Translation of non-standard codon nucleotides reveals minimal requirements for codon-anticodon interactions

Thomas Philipp Hoernes1, Klaus Faserl2, Michael Andreas Juen3, Johannes Kremser3, Catherina Gasser3, Elisabeth Fuchs3, Xinying Shi4, Aaron Siewert5, Herbert Lindner2, Christoph Kreutz3, Ronald Micura3, Simpson Joseph4, Claudia Höbartner5, Eric Westhof6, Alexander Hüttenhofer1 & Matthias David Erlacher1

The precise interplay between the mRNA codon and the tRNA anticodon is crucial for ensuring efficient and accurate translation by the ribosome. The insertion of RNA nucleobase derivatives in the mRNA allowed us to modulate the stability of the codon-anticodon interaction in the decoding site of bacterial and eukaryotic ribosomes, allowing an in-depth analysis of codon recognition. We found the hydrogen bond between the N1 of purines and the N3 of pyrimidines to be sufficient for decoding of the first two codon nucleotides, whereas adequate stacking between the RNA bases is critical at the wobble position. Inosine, found in eukaryotic mRNAs, is an important example of destabilization of the codon-anticodon interaction. Whereas single inosines are efficiently translated, multiple inosines, e.g., in the serotonin receptor 5-HT2C mRNA, inhibit translation. Thus, our results indicate that despite the robustness of the decoding process, its tolerance toward the weakening of codon-anticodon interactions is limited.
n RNA, Watson–Crick (W–C) base pairing is ubiquitous but is only one of numerous possible interactions that can be formed due to the single-stranded nature of RNA. This structural versatility enables single-stranded RNA not only to contain and transport simple sequence information in the form of messenger RNAs (mRNAs) but also to execute enzymatic functions as ribozymes. During protein synthesis, both the structural variability and the sequence information of RNA are absolutely essential. Transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) form characteristic and elaborate tertiary structures enabling the optimized and fine-tuned translation of mRNAs into proteins.

However, the very basis for the decoding of mRNA sequences during protein synthesis is W–C base pairing. At the ribosomal A site, the mRNA codon is presented to the incoming tRNA anticodon, thereby forming W–C interactions. Whereas W–C base pairing at the first and second codon nucleotides is strictly required, the conformation at the third, the so-called wobble position, is structurally more flexible.

Extensive structural, physico-chemical, and kinetic studies of the decoding processes revealed that an integrative interaction network between mRNA, tRNA, and rRNA ensures efficient and accurate translation. Additional factors have been identified that modulate the decoding process by impacting the quality of the codon–anticodon interaction. A significant and essential influence on decoding derives from RNA modifications. In tRNAs, in particular, contain many post-transcriptionally modified nucleotides. Also, rRNAs and mRNAs harbor numerous base and ribose modifications, implying that they may play an important role during protein synthesis. Their functions, however, are still largely unknown.

A well-characterized non-standard nucleoside in mRNAs is inosine (I). Is a result of a hydrolytic deamination of adenosine by a family of proteins called ADARs. This editing event leads to a switch from a hydrogen donor (amino group) to a hydrogen acceptor (carbonyl oxygen) at position 6, generating a W–C edge reminiscent of guanosine (G), thereby altering the genetic information through the preferential base pairing of I with cytosine (C). In addition, the I–C interaction is less stable than that of G–C, mainly due to the loss of one hydrogen bond (H-bond) since inosine has also been revealed in coding sequences (CDSs) of mRNAs, it is remarkable that, so far, inosine has not been observed to impair protein synthesis. This potentially implies that the number of H-bonds between codon and anticodon is less critical during translation than previously assumed or that a loss in stability of the codon–anticodon interaction can be compensated for by other means.

Decoding of mRNAs has been extensively studied during the last decades. Recently, complementary packing and hydrophobic forces have been demonstrated to be crucial for decoding; however, the contribution of single H-bonds to the codon–anticodon interaction has not yet been systematically addressed. By specific insertions of various non-natural modifications into mRNA codons, we intended to define the limits for the stability of the W–C interaction in the ribosomal decoding site during protein synthesis. Strikingly, in bacteria as well as in eukaryotes, the number of H-bonds at single codon positions only marginally affected protein synthesis. Thus, it is not the number of H-bonds, but rather the contact interactions that maintain the overall geometry and shape of the base pair that are critical for translation.

**Results**

**Non-standard nucleotides as tools to investigate decoding.** To manipulate W–C interactions at the decoding site of the ribosome, a variety of non-natural RNA nucleobase derivatives were introduced site-specifically into reporter mRNAs (Fig. 1). This was achieved by employing chemically synthesized oligonucleotides harboring the desired base modification. Due to the length limitations of these oligonucleotides, a splintered ligation was carried out covalently linking the modified 3’-fragment to a 5’-transcript, resulting in a full-length mRNA. The ligation products were purified and subsequently served as templates for translation. To investigate bacterial and eukaryotic translation processes, the recombinant PURExpress system (NEB) and HEK293T cells were employed, respectively.

