

ARTICLE

Received 25 Feb 2014 | Accepted 19 Jun 2014 | Published 21 Jul 2014

DOI: 10.1038/ncomms5459

# High-efficiency translational bypassing of non-coding nucleotides specified by mRNA structure and nascent peptide

Ekaterina Samatova<sup>1</sup>, Andrey L. Konevega<sup>1,2</sup>, Norma M. Wills<sup>3</sup>, John F. Atkins<sup>3,4</sup> & Marina V. Rodnina<sup>1</sup>

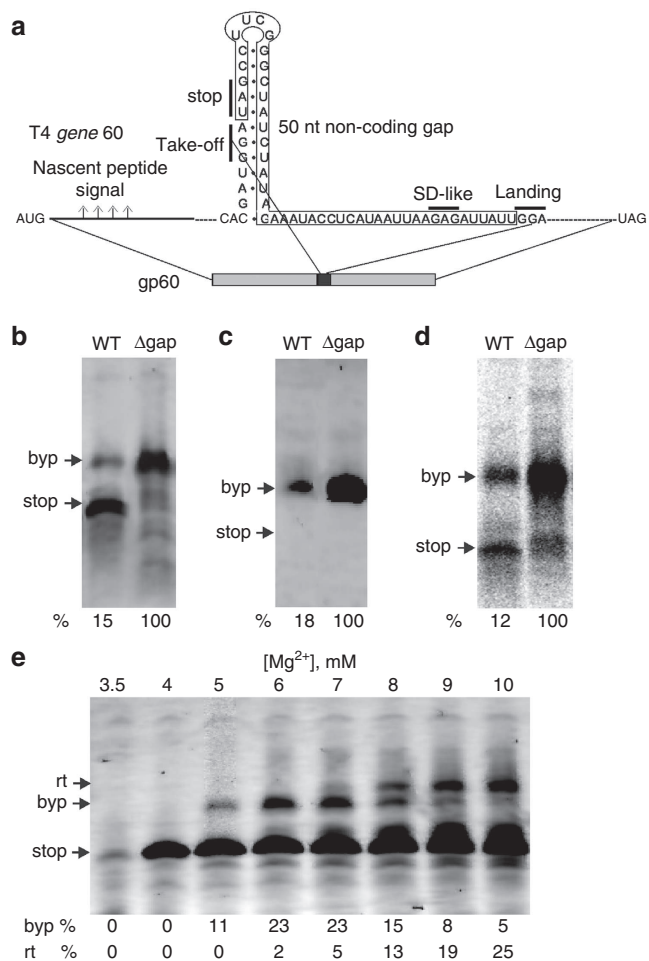
The gene product 60 (gp60) of bacteriophage T4 is synthesized as a single polypeptide chain from a discontinuous reading frame as a result of bypassing of a non-coding mRNA region of 50 nucleotides by the ribosome. To identify the minimum set of signals required for bypassing, we recapitulated efficient translational bypassing in an *in vitro* reconstituted translation system from *Escherichia coli*. We find that the signals, which promote efficient and accurate bypassing, are specified by the *gene 60* mRNA sequence. Systematic analysis of the mRNA suggests unexpected contributions of sequences upstream and downstream of the non-coding gap region as well as of the nascent peptide. During bypassing, ribosomes glide forward on the mRNA track in a processive way. Gliding may have a role not only for gp60 synthesis, but also during regular mRNA translation for reading frame selection during initiation or tRNA translocation during elongation.

<sup>1</sup>Department of Physical Biochemistry, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany. <sup>2</sup>Molecular and Radiation Biophysics Department, B.P. Konstantinov Petersburg Nuclear Physics Institute, 188300 Gatchina, Russia, and St. Petersburg State Polytechnical University, Polytechnicheskaya 29, 195251 St. Petersburg, Russia. <sup>3</sup>Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112-5330, USA. <sup>4</sup>School of Biochemistry and Cell Biology, University College Cork, Cork, Ireland. Correspondence and requests for materials should be addressed to M.V.R. (email: rodnina@mpibpc.mpg.de).

The ribosome is a molecular machine that translates the nucleotide (nt) sequence of the mRNA into the co-linear amino-acid sequence of a protein. In general, the co-linearity is maintained by sophisticated mechanisms that ensure that the ribosome moves on the mRNA track in a stepwise manner, by precisely one codon at a time, and each codon in the open reading frame (ORF) is translated into an amino acid in the protein. There are only a few cases known where the amino-acid sequence of a protein is not co-linear with the mRNA. Until recently, one well-documented case for a programmed non-co-linear mRNA translation was provided by the *gene 60* mRNA of bacteriophage T4 that codes for a subunit of a DNA topoisomerase (gene product 60, gp60). Recent evidence suggests that bypassing phenomena may be more common than assumed so far; for example, the prevalence of ribosome bypassing was demonstrated in mitochondria of the yeast *Magnusiomyces capitatus*<sup>1</sup>, and there are further strong candidates for novel cases of translational bypassing<sup>2</sup>.

