode shift of the voltage sensors in Shaker K⁺ channels is caused by energetic coupling to the pore domain. *J Gen Physiol* **137**, 455–472, doi: 10.1085/jgp.201010573 (2011).
44. Bezanilla, F. & Armstrong, C. M. Inactivation of the sodium channel. I. Sodium current experiments. *J Gen Physiol* **70**, 549–566 (1977).
45. Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K⁺ channel. *Neuron* **16**, 1159–1167 (1996).
46. Schoppa, N. E., McCormack, K., Tanouye, M. A. & Sigworth, F. J. The size of gating charge in wild-type and mutant Shaker potassium channels. *Science* **255**, 1712–1715, doi: 10.1126/science.1198888 (1992).
47. Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K⁺ channel. *Neuron* **16**, 1169–1177 (1996).
48. Geering, K., Theulaz, I., Verrey, F., Hauptle, M. T. & Rossier, B. C. A role for the beta-subunit in the expression of functional Na⁺-K⁺-ATPase in *Xenopus* oocytes. *The American journal of physiology* **257**, C851–858 (1989).

Acknowledgements

We thank Mireille Marsolais for technical assistance and Dr. H el ene Klein for valuable advice. Anap and the plasmid encoding the tRNA/AnapRS pair (pAnap) were kind gifts from Dr. Peter G. Schultz (Scripps Research Institute, La Jolla, CA). This work was funded by the Canadian Institutes for Health Research Grants MOP-102689 and MOP-136894 (to R.B.) and Canadian Foundation for Innovation Grant 950-225005. Groupe d' tudes des Prot ines Membranaires is a research group funded by the Fonds de recherche du Qu bec-Sant . R.B. holds a Canada Research Chair on Molecular Mechanisms of Membrane Proteins.

Author Contributions

T.K. performed experiments; T.K. and R.B. designed research, analyzed data and wrote the manuscript.

Additional Information

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

How to cite this article: Kalstrup, T. and Blunck, R. Reinitiation at non-canonical start codons leads to leak expression when incorporating unnatural amino acids. *Sci. Rep.* **5**, 11866; doi: 10.1038/srep11866 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>