**Defining the basic rules of the codon–anticodon interactions.** To define the boundaries for efficient decoding, the codon–anticodon interactions were drastically weakened by inserting base modifications within a GGG (Gly) codon. Thereby, the codon context is strong in terms of hydrogen bonding, and the respective tRNA only carries a comparably low number of modifications, thus reducing the contribution of RNA modifications to decoding. Initially, benzimidazole (Benz) and a ribose abasic site (Rab) were site-specifically introduced into this codon. Benzimidazole (Fig. 1) cannot form H-bonds with a pyrimidine base due to the absence of the N1 and the 6-amino group, assuming that a W–C geometry is formed. Due to the missing heterocycle of the purine or pyrimidine base, Rab-sites do not provide any stabilization of the decoding site via base stacking, which further weakens the codon–anticodon interaction (Fig. 1). Not unexpectedly, in both bacterial and eukaryotic systems the introduction of Rab-sites did not allow translation of the modified mRNA, independent of its position within the GGG codon. Unexpectedly, in both bacterial and eukaryotic systems the introduction of Rab-sites did not allow translation of the modified mRNA, independent of its position within the GGG codon. In HEK293T cells, however, Py was not translated when it was located at the first or second codon position. However, at the wobble position, Benz allowed for protein synthesis comparable to that of unmodified mRNAs.

A less drastic modification, in respect to exocyclic groups potentially participating in H-bonding, is pyridone (Py) (Fig. 1) in a UUU codon. In a W–C geometry, however, Py should not favor base pairing with A because of the close contact between the amino group of A and the C3-H of Py. This interaction is less critical during translation than previously assumed or that a loss in stability of the codon–anticodon interaction can be compensated for by other means.

Decoding of mRNAs has been extensively studied during the last decades. Recently, complementary packing and hydrophobic forces have been demonstrated to be crucial for decoding; however, the contribution of single H-bonds to the codon–anticodon interaction has not yet been systematically addressed. By specific insertions of various non-natural modifications into mRNA codons (Fig. 1), we intended to define the limits for the stability of the W–C interaction in the ribosomal decoding site during protein synthesis. Strikingly, in bacteria as well as in eukaryotes, the number of H-bonds at single codon positions only marginally affected protein synthesis. Thus, it is not the number of H-bonds, but rather the contact interactions that maintain the overall geometry and shape of the base pair that are critical for translation.

**Results**

**Non-standard nucleotides as tools to investigate decoding.** To manipulate W–C interactions at the decoding site of the ribosome, a variety of non-natural RNA nucleobase derivatives were introduced site-specifically into reporter mRNAs (Fig. 1). This was achieved by employing chemically synthesized oligonucleotides harboring the desired base modification. Due to the length limitations of these oligonucleotides, a splintered ligation was carried out covalently linking the modified 3’-fragment to a 5’-transcript, resulting in a full-length mRNA. The ligation products were purified and subsequently served as templates for translation. To investigate bacterial and eukaryotic translation processes, the recombinant PURExpress system (NEB) and HEK293T cells were employed, respectively.

**Defining the basic rules of the codon–anticodon interactions.** To define the boundaries for efficient decoding, the codon–anticodon interactions were drastically weakened by inserting base modifications within a GGG (Gly) codon. Thereby, the codon context is strong in terms of hydrogen bonding, and the respective tRNA carries a comparably low number of modifications, thus reducing the contribution of RNA modifications to decoding. Initially, benzimidazole (Benz) and a ribose abasic site (Rab) were site-specifically introduced into this codon. Benzimidazole (Fig. 1) cannot form H-bonds with a pyrimidine base due to the absence of the N1 and the 6-amino group, assuming that a W–C geometry is formed. Due to the missing heterocycle of the purine or pyrimidine base, Rab-sites do not provide any stabilization of the decoding site via base stacking, which further weakens the codon–anticodon interaction (Fig. 1). Not unexpectedly, in both bacterial and eukaryotic systems the introduction of Rab-sites did not allow translation of the modified mRNA, independent of its position within the GGG codon. In HEK293T cells, however, Py was not translated when it was located at the first or second codon position. However, at the wobble position, Benz allowed for protein synthesis comparable to that of unmodified mRNAs.

A less drastic modification, in respect to exocyclic groups potentially participating in H-bonding, is pyridone (Py) (Fig. 1) in a UUU codon. In a W–C geometry, however, Py should not favor base pairing with A because of the close contact between the amino group of A and the C3-H of Py. This interaction is less critical during translation than previously assumed or that a loss in stability of the codon–anticodon interaction can be compensated for by other means.