During translation of gp60 mRNA, the ribosome produces a single polypeptide chain from a discontinuous reading frame<sup>3,4</sup>. The interrupted ORF is unique to the T4 genome and is caused by the disruption of *gene 39*, which encodes the DNA topoisomerase subunit in other T-even bacteriophages<sup>5</sup>, by the insertion of a mobile DNA element<sup>6</sup>. The mobile element splits *gene 39* into a sequence coding for the N-terminal part of gp39 and the newly formed *gene 60*, which codes for the C-terminal portion of the topoisomerase. One consequence of the mobile element insertion is the addition within the *gene 60* coding sequence of a 50-base pair foreign sequence<sup>6</sup>, which is copied into the *gene 60* mRNA. The non-coding sequence is not spliced but is skipped during translation<sup>3,4</sup>. The ribosome translates the first 46 codons of *gene 60* mRNA up to the glycine codon GGA followed by a stop codon UAG (Fig. 1a). However, instead of terminating translation at the stop codon, peptidyl-tRNA<sup>Gly</sup> disengages from pairing with the GGA codon 5' adjacent to the coding gap, is retained within the ribosome as it slides over the 50-nt non-coding gap and re-pairs to mRNA at the triplet, also GGA, 5' adjacent to the coding resumption site<sup>7</sup>. Disengagement is referred to as occurring at the 'take-off' codon with re-pairing to mRNA being at the 'landing site' codon, which is complementary to the anticodon. The take-off codon is 5' adjacent to UAG and both are near the top of a stem loop (SL), the 'take-off SL' that is critical for bypassing<sup>4,7-10</sup>. The analysis of the mRNA structure suggested that the gap region is structurally autonomous from the upstream and downstream coding regions<sup>10</sup>, consistent with the evolutionary origin of the gap from a mobile element<sup>6</sup>. The presence of the take-off GGA and the matching landing codons, the UAG stop codon next to the take-off GGA and the SL structure in the 5' gap are essential for bypassing. Also the identity of peptidyl-tRNA<sup>Gly</sup> that remains bound to the ribosome during bypassing is crucial, as other tRNAs support bypassing to different, but generally substantially lesser extent<sup>11,12</sup>. A short potential Shine-Dalgarno (SD)-like GAG sequence 6 nt 5' of the GGA landing site contributes to the precision of landing<sup>9</sup>. Furthermore, part of the nascent peptide is important for bypassing efficiency<sup>4</sup>, with preliminary evidence pointing to the significance of a KKYK motif<sup>13</sup>. Finally, genetic experiments with mutants containing an enlarged SL indicated an important role of ribosomal protein L9 (ref. 14), which subsequently was shown to have a role in normally restraining forward mRNA slippage<sup>11</sup>. These diverse stimulatory elements are likely to provide a complex regulatory network of physical interactions and, presumably, kinetic effects that influence bypassing<sup>15</sup>.

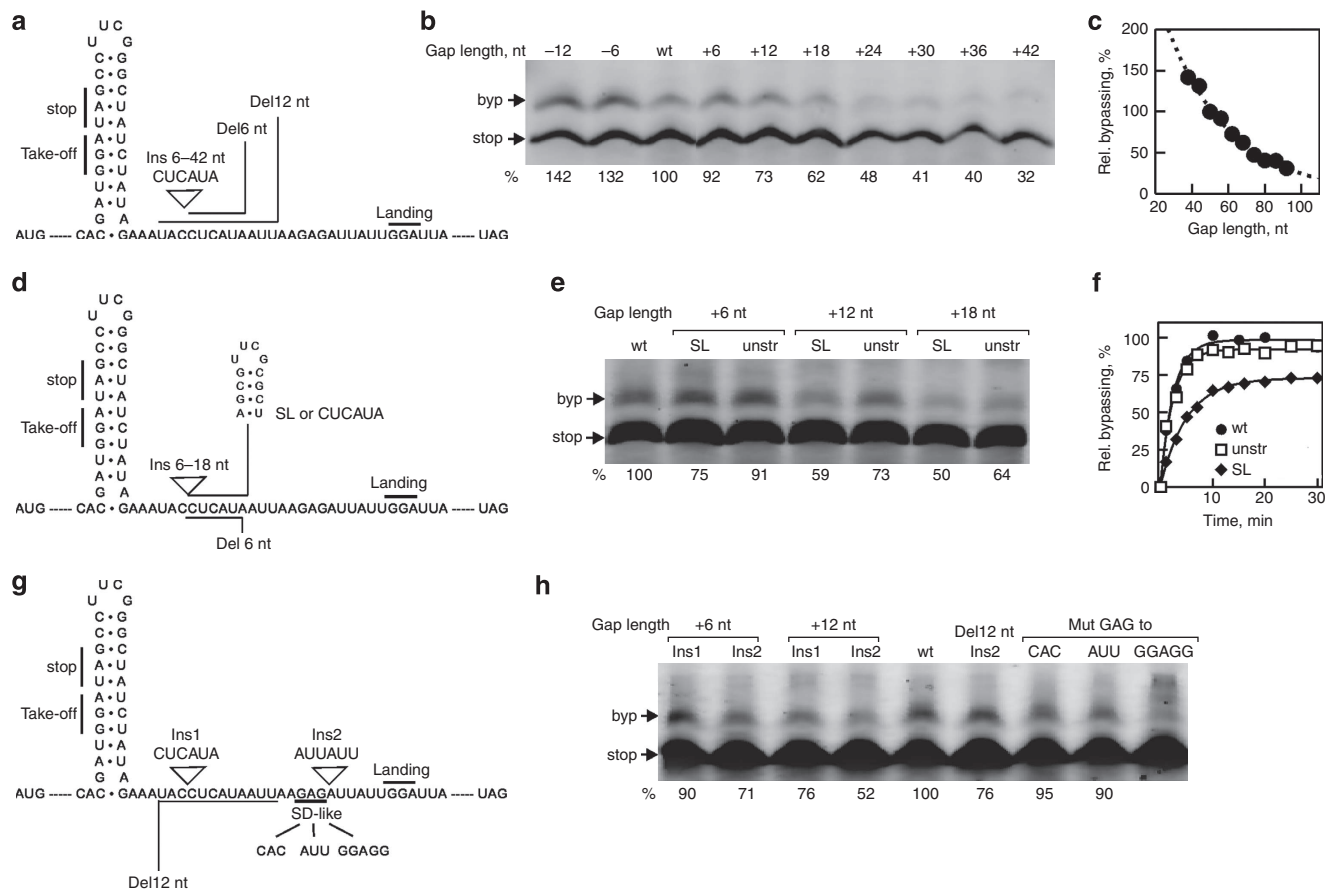
The exact mechanism of how the ribosome traverses the gap sequence and why it lands at the particular landing codon is not known. Important unresolved questions are how and why



**Figure 1 | Peptide products upon translation of *gene 60* mRNA *in vitro*.**

(a) Schematic of phage T4 *gene 60* mRNA and the respective full-length peptide product gp60. The take-off and landing codons (GGA), the stop codon (UAG) and the potential SD-like sequence (GAG) preceding the landing codon are indicated; the putative enhancing signals of the nascent peptide are shown by arrows. The non-coding 50-nt gap encompasses the sequence between the UAG stop codon and the codon preceding the landing GGA. The resulting product, the topoisomerase subunit gp60, is composed of the N-terminal part (grey area) corresponding to the mRNA sequence upstream of the internal stop codon, a Gly encoded by the take-off and landing site codons (black square), and the C-terminal part (grey area) encoded by the downstream mRNA sequence; the N- and C-terminal parts are not shown to scale. (b-d) Products of the *in vitro* translation of a 455-nt-long fragment of *gene 60* mRNA. Proteins were separated by SDS-polyacrylamide gel electrophoresis and detected using the fluorescence of the N-terminal BPY-Met (b); the C-terminal LUMIO-tag (c); or radioactivity of [<sup>14</sup>C]Leu in peptides (d); note that there is a single Leu residue in the N-terminal part of gp60 preceding the Gly at position 46, compared with five Leu residues in the full-length product. WT, the construct containing the gap region.  $\Delta$ gap, the construct in which the gap region was deleted and the N- and C-terminal parts of gp60 are fused in frame. (e) Effect of Mg<sup>2+</sup> concentration on the efficiency and fidelity of bypassing. stop, the N-terminal 46 amino acids long peptide terminated at the UAG stop codon immediately following the take-off GGA; byp, the protein product of bypassing (88 amino acids); read-through (rt), the protein product synthesized when an amino acid is incorporated in place of stop codons (105 amino acids). For each translation experiments, two independent repeats were carried out ( $n = 2$ ).





**Figure 4 | Variations in the coding gap.** (a) Sequence of the coding gap region. Arrowheads indicate the position where the (CUCAUA)<sub>n</sub> cassette (1 ≤ n ≤ 7) was inserted (Ins6–42 nt). Deletions of 6 or 12 nucleotides between the SL structure and potential SD-like GAG in the non-structured region are indicated (Del6 nt and Del12 nt). (b) Effect of insertions and deletions in the gap sequence. Changes introduced into the gap sequence are indicated on top, the bypassing efficiency relative to the WT construct at the bottom. (c) Bypassing efficiency as a function of the gap length. Smooth line represents the results of exponential fitting (see text). (d) Introducing structures in the gap. The CUCAUA sequence was deleted (Del 6 nt) and replaced by either a 12-nt step-loop structure (SL; in one, two or three copies) or by an unstructured CUCAUA sequence of the same length as the respective SL constructs (Ins 6–18). (e) Changes in bypassing by increasing the gap length by 6 (+6 nt), 12 (+12 nt) or 18 (+18 nt) nucleotides; the inserted fragment contained either an SL element or a presumably unstructured CUCAUA repeats (unstr) as indicated. The bypassing efficiency with the WT construct was set to 100%. (f) Time courses of synthesis of the bypassing product with the WT mRNA (circles), and mRNAs (+6 nt) containing an additional CUCAUA sequence (open squares) or a SL structure (diamonds). (g) Insertions upstream (Ins1) or downstream (Ins2) of the SD-like GAG codon and the replacements of GAG to CAC, AUU (weak SD sequences) or GGAGG (strong SD sequence). To change the position of the GAG sequence relative to the landing codon, 12 nt were deleted where indicated and inserted at position of Ins2 (Del12/Ins2). (h) Bypassing on mRNAs with mutations in the SD-like sequence (n = 8).

byp was validated using a construct in which the two reading frames were fused by omission of the bypassing gap ( $\Delta$ gap). The correct reading frame of the byp was verified using the C-terminal Lumio-tag introduced in-frame with gp60 (Fig. 1c). The efficiency of translation bypassing was quantified using the fluorescence of Bpy or Lumio, and by autoradiography based on [<sup>14</sup>C]Leu incorporated into the WT or  $\Delta$ gap products (Fig. 1d). According to the three reporters (Fig. 1b–d), and comparing with the WT control in the experiments described below (Fig. 1e and 4b, etc.), bypassing was observed on about 15–25% of translating ribosomes, comparable to bypassing efficiencies measured *in vivo*. These results suggest that the synthesis of gp60 in a well-defined reconstituted *in vitro* translation system results in efficient bypassing, which does not require any additional auxiliary factors. Thus, the major signals required for translational bypassing must reside in the mRNA itself.