Decoding of mRNAs has been extensively studied during the last decades. Recently, complementary packing and hydrophobic forces have been demonstrated to be crucial for decoding; however, the contribution of single H-bonds to the codon–anticodon interaction has not yet been systematically addressed. By specific insertions of various non-natural modifications into mRNA codons (Fig. 1), we intended to define the limits for the stability of the W–C interaction in the ribosomal decoding site during protein synthesis. Strikingly, in bacteria as well as in eukaryotes, the number of H-bonds at single codon positions only marginally affected protein synthesis. Thus, it is not the number of H-bonds, but rather the contact interactions that maintain the overall geometry and shape of the base pair that are critical for translation.
encoding for Arg was efficiently translated (Supplementary Fig. 5). To shed more light on this observation and to examine not only the number but also the position of the H-bonds, 1-deazaadenosine (c1A) was investigated (Fig. 1f). In analogy to P, c1A can potentially only form a single H-bond but, in this case, through the 6-amino group and not through the N1 (Supplementary Fig. 2c). c1A was introduced separately at every position within an AAA codon. In such codons, c1A did not enable protein synthesis at the first or the second position (Fig. 2e, f). This can be rationalized not only by the impact of the position of the single H-bond but also by the close proximity of the C1-H and N3-H of the c1A and the U, respectively (Supplementary Fig. 2c). Consistently, the c1A–U interaction was drastically destabilized in the minimal codon–anticodon model system, decreasing the Tm below the detection limit (Supplementary Fig. 4). Unexpectedly, the third codon position c1A could also not be decoded, presumably, because a proper wobble base pair could not be formed (Fig. 2e, f; Supplementary Fig. 2d).
**Fig. 2** Translation of mRNAs carrying non-standard nucleotides. Left panel a, c, e, g: Autoradiography of ErmCL peptides synthesized using the bacterial in vitro translation system. Right panel b, d, f, h: Western blot analyses of modified mRNAs translated in HEK293T cells. An unmodified eGFP mRNA and α-tubulin served as an internal transfection control and as a loading control, respectively. Amino acids incorporated at the respective positions of peptides purified from i the *E. coli* translation system and j HEK293T cells were analyzed by mass spectrometry. Error bars depict the SDs from the mean of at least three independent experiments.
Since we observed that not only the number but also the positions of the formed H-bonds were important, 2-aminopurine (Ap) and 2,6-diaminopurine (Dap) (Fig. 1d, e) were incorporated to evaluate the importance of the exocyclic group for base pairing. In the bacterial and eukaryotic translation systems, both modifications enabled efficient protein synthesis (Fig. 2g, h; Supplementary Fig. 1). In the case of Dap, even the increased number of H-bonds did not significantly alter the translation process (Fig. 2g, h). P, Ap, and Dap had similar translation levels in both bacteria and eukaryotes.

**Codon–anticodon interactions that define sense codons.** In addition to the protein products that are formed in dependence of the modified codons, the interpretation of these RNA nucleobase derivatives by the ribosomal decoding site was also determined. For each modification, the translated peptides from the bacterial and eukaryotic systems were purified and analyzed by mass spectrometry (MS) (Fig. 2i, j; Supplementary Tables 2–4). In respect to the pyridine modifications, Zc was decoded as C since ZcU.U and UZcE codons resulted in the incorporation of leucine (Leu) and serine (Ser), respectively. In bacteria, UZcE was also partly decoded (~8%) as a phenylalanine (Phe) codon, indicating that Zc can also base pair with A to a limited extent (Fig. 2i; Supplementary Table 2). Although the base pair Zc–A is also observed in HEK293T cells, it is less abundant than in bacteria (Fig. 2j; Supplementary Tables 3 and 4).

P within AAA codons was always translated as lysine (Lys) independent of its position within the codon and the translation system (Fig. 2i, j; Supplementary Tables 2 and 3). By placing an exocyclic amino group at position 2, a shift in decoding from A to G was assumed, since the 2-amino group is also present in G. In eukaryotes, however, Ap was exclusively read as A at the first two codon positions (Fig. 2j; Supplementary Table 3). In bacteria, ApGG led to the incorporation of tryptophan (Trp), which was a result of decoding ApGG as UGG (Fig. 2i; Supplementary Table 2). This is remarkable, since the Trp incorporation would require a purine–purine base pair within the codon–anticodon interaction. Ap at the second nucleotide of the codon led to low levels of Gly in addition to Glu, caused by reading Ap as G (Fig. 2i; Supplementary Table 2). Placing a 6-amino group in addition to the 2-amino group into the mRNA, by integrating Dap, led to an unambiguous base pairing of Dap with U in both tested translation systems (Fig. 2i, j; Supplementary Tables 2 and 3).

**Decoding of non-standard nucleotides at the wobble position.** Due to the degenerate nature of the genetic code, addressing the decoding of modified bases at the wobble position is limited to the AUG codon. AUG is the only codon that enables the differentiation of whether the ribosome interprets the modified purine base as A or G, since AUG decodes for methionine (Met) and AUA/C/G for isoleucine (Ile).

The codon AURab was not efficiently translated (Fig. 3a, b). Nevertheless, the bacterial translation system provided sufficient amounts of the peptide for MS analysis, revealing that the majority of peptides contained Ile and only a fraction of the peptides contained Met (Fig. 3a, c; Supplementary Tables 2 and 4), possibly reflecting the abundance of the respective tRNAs. In accordance with the results obtained when P was positioned at the first two codon nucleotides, this base derivative was decoded almost exclusively as an A in the AUP codon (Fig. 3c, d; Supplementary Table 2–4). Remarkably, the AUP codon was not as efficiently translated as the AAP codon (Figs. 3a, b and 2e, f). Furthermore, the introduction of c1A at the third codon position...
of the Lys codon did not result in any translation product, whereas sufficient amounts of peptide could be detected for AUc1A, which was decoded as isoleucine (Fig. 3a, c; Supplementary Table 2).

At the wobble position, the relocation of the exocyclic amino group from position 6 to 2 indeed increased the incorporation of Met, caused by Ap being decoded as G (Fig. 3c, d; Supplementary Tables 2 and 3). This effect was more pronounced in the bacterial than in the eukaryotic translation system. The simultaneous presence of 2- and 6-amino groups reduced the incorporation of Met, indicating a compensatory effect of the 6-amino group in the presence of the 2-amino group (Fig. 3c, d; Supplementary Tables 2 and 3).