Both efficiency and fidelity of bypassing were further tested by varying the Mg<sup>2+</sup> concentration (Fig. 1e). The Mg<sup>2+</sup> concentration is crucial for speed and fidelity of decoding<sup>16</sup>; however, its

effect on recoding events, such as bypassing, is not known. At low Mg<sup>2+</sup> concentration, the *gene 60* mRNA was translated up to the stop codon at the beginning of the gap, but the full-length byp was not formed (we have added 3.5–4.0 mM Mg<sup>2+</sup>, corresponding to about 1.0–1.5 mM of free Mg<sup>2+</sup>, taking into account the Mg<sup>2+</sup> bound to GTP and phosphoenol pyruvate present in the reaction mixture<sup>16</sup>; see Methods). At higher concentrations of Mg<sup>2+</sup> (5–7 mM added), the byp was formed (23%). When the Mg<sup>2+</sup> concentration was increased further, additional translation products appeared, including the product of the stop-codon read-through (upper band, ‘rt’), suggesting that the fidelity of decoding was compromised.

Efficiency and fidelity of translation and bypassing can be additionally tuned by polyamines (Supplementary Fig. 1). Generally, at low Mg<sup>2+</sup> concentrations, the addition of spermidine and putrescine facilitated bypassing; however, the read-through efficiency was increased as well. Spermine induced significant misreading even at low concentrations. Given the complex interplay between Mg<sup>2+</sup> and polyamines and the

narrow optimum in polyamine concentrations, we chose to work without polyamines at conditions of maximum efficiency and fidelity of recoding (7 mM added  $Mg^{2+}$ , which corresponds to about 4.5 mM free  $Mg^{2+}$  (ref. 16).

**Role of the nascent peptide.** Having established conditions for efficient bypassing, we next examined the roles of distinct regions of the *gene 60* mRNA and of the nascent peptide. To study the effect of the N-terminal part of the nascent peptide, we deleted the mRNA sequence coding for amino acids 1–13 (Nascent Peptide construct 1, NP1; Fig. 2a); the deletion reduced the bypassing efficiency by >50%, in agreement with an earlier report<sup>4</sup>. In order to distinguish effects potentially caused by alterations in the mRNA structure from those of amino-acid replacements, we introduced non-synonymous mutations (NP2) or synonymous replacements by other frequent codons (NP3). Both variants supported efficient bypassing, suggesting that the amino-acid composition of the N-terminal part of the nascent peptide, as well as any secondary structure of the respective mRNA region, did not affect bypassing. Together with the results of the deletion experiment, these data suggest that the contribution of the N-terminal part of gp60 may be unspecific, for example, contributing to anchoring the peptidyl moiety of peptidyl-tRNA<sup>Gly</sup> during traversing the gap, when the tRNA on the ribosome is not stabilized by codon–anticodon interaction.

Following previous reports<sup>4,9,17</sup>, we next tested the effect of non-synonymous (NP4 and NP5) and synonymous (NP6 and NP7) replacements introduced in codons 14–21 and 22–30 (Fig. 2b). Amino-acid exchanges inhibited bypassing, whereas synonymous replacements had very little effect (Fig. 2b), indicating that the sequence of the nascent peptide, rather than the mRNA sequence, was important. We note that alterations in the mRNA sequence resulted in decreased overall translation levels (see the stop bands in NP4 and NP6). As none of the synonymous replacements involved rare codons, the generic effect on translation probably reflects an altered mRNA secondary structure.

To further identify the amino-acid positions that are critical for bypassing, we constructed mRNAs with single amino-acid replacements at each position between amino acids 14 and 30 positions (Fig. 2b,c). Given the length of the ribosomal peptide exit tunnel<sup>18</sup>, amino acids 14–30 and the following 17 amino acids of the nascent peptide up to the take-off Gly should reside within the ribosome tunnel. The mutations in the KKYK motif resulted in about 50% inhibition, mostly due to Y16, whereas substitutions of Q19, V22, R23 and S28 decreased the bypassing efficiency by 60–70%, whereas R24, S25, K27 essentially abolished bypassing. Secondary structure predictions suggest that the 14–30 region of the nascent peptide has a high propensity of helix formation with amino acids Y16, Q19, R23, R24, K27 and S28 found on the same face of the potential  $\alpha$ -helix (Fig. 2c, inset). The presumed  $\alpha$ -helix may interact with the exit tunnel wall, thereby further stabilizing peptidyl-tRNA<sup>Gly</sup> binding while the ribosome traverses the gap. When peptidyl-tRNA<sup>Gly</sup> is formed, the length of the nascent peptide is about 20 amino acids, placing the exchanged residues in the middle part of the peptide exit tunnel, beyond the constriction formed by ribosomal proteins L4 and L22 (ref. 19). Although it is difficult to predict the exact location of the nascent peptide 14–30 region in the tunnel, we tested whether mutations of the proteins L4 and L22, as well as of ribosomal protein L23, which resides at the tunnel exit and is involved in signalling from inside the tunnel to the outside<sup>20</sup>, affect bypassing. Deletion in the L23 signalling loop<sup>20</sup> has very little effect on bypassing (Fig. 2d). Deletion of the L22 loop consisting of residues 84–97, which forms part of the constriction,

reduced bypassing by 50%. Additional deletions of constriction site residues in L4 (residues 63–64), did not have any further effect (Fig. 2d). The effect of one large deletion from L22 ( $\Delta$ 84–97) on bypassing may be explained by assuming that this deletion removed (part of) the potential interaction partner for the critical residues in the 14–30 region of gp60. Thus, L22 may constitute one of the contacts; however, further interaction partners (for example, 23S rRNA) may be involved.

**Putative elements upstream of the take-off site.** Next, we asked whether the mRNA region following nt 90 (after codon 30) but preceding the SL element in the 5' part of the gap region (nt 91–138) affects bypassing. Prediction of the mRNA structure using mfold and the results of chemical probing<sup>10</sup> suggested the existence of a SL structure upstream of the gap (Fig. 3a; Supplementary Fig. 2a,b). Introducing two synonymous replacements in the upper part of the putative SL (2M) or four synonymous replacements (4M) that should disrupt the SL essentially abolished bypassing (Fig. 3b), which suggested a critical role of the mRNA sequence upstream of the gap. Introducing compensatory mutations, which were expected to restore the secondary structure of the upper part (2MC), did not restore bypassing, whereas compensatory mutations for the whole SL (4MC) resulted in partial rescue. The latter observation suggests that the secondary structure of the lower part of the region is important. Why compensatory mutations in the upper part did not restore bypassing is more difficult to explain. One reason may be that the mutations induced formation of an alternative secondary structure, which did not support bypassing. It is also possible that the secondary structure of the region upstream of the gap is different in solution and on the ribosome, such that 'compensatory' mutations in the upper part of the SL do not restore the secondary structure. Alternatively, a specific sequence in the region may be important, for example, for interactions with the ribosome. A structure of the mRNA bound to the ribosome during bypassing would be required to distinguish between these possibilities.

We also sought to disrupt putative 5' SL structures by adding two complementary oligonucleotides (Fig. 3a). The oligonucleotides did not affect translation as such, because the ribosome can efficiently dissolve duplexes of moderate stability<sup>21</sup>. However, bypassing was abolished (Fig. 3c), underscoring the importance of mRNA structure in the region upstream of the gap.

**The non-coding gap sequence.** The crucial role for translational bypassing of the SL structure formed by the GGA codon, the stop codons and the 5' part of the non-coding gap region is well documented<sup>4,7–9,17</sup>. In comparison, the importance and structure of the 3' part of the gap are less clear. We varied the length of the 3' gap by deleting 6 or 12 nt or inserting up to seven copies of the unstructured 6-nt CUCAUA cassette, which extended the 3' gap region by 6 to 42 nt; in these experiments the SD-like sequence GAG remained at its original distance from the landing GGA codon (Fig. 4a). Shortening the 3' gap increased bypassing, whereas increasing the length of the gap reduced bypassing (Fig. 4b), and the yield of byp decreased monotonously with increasing gap length (Fig. 4c). Such a dependence can be described by an exponential function ( $Y = A \times R^n$ ), where  $Y$  is the observed bypassing efficiency,  $n$  is the gap length in nucleotides,  $A$  is the portion of ribosomes initiating bypassing at the take-off site and resuming translation once they reached the landing codon and  $R$  is the processivity coefficient, that is, the efficiency of traversing a single nucleotide in the gap. The fit yields  $A = 440 \pm 30\%$  and  $R = 0.972 \pm 0.001$ . Given the absolute bypassing efficiency on the WT construct (22%), the fraction of

the ribosomes that take part in take-off and landing (provided they reached the landing codon) is close to one ( $440\% \times 0.22$ ). This suggests that all ribosomes that initiated bypassing continued translation after landing, and  $R = 0.972$ . This implies that at each step  $>97\%$  of ribosomes moved forwards, whereas only  $<3\%$  dissociated from the track or lost their peptidyl-tRNA<sub>2</sub><sup>Gly</sup>. The uniformity of sliding suggests that ribosomes traversing the gap are not affected by pausing because of mRNA secondary structures or other potential stalling interactions with the mRNA, such as attempts of codon reading.

We also tested whether inserting of defined SL structures into the 3' part of the gap affected bypassing. Six nucleotides (CUCAUA) were deleted as shown in Fig. 4a (Del6) and instead a 12-nt SL was inserted in one, two or three copies (SL +6 nt, +12 nt and +18 nt, respectively; Fig. 4d). For comparison, analogous constructs were made with unstructured sequences (unstr) by inserting a CUCAUA cassette. Introducing a SL motif in up to three copies had a moderate inhibitory effect (Fig. 4e). Assuming that all ribosomes could take-off and land (which appears a reasonable assumption, because take-off and landing sites were not altered), the average processivity coefficient was  $R = 0.969$  for the mRNAs containing three copies of the SL motif, compared with 0.972 for the unstructured constructs. With the WT mRNA and the +6 nt mRNA construct containing an unstructured CUCAUA sequence, time courses of gp60 synthesis were identical, with half-life time  $\tau_{1/2} = 2.5$  min (Fig. 4f). In the presence of the SL structure in the gap, this half-life time increased to  $\tau_{1/2} = 5$  min, suggesting that structures in the 3' part of the gap slow down the appearance of the byp.

Furthermore, we examined whether the position of the SD-like codon GAG affects bypassing. The insertion of a CUCAUA sequence in one or two copies upstream of the GAG, duplication of an unstructured sequence AUUAUU downstream of GAG or deletion of 12 nt (CUCAUA) preceding the GAG codon had a moderate effect, which did not exceed 50%, suggesting that the positioning of the GAG sequence relative to the landing site is not crucial (Fig. 4h). Replacements of GAG with CAC or AUU sequences, which cannot engage in SD-aSD (anti-Shine Dalgarno) interactions, had essentially no effect on bypassing, whereas introducing a strong SD sequence, GGAGG, resulted in the synthesis of alternative, longer products. The latter products may result from landing on the two additional GGA codons that are introduced with the enhanced SD sequence<sup>9</sup>.