The discrimination between AUA and AUG codons is different between bacteria and eukaryotes. In bacteria, AUA is decoded to the incorporation of arginine instead of glutamine (Fig. 4b, c); editing at these sites affects G-protein coupling and, consequently, the receptor’s activity. One of the most prominent and best-studied examples of A-to-I editing is the serotonin 5-HT2C receptor (5-HT2CR) mRNA. In total, five editing positions on exon V, designated site A to site E, have been reported (Fig. 4b, c); editing at these sites affects G-protein coupling and, consequently, the receptor’s activity. Another transcript that is almost quantitatively edited is the glutamate receptor GluR-B mRNA expressed in the brain, where at the protein level, the editing event leads to the incorporation of arginine instead of glutamine (Fig. 4b, c). This amino acid exchange is associated with the altered calcium permeability of the receptor. We incorporated single inosines in the mRNA contexts of either 5-HT2CR or GluR-B and investigated their influence on eukaryotic translation. We found that single inosines did not affect the yield of the translated peptide product (Fig. 4d). As expected, the inosines were exclusively decoded as G, resulting in an amino acid change from Gln to Arg and from Asn to Ser in the case of the modified 5-HT2CR and GluR-B mRNAs, respectively (Fig. 4e; Supplementary Table 3). However, the simultaneous presence of five inosines within the 5-HT2CR mRNA completely abolished its translation (Fig. 4d).

Therefore, we studied the effect of I on the stability of the codon–anticodon interaction in both the absence and the presence of the ribosome (Fig. 5). Initially, we employed a minimal codon–anticodon system in solution and measured the

**Inosine affects the genetic variability and tRNA binding.** A naturally occurring example of an altered number of H-bonds during decoding is the translation of I-containing codons. The I–C base pair forms only two H-bonds, whereas the canonical G–C base pair exhibits three H-bonds (Fig. 4a). The conversion of adenosines to inosines, designated as A-to-I editing, is the most prevalent form of RNA editing, with possibly more than 100 million potential modification sites in the human transcriptome.

One of the most prominent and best-studied examples of A-to-I editing is the serotonin 5-HT2C receptor (5-HT2CR) mRNA. In total, five editing positions on exon V, designated site A to site E, have been reported (Fig. 4b, c); editing at these sites affects G-protein coupling and, consequently, the receptor’s activity. Another transcript that is almost quantitatively edited is the glutamate receptor GluR-B mRNA expressed in the brain, where at the protein level, the editing event leads to the incorporation of arginine instead of glutamine (Fig. 4b, c). This amino acid exchange is associated with the altered calcium permeability of the receptor. We incorporated single inosines in the mRNA contexts of either 5-HT2CR or GluR-B and investigated their influence on eukaryotic translation. We found that single inosines did not affect the yield of the translated peptide product (Fig. 4d). As expected, the inosines were exclusively decoded as G, resulting in an amino acid change from Gln to Arg and from Asn to Ser in the case of the modified 5-HT2CR and GluR-B mRNAs, respectively (Fig. 4e; Supplementary Table 3). However, the simultaneous presence of five inosines within the 5-HT2CR mRNA completely abolished its translation (Fig. 4d).

Therefore, we studied the effect of I on the stability of the codon–anticodon interaction in both the absence and the presence of the ribosome (Fig. 5). Initially, we employed a minimal codon–anticodon system in solution and measured the
interaction strength of inosines within a glycine codon (GGG) using CCC as an anticodon (Fig. 5a)\textsuperscript{43,44}. In such constructs, the stability of the codon–anticodon interaction changed depending on the position of inosine. The respective \( T_m \) values (IGG: 47.5 °C, GIG: 46.6 °C, GGI: 51.6 °C) indicated that I–C base pairs, especially in the first and the second codon positions, are significantly less stable compared to an unmodified codon (GGG: 64.5 °C).

We then asked whether this is also the case when an inosine-containing codon is present in the ribosomal A site. To enable an analysis of initial selection at inosine-encoding codons, we measured binding rates of EF–Tu ternary complexes in the A/T state of the 70S ribosome\textsuperscript{15}. Due to the limited availability of purified \( E. \) coli tRNAs, we synthesized a fully modified tRNA\textsuperscript{Gly} (Supplementary Fig. 6) that was subsequently radioactively labeled and charged. Ternary complexes of the Gly–[\textsuperscript{32P}]tRNA\textsuperscript{Gly} with EF–Tu (His84Ala) and GTP were formed (Supplementary Fig. 7). The subsequent filter-binding experiments were performed using varying concentrations of the ribosome complexes (Fig. 5b; Supplementary Table 1). The equilibrium dissociation constant \( (K_D, \text{depicted as } K_P, \pm \text{standard deviation from the mean of at least three independent experiments}) \) with an unmodified GGG codon in the A site was found to be 49.1 ± 2.0 nM. Binding of mRNAs with IGG or GIG codons to the ribosome increased \( K_D \) values compared to GGG more than 4-fold or 2.6-fold, respectively (IGG: \( K_D = 205.3 \pm 8.9 \text{ nM and GIG: } K_D = 127.9 \pm 7.8 \text{ nM} \); Fig. 5b; Supplementary Table 1). GGI codons only marginally increased \( K_D \) values to 64.8 ± 3.5 nM. Thus, the binding data are in line with the results obtained in solution and indicate that inosines in the A site moderately but significantly interfere with binding of EF–Tu ternary complexes. In particular, inosines in the first and the second codon positions decreased the stability of the codon–anticodon interaction.

Next, we investigated whether this position-dependent effect of inosine could be reproduced in an authentic translation setting. We analyzed the translation of IGG, GIG, and GGI-encoding mRNAs with the PURExpress translation system (Fig. 5c) and HEK293T cells (Fig. 5d). In bacteria, inosine in the first codon position reduced translation rates by approximately 30% (Supplementary Fig. 1); this effect could not be observed in mammalian cells, however. Furthermore, we determined the amino acids that were incorporated at the respective positions of the peptides (Fig. 5e; Supplementary Tables 2 and 3). In both bacterial and eukaryotic translation systems, miscoding was observed at low levels when inosine was decoded in the first codon position: the IGG codon was read 0.4% and 1.9% of the time as AGG (Arg) (or CGN) in bacteria and eukaryotes, respectively. Further, in bacteria, IGG was read 1.4% of the time as CAG (Gln) (Fig. 5e; Supplementary Table 4). GIG codons were also decoded as Val (GUG) in bacteria. Thus, especially in bacterial translations, purine–purine interactions (either G–I or A–I) are observed to a minor extent. Inosine in the wobble position of an AUG codon was almost exclusively decoded as guanosine in bacteria and eukaryotes (Fig. 3c, d; Supplementary Tables 2 and 3).
Although single inosines impair binding of ternary complexes in a codon position-dependent manner, ribosomal elongation was only moderately affected. However, a simultaneous incorporation of two inosines within one codon (A- and B-sites of the 5-HT_{2CR} mRNA) completely abolished translation of the edited mRNA (Figs. 4b and 5f). Even inosines within two distinct codons (C- and D-sites of the 5-HT_{2CR} mRNA) drastically reduced translation rates, indicating that inosines exhibit an additive inhibitory effect, at least when present in close proximity to each other.

**Discussion**

The codon–anticodon interaction is without doubt one of the most crucial interactions in molecular biology. Decades of research have evaluated numerous aspects and factors contributing to the speed and accuracy of the decoding process during protein synthesis. Although these efforts have led to a detailed picture of the translation process, the contribution of single H-bonds between the W–C edges has not been biochemically investigated so far\(^{16,46,47}\). Since the formation of stable complexes between the codon and anticodon is apparently crucial for decoding, we investigated the contributions of H-bonds in respect to their number and positions within the W–C geometry by inserting non-natural RNA nucleobase derivatives in the mRNA codons. We find that the translation process in bacteria and eukaryotes is astonishingly robust against the loss of single H-bonds and resulting in the destabilization of the W–C base pairs (Supplementary Table 4). Translation of the respective codon is only modestly impaired, when H-bonds between the purine-N\(^1\) and the pyrimidine-N\(^3\) at the first two codon positions are formed. Thus, in the cases of pyridone or \(^{1}A\), the single H-bond is at a different location and translation efficiencies are drastically reduced. The only exception was the translation of single purines within an AAA codon (Lys) in HEK293T cells. Whereas in *E. coli*, the codon APA could be efficiently decoded, it did not serve as an efficient template for translation in HEK293T cells. This may be explained by the presence of tRNA modifications in human tRNA\(^{Lys}_{UUU}\) that potentially destabilize the interaction with APA\(^{48}\), since this effect is not observed in the case of other codons (Supplementary Fig. 5).

Another difference between translation in bacteria and eukaryotes is the accuracy in the decoding of certain RNA nucleobase derivatives at defined codon positions. Generally, bacterial translation seems to be less restrictive towards the translation of modified codon nucleotides, as observed for the codons UZeU, ApGG, and GIG. In case of ApGG, the bacterial ribosome incorporates more than 10% of Trp into the peptide (Fig. 2i). However, this requires the formation of a purine–purine base pair at the first codon position. Another purine–purine pair was found when translating the GIG codon, but, again, only in the bacterial system. In the case of UZeU, the bacterial translation system incorporates Phe instead of Ser (Fig. 2i). Zc, in all other cases, was read as a C due to the hybridization state of N\(^3\). Interestingly, the eukaryotic translation system appears to be more stringent against changes within CDSs of mRNAs. Noteworthy, due to the lower amounts of purified translation products from HEK293T cells, we cannot completely exclude the existence of low-level peptide variants (below the detection limit; typically <1%). One could speculate that—among other underlying factors—an increased protein length\(^{49}\) or the expanded lifespan of eukaryotes\(^{50,51}\) could require a more accurate decoding in higher organisms.

At the wobble position, a generally less restrictive decoding behavior was observed for both bacteria and eukaryotes, as reported previously\(^{47}\). Even nucleotides that could not form an H-bond between N\(^1\) and N\(^3\) or were completely lacking any H-bonds (i.e., Py and Benz, respectively) provided efficient protein synthesis. In addition, the presence of the 2-amino group within Ap and Dap could compete with the otherwise determining N\(^3\)–N\(^3\) H-bond, confirming a higher structural flexibility at this position. However, decoding of the wobble position also supports the general observation that protein synthesis in eukaryotes is more stringent than in bacteria.