**Bypassing signals downstream of the landing site.** Previous work indicated that replacing the *gene 60* sequence following the landing site by reporter constructs did not abolish bypassing<sup>4,22</sup>. On the other hand, the bypassing efficiency on truncated *gene 60* mRNA in cell lysates was strongly diminished<sup>10</sup>. To assess the minimum length required for bypassing, we tested several mRNA constructs that were truncated at different positions downstream of the landing codon (Fig. 5a). The shortest construct that was efficient in bypassing contained 69 nt following the landing site (Fig. 5b) and included the whole length of a potential secondary structure element (3'SL) predicted by computational analysis (mfold) and chemical probing<sup>10</sup>. Translation of the T0 control mRNA, which had the same overall length as the WT construct but contained a stop codon at the indicated position (Fig. 5a), resulted in formation of a peptide distinct from the stop product (Fig. 5b), which shows that the difference in the size between the stop product and of the shortest expected byp from the T1 construct should have been sufficient to resolve the two translation products on the gel.

Next, we tested which features of the 3'SL structure are essential (Fig. 6a). A single G-to-C mutation 64 nt downstream of

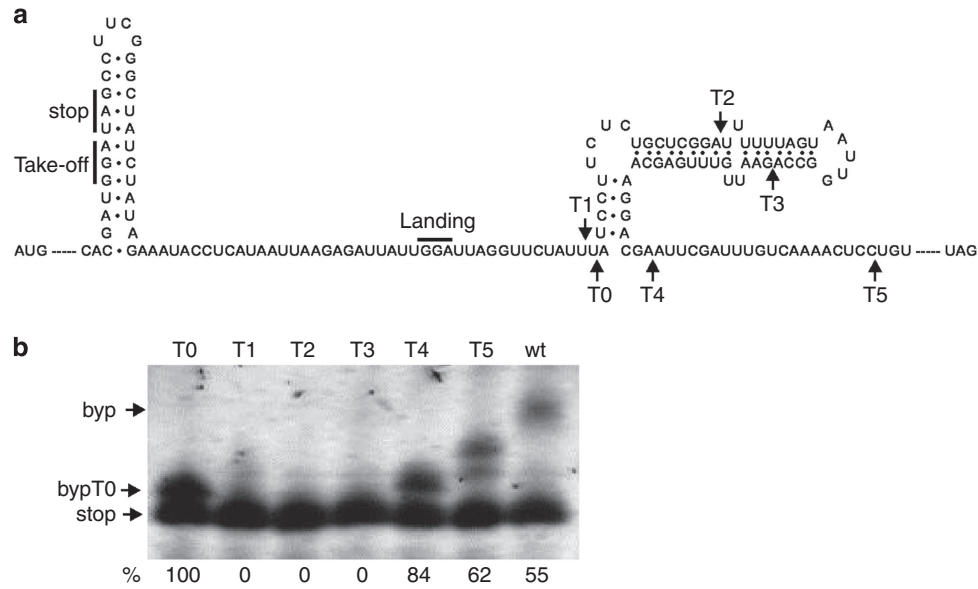
the landing codon, which is expected to weaken the lower stem of the 3'SL, reduced the bypassing efficiency by 50% (Fig. 6b). Deletion of a 6-nt GGUUCU sequence four nucleotides after the landing site (Del1) did not diminish bypassing (Fig. 6c); however, additional translation products appeared, suggesting that the fidelity of landing was compromised. In contrast, insertion of a CCGCCG sequence 3 nt downstream of the landing site (Ins1), deletion of a part of the putative secondary structure element (Del 2) or a single U to AA mutation that is expected to distort the local mRNA structure (Mut2) all abolished bypassing (Fig. 6c). Surprisingly, removing the putative secondary structure element (Del3) or inserting the same sequence in reverse direction (Rev1) had almost no effect (Fig. 6d). Although a 3'SL element probably can form even when the sequence is reversed (Rev1), efficient bypassing in the absence of the presumed 3'SL structure (Del3) can only be explained by the formation of alternative mRNA structures. Mfold predictions suggested that Del3 mRNA may form an alternative secondary structure, which is only four nucleotides closer to the landing codon (Supplementary Fig. 2c), which does not abolish bypassing (see Del1). In contrast, Mut2 and Del2 mutations appear to disfavour the WT 3'SL structure and may form alternative helices, which are further away from the landing site (Supplementary Fig. 2d). By analogy with the Ins1 mutant, such a shift in the position of the putative helix should abolish bypassing. Thus, efficient bypassing may depend on the presence of a secondary structure element within a particular distance from the landing site.

To further test the importance of secondary structure, we replaced the mRNA sequence following the landing codon by an unstructured stretch consisting of 17 UUC repeats ((UUC)<sub>17</sub>). With that mRNA, no byp of the expected length was observed. Instead, some long, low-abundance peptides appeared, which may stem from incorrect landing at the glycine codons preceding the landing codon. These results suggest that the mRNA region downstream of the coding gap is important for bypassing and that secondary structure formation is crucial. However, there is no stringent requirement for one particular structure of the 3'SL, as different alternative secondary structures appear to support bypassing.