In eukaryotes, A-to-I editing causes a rewiring of the genetic code accompanied by a reduced number of H-bonds during decoding, thereby leading to an amino acid exchange in the respective protein or peptide, as in the cases of GluR-B and 5-HT_{2CR}\(^{39,42}\). However, this gain in flexibility might come at a cost. The I–C base pair is significantly destabilized in comparison to the standard G–C base pair due to the loss of two H-bonds (one to the C and one to either A1492 or A1493 of the ribosomal A site; *E. coli* numbering), but this still allows for efficient and accurate translation (Figs. 4 and 5). This is remarkable, since the stability of the codon–anticodon helix at the A site is an important factor for decoding\(^{29}\). Indeed, inosine weakens the RNA–mRNA interactions during decoding, especially if placed at the first or second position of the mRNA codon, as demonstrated using a minimal codon–anticodon model system (Fig. 5a) and by EF–Tu A site binding (Fig. 3b). During elongation, however, a single inosine had no apparent effect, whereas the incorporation of multiple inosines into an mRNA strongly hindered its translation in HEK293T cells (Figs. 4d and 5f). This was unexpected, since it has been reported that up to five editing sites are located in the 5-HT_{2CR} mRNA and are efficiently edited in different combinations, depending on the brain areas and the developmental states in which they are expressed\(^{40,52}\). In particular, the presence of two inosines within one codon (i.e., concurrent editing of the A- and B-sites), drastically reduced peptide synthesis to levels below the detection limit (Fig. 5f). This indicates that a reduction of the codon–anticodon interaction strength by the inclusion of inosines is only tolerated to a certain extent. Although the absolute editing levels are difficult to assess, strong evidence indicates that multiple inosines are present simultaneously within the 5-HT_{2CR} mRNA\(^{40,52}\).

Over the last decades, different nucleotide analogs and base pairs have been screened for their ability to form stable W–C base pairs, predominately during replication and transcription\(^{53,54}\). In line with our findings, most of these nucleotide derivatives provided the structural prerequisites to form at least one H-bond. Interestingly, the pair 6-methoxypurine-thymine forms a central H-bond between N\(^1\) (purine) and N\(^3\) (pyrimidine) and was reported to have the highest incorporation efficiencies (among the three described base pairs), while the 2-amino-uracil-uracil pair that forms the H-bond between the 2-amino group and the carbonyl oxygen at C\(^2\) is less efficiently incorporated into the DNA\(^{54}\). The only exceptions are fluorine-containing bases. Although not forming H-bonds, they were incorporated during DNA replication\(^{55,56}\). More recently, an artificial base pair was identified that did not depend on the presence of H-bond interactions but still allowed efficient transcription and subsequent translation\(^{51}\). In this artificial base pair, the components are highly hydrophobic and their interaction leads to a base pair isosteric to W–C pairs. The missing H-bonds can most likely be compensated by different hydrophobic and stacking interactions\(^{37}\). In contrast to these studies, we systematically eliminated potential H-bond partners only from the mRNA codon side in the codon–anticodon helix, revealing the robustness of the decoding process in an authentic setting for protein synthesis. Clearly, changes within purines or pyrimidines impact also polarity, stacking, the syn/anti equilibrium and can lead to steric effects. These parameters contribute to the binding strength in a complex manner and would require
accommodate several natural and non-natural base modifications within mRNAs as long as the resulting base pair is isosteric to W–C geometry. The data show that the interactions, in the middle of the base pair, between N1 and N3 at the eukaryotic translation systems with the eukaryotic system of interactions between the codon and the anticodon, which regulate expression by inhibiting protein synthesis. Nonetheless, the flexibility of mRNAs is a potent way to increase the genetic choice. 

Ethyl- and 5′-GAGCAAGGGCGAGG-3′ (the sense codon that was modified or exchanged is underlined; UAA stop codon in bold.) were sequenced. In addition, we ligated a capped transcript to the poly(A)-tailed ErmCL oligonucleotide bridged by splinter 5′-TTTTTTGGTCTTTATATGTTTGTCTCTGTAGGA-3′ (the A-to-I editing site is underlined; UAA stop codon in bold) and 5′-HT2CR-5′-P-AUUAUUGGAAAGACAAACAAACAAUAAUUAUA-3′ (the 3′-end editing site is underlined; UAA stop codon in bold). The ligation sites were bridged with DNA oligonucleotides (GluR-B: 5′-GAGC AAAGGAAUCAUCUAGGACAUAGGCUA-3′ (the A-to-I editing sites are underlined; UAA stop codon in bold)).

Prokaryotic in vitro translation. In vitro translation (IVT) employing the PUREExpress Δ ribosome system (NEB, E3313S) was described as performed by the manufacturer. Briefly, 1 μM mRNAs were translated in the presence of 10 μCi [35S]Met (Hartmann Analytical, ACM-01H) and 1 μM 70S ribosomes. The IVT reactions were incubated for 1 h at 37 °C and were then resolved on Novex 16% Tricine gels (Thermo Fisher Scientific, EC66953BOX) and exposed to phoshorimage screens, which were scanned using a STORM 840 scanner.

E. coli tRNA synthesis. Standard RNA nucleoside building blocks were used in form of labile base- and 2′-cyanoethoxymethyl (CEM)-protected phosphor- amidites, synthesized according to published procedures. Modified tRNAs were exchanged via self-synthesized (DHU, 5′-U [will be published elsewhere], m5U) or purchased as 2′-TBDMS-protected phosphoramidites (Pseudouridine, Glen Research) (Supplementary Fig. 6).

Deprotection and cleavage was performed under extra mild conditions (2 M NH4 in MeOH, 37 °C, 19 h) to prevent DHU ring opening or substitution products at 5′U. Removal of the different 2′-protecting groups was achieved via treatment with 1 M TBAF in anhydrous dimethyl sulfoxide (DMSO) (2 ml) after dissolving the partially liberated RNA in anhydrous DMSO (0.5 ml) and adding 20 μl CH3NO2. After 20 h at 37 °C, the mixture was quenched with 2 ml Tris Buffer solution (Glen Research), and desalted using a HiPrep 26/10 desalting column (GE Healthcare). Purification of the desired RNA was carried out by anion exchange chromatography on a HiLoad 16/60 Superdex 200 10/300 GL column (GE Healthcare). The purified RNA was visualized using the Picogreen assay and quantified using a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific).

For the solid-phase syntheses of 1-deazaadenosine or inosine-containing oligonucleotides, either 2′,3′-dideoxy-3′-O-(2-cyanoethyl)-5′-O-(2-thio) adenosine (OMAD, Glen Research) or 2′,3′-dideoxy-3′-O-(2-cyanoethyl)-5′-O-(2-thio) inosine (OMID, Glen Research) were used.

... 2′-cyanoethyl-N,N-dimethylphosphoramidite (OMAD, Glen Research) or 2′,3′-dideoxy-3′-O-(2-cyanoethyl)-5′-O-(2-thio) inosine (OMID, Glen Research) were used. Oligonucleotide synthesis, deprotection, and quality control were carried out as previously described. The synthesis of the benzimidazole nucleotide will be published elsewhere.
Received: 16 July 2018 Accepted: 25 October 2018
Published online: 19 November 2018

**Data availability**

All mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE database with the data set identifiers PXD011301 (E. coli PURExpress translation assay) and PXD011300 (translation in HEK293T cells). All other data supporting the findings of this study are available within this article and in the Supplemental Information or from the corresponding author upon reasonable request. A reporting summary for this article is available as a Supplementary Information file.