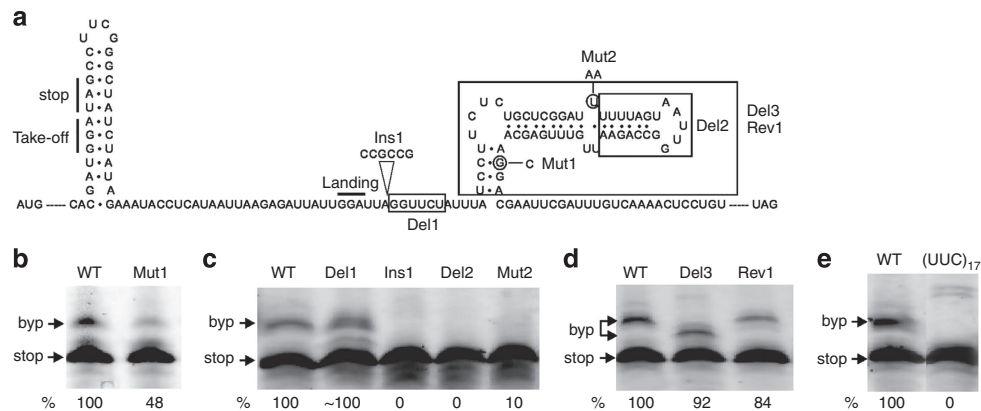
## Discussion

The present experiments recapitulate translational bypassing in a fully reconstituted translation system consisting of purified ribosomes, translation factors and aminoacyl-tRNAs. Our results indicate that all major determinants for bypassing are specified by the *gene 60* mRNA itself and the efficiency of synthesis of the full-length product depends on the interplay between signals in the mRNA and the ability of the ribosome to slide over the coding gap in a processive way. We found that the nascent peptide and the mRNA structure upstream of the gap region (5'SL) and downstream of the landing site (3'SL) have an important role (Fig. 7). Mutations of any of these elements led to a strong reduction in the bypassing efficiency. The relative contributions of each of these elements *in vivo* may vary, depending on genetic background<sup>22</sup> and changes in the mRNA or tRNA<sub>2</sub><sup>Gly</sup> (refs 4,9,11,17,23) or on the ribosome elements that are known to affect bypassing, such as protein L9 (refs 11,14).

Our data suggest the following mechanism of translational bypassing (Fig. 7). When translating the 5' ORF up to the take-off GGA codon, the ribosome has to unwind the 5'SL upstream of the gap and the SL element in the 5' part of the gap region. With the take-off GGA in the P site, the whole length of the 5'gap SL and a large part of the downstream 5'SL are probably occluded in the mRNA-binding tunnel of the ribosome, and peptidyl-tRNA<sub>2</sub><sup>Gly</sup> is bound in the P site. For the mRNA to become mobile,



**Figure 5 | Assessment of the minimum mRNA length required for bypassing.** (a) Sequence and putative secondary structure of the downstream part of mRNA. Positions of mRNA truncations (Truncations T1-T5) are indicated by arrows. T0, control construct of the same length as the parent mRNA (WT), but with a stop codon after the landing site preceding the putative downstream mRNA structured element. (b) Bypassing efficiency with the truncated mRNA constructs ( $n = 2$ ).

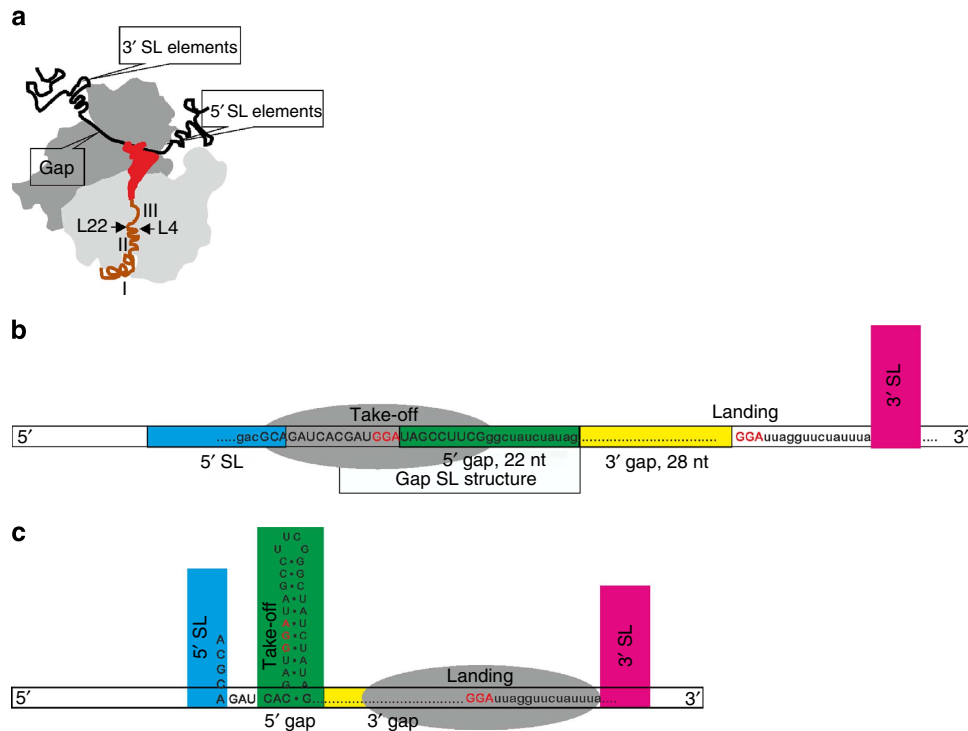


**Figure 6 | Structure requirements for the mRNA region downstream the landing codon.** (a) Sequence and putative secondary structure of the region downstream the landing site (from mfold). Positions of mutations (Mut1 and Mut2), deletions (Del1 and Del2) and insertion (Ins1) are indicated. (b) Effect of a single G to C replacement 65 nt downstream of the landing codon. (c) Effects of deletions, insertions and mutations in the 3'SL. (d) Deletion of the putative mRNA secondary structure element (Del3) or insertion of the same element in a reverse direction (Rev1). (e) Replacement of the mRNA sequence downstream of the landing codon by a poly(UUC)<sub>17</sub> sequence. Efficiency of translational bypassing with the WT mRNA construct was set to 100% ( $n = 8$ ).

the codon-anticodon duplex must be disrupted. This may be brought about by a conformational rearrangement of the ribosome, for example, the formation of a highly rotated chimeric intermediate similar to that formed during EF-G-catalysed translocation<sup>24-26</sup> or frameshifting<sup>27</sup>. Although the codon-anticodon interactions are weakened, the retention of peptidyl-tRNA<sup>Gly</sup> in the ribosome is ensured by the interactions of the nascent peptide with the ribosome, presumably through both unspecific anchoring of the N-terminal part of gp60 emerging from the ribosome peptide exit tunnel and specific interactions of the part of nascent peptide residing in the exit tunnel. Ribosome sliding over the gap is uniformly processive and rapid. The high speed and processivity of sliding is consistent with the idea that the ribosome does not systematically scan the codons in the gap<sup>9</sup>, as scanning would require multiple attempts of forming codon-anticodon interactions in the P site, which—judging from codon

recognition in the P site during initiation—is a relatively slow process<sup>28</sup>. Our data suggest that efficient landing requires that the mRNA downstream of the landing site is structured. Presumably, the 3'SL stalls the ribosome, thereby facilitating the landing. At the same time, the 5'SL in the coding part of the mRNA and the SL element in the 5' part of the gap leave the ribosomal mRNA tunnel and may re-fold, thereby preventing backward sliding of the ribosome (Fig. 7).

The exact role of the 5'SL upstream of the gap region is not known. We speculate that it may help to maintain the directionality of ribosome gliding over the gap by re-forming the 5'SL as soon as the ribosome vacates the respective mRNA stretch. The 5'SL may also help re-folding or maintaining the SL in the 5' part of the gap, which is essential for bypassing. The 5'SL alone or in combination with the SL element in the 5' gap may interact with protein L9, thereby contributing to the processivity



**Figure 7 | mRNA-encoded signals for translational bypassing in T4 gene 60 mRNA.** (a) Schematic of the ribosome with putative mRNA secondary structures upstream of the gap (5'SL), the gap region and putative structures downstream the gap (3'SL) at the moment when bypassing is initiated. Peptidyl-tRNA<sup>Gly</sup> is shown in red. The nascent peptide (46 amino acids, brown) can be divided into three regions. Amino acids 1–13 (region I) may anchor the nascent peptide to the surface of the ribosome. Amino acids 14–30 (region II) are located within the exit tunnel after the constriction formed by ribosomal proteins L4 and L22; amino acids on one side of the putative peptide helix within the tunnel have a strong effect on bypassing. Amino acids 31–47 (region III) reside in the cavity between the peptidyl transferase region and the L4–L22 constriction. The 30S and 50S ribosomal subunits are shown in dark and light grey, respectively. (b) The position of the ribosome (grey oval) at the onset of bypassing. 5'SL and 3'SL structures are shown in blue and red, respectively, the 5' and 3' parts of the gap sequence are green and yellow, respectively. A large part of the 5'SL, and the SL part of the gap are unfolded (indicated by horizontal bars) and reside in the mRNA channel of the ribosome. The 3' part of the gap region is predominantly unfolded as well (horizontal bar), whereas the 3'SL element is structured<sup>10</sup> (vertical bar). (c) Upon movement of the ribosome onto the landing site, the 5'SL upstream of the gap and the 5'SL in the gap re-fold (vertical bar), whereas the ribosome covers most of the unstructured 3' part of the gap up to the presumably structured 3'SL (vertical bar).

of sliding or the fidelity of landing codon selection. Finally, it may affect the function of L1, the protein which influences the retention of tRNA<sup>29</sup> and the formation of a particular, over-rotated conformation of the ribosome<sup>27</sup>. Previous *in-vivo* experiments did not implicate the region immediately upstream the gap in the mechanism of bypassing either because the effect of specific mutations was not tested directly or because *in vivo* the effect is mimicked by contributions of other elements, which are more important under *in-vivo* conditions. In this context, we note an as yet unexplained effect on bypassing of sequences at the ribosome-binding site 5' to the start codon<sup>22</sup>, which was interpreted in terms of a potential influence of neighbouring ribosomes in a polysome. The 5'SL may take over this role on the leading ribosome or when the density of ribosomes on the gene60 mRNA is low.

The role of the 3'SL downstream the gap is puzzling. Our present experiments and recent *in-vitro* translation experiments performed in S30 extracts<sup>10</sup> showed that essentially no bypass was observed with mRNAs that were truncated shortly after the coding gap. These findings appear to contradict previous results obtained using different reporters *in vivo*, which suggested that translational bypassing was observed even when the sequence beyond the fifth nucleotide after the landing codon was replaced by the reporter sequence<sup>4</sup>. Also in our translation system, the deletion of nt 264–316 in the 3' region did not affect bypassing when the remaining mRNA sequence was present. One possible

explanation for these observations