**References**

1. Leontis, N. B., Stombaugh, J. & Westhof, E. The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* **30**, 3497–3531 (2002).
2. Moore, P. B. Structural motifs in RNA. *Annu. Rev. Biochem.* **68**, 287–300 (1999).
3. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
4. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
5. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
6. Kruger, K. et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–157 (1982).
7. Khatter, H., Myasnikov, A. G., Natchiar, S. K. & Klaholz, B. P. Structure of the human 80S ribosome. *Nature* **520**, 640–645 (2015).
8. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **299**, 905–920 (2003).
9. Yusupov, M. M. et al. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
10. Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new understanding of the decoding principle on the ribosome. *Nature* **484**, 256–259 (2012).
11. Ogle, J. M. et al. Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).
12. Rozov, A. et al. Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nat. Commun.* **7**, 10457 (2016).
13. Crick, F. H. Codon–anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548–555 (1966).
14. Ogle, J. M. & Ramakrishnan, V. Structural insights into translational fidelity. *Annu. Rev. Biochem.* **74**, 129–177 (2005).
15. Grosjean, H. & Westhof, E. An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res.* **44**, 8020–8040 (2016).
16. Wohlgenuth, I., Pohl, C., Mittelstaedt, J., Konevea, A. L. & Rodina, M. V. Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2979–2986 (2011).
17. Yusupova, G. Z., Yusupov, M. M., Cate, J. H. & Noller, H. F. The path of messenger RNA through the ribosome. *Cell* **106**, 233–241 (2001).
18. Boccaletto, P. et al. MODOMICS: a database of RNA modification pathways. *2017 update. Nucleic Acids Res.* **46**, D303–D307 (2018).
19. Schimmel, P. The emerging complexity of the RNA world: mammalian tRNAs beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **19**, 45–58 (2018).
20. Frigo, A. P., Narendran, A., Sarachan, K., Vare, V. Y. & Eryusal, E. The importance of being modified: the role of RNA modifications in translational fidelity. *Enzymes* **41**, 1–50 (2017).
21. Grosjean, H., de Crécy-Lagard, V. & Marck, C. Deciphering synonymous codons in the three domains of life: co-evolution with specific RNA modification enzymes. *FEBS Lett.* **584**, 252–264 (2010).
22. Sloan, K. E. et al. Tuning the ribosome: the in vivo decoding of mRNA through the ribosome. *Nat. Rev. Mol. Cell Biol.* **19**, 45–58 (2018).
23. Hoernes T. P. & Erlacher M. D. Translating the Epitranscriptome. *Wiley Interdiscip. Rev. RNA* 8, doi: 10.1002/rrna.1375 (2017).
24. Hoernes, T. P., Hüthenhofer, A. & Erlacher, M. D. mRNA modifications: dynamic regulators of gene expression? *RNA Biol.* **13**, 760–765 (2016).
25. Davalos, V., Blanco, S. & Esteller, M. SnapShot: messenger RNA modifications. *Cell* **174**, 498–499 e491 (2018).
26. Paul, M. S. & Bass, R. L. Incisor exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* **17**, 1120–1127 (1998).
27. Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **17**, 83–96 (2016).
28. Krepl, M., Otyepka, M., Banáš, P. & Sponer, J. Effect of guanine to inosine substitution on stability of canonical DNA and RNA duplexes: molecular dynamics thermodynamics integration study. J. Phys. Chem. B 117, 1872–1879 (2013).
29. Manickam, N., Josh, K., Bhatt, M. J. & Farabaugh, P. J. Effects of RNA modification on translational accuracy depend on intrinsic codon-anticodon strength. Nucleic Acids Res. 44, 1871–1881 (2016).
30. Ledoux, S. & Uhlenbeck, O. C. Different aa-tRNAs are selected uniformly on the ribosome. Mol. Cell 31, 114–123 (2008).
31. Zhang, Y. et al. A semi-synthetic organism that stores and retrieves increased genetic information. Nature 551, 644–647 (2017).
32. Hoernes, T. P. et al. Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewrite the genetic code. Nucleic Acids Res. 44, 862–866 (2016).
33. Hoernes, T. P. et al. Atomic mutagenesis of stop codon nucleotides reveals the chemical prerequisites for release factor-mediated peptide release. Proc. Natl. Acad. Sci. USA 115, E382–E389 (2018).
34. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. Nat. Biotechnol. 19, 731–735 (2001).
35. Murakami, H., Ohita, A. & Suga, H. Bases in the anticodon loop of tRNAAla,Glu prevent misreading. Nat. Struct. Mol. Biol. 16, 353–358 (2009).
36. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of RNA abundance and codon usage in Escherichia coli at different growth rates. J. Mol. Biol. 260, 649–663 (1996).
37. Grosjean, H. & Bjork, G. R. Enzymatic conversion of cytidine to lysidine in replication. Trends Biochem. Sci. 29, 165–168 (2004).
38. Bazzak, L. et al. A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res. 24, 365–376 (2014).
39. Burns, C. M. et al. Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature 387, 303–308 (1997).
40. Fritzgerald, L. W. et al. Messenger RNA editing of the human serotonin 5-HT2C receptor. Neuropharmacology 21, 825–905 (1999).
41. Niwender, C. M., Copeland, S. C., Herrick-Davis, K., Emeson, R. B. & Sanders-Bush, E. RNA editing of the human serotonin 5-hydroxytryptamine (5-HT) receptor. J. Biol. Chem. 274, 9472–9478 (1999).
42. Sommer, B., Kohler, M., Sprengel, R. & Seeburg, P. H. RNA editing in brain and its roles for an active-site adenosine and hydrated Mg2+ in pistol ribozymes. Angew. Chem. Int. Ed. Engl. 55945–5958 (2017).
43. Dutuk, A. M. et al. Design and synthesis of RNA miniduplexes via a synthetic linker approach. Biochemistry 32, 1751–1758 (1993).
44. Micura, R. Small interfering RNAs and their chemical synthesis. Angew. Chem. Int. Ed. Engl. 41, 2265–2269 (2002).
45. Kremoser, J. et al. Chemical synthesis and NMR spectroscopy of long stable isooleucine RNA. Chem. Commun. (Camb.) 53, 12939–12941 (2017).
46. Hoernes, T. P. & Erlacher, M. D. Methylated mRNA Nucleotides as Regulators for Ribosomal Translation. Methods Mol. Biol. 1562, 283–294 (2017).
47. Faserl, K., Sarg, B., Sola, L. & Lindner, H. H. Enhancing protein throughput in capillary electrophoresis-mass spectrometry by sequential sample injection. Proteomics 17, doi: 10.1002/pmic.201707310 (2017).
48. Marky, L. A. & Beeslaer, K. J. Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. Biopolymers 26, 1601–1620 (1987).
49. Santalucia, J. Jr. & Turner, D. H. Measuring the thermodynamics of RNA secondary structure formation. Biopolymers 44, 309–319 (1997).
50. Hetrick, B. et al. Polyamines accelerate codon recognition by transfer RNAs on the ribosome. Biochemistry 49, 7179–7189 (2010).
51. Bilgin, N. et al. Solution structure of the ternary complex between aminoacyl-tRNA, elongation factor Tu, and guanosine triphosphate. Biochemistry 37, 8163–8172 (1998).
52. Ledoux, S. & Uhlenbeck, O. C. [3′–32P]-labeling RNA with nucleotidyltransferase for assaying aminoacylation and peptide bond formation. Methods 44, 70–84 (2008).
53. Feinberg, J. S. & Joseph, S. Ribose 2′-hydroxy groups in the 5′-strand of the acceptor arm of P-site tRNA are not essential for EF-G catalyzed translocation. RNA 12, 580–588 (2006).
54. Wong, I. & Lohman, T. M. A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. Proc. Natl. Acad. Sci. USA 90, 5428–5432 (1993).
55. Vizcaino, J. A. et al. 2016 update of the PRIDE database and its related tools. Nucleic Acids Res. 44, 11033 (2016).

Acknowledgements
We would like to thank Nina Clementi for valuable comments on the manuscript. We also thank the entire Joseph laboratory for providing helpful discussions and a stimulating environment. This work was supported by the Austrian Science Fund (P 28494-BBL to M.D.E.). We would like to thank Nina Clementi for valuable comments on the manuscript. We also thank the entire Joseph laboratory for providing helpful discussions and a stimulating environment. This work was supported by the Austrian Science Fund (P 28494-BBL to M.D.E.).

Author contributions
T.P.H. and M.D.E. designed the experiments. T.P.H., K.F., M.A.J., J.K., C.G., E.F., X.S., M.D.E. analyzed the data. T.P.H. and M.D.E. wrote the manuscript. All authors contributed to the production of the final manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-07321-8.

Competing interests: The authors declare no competing 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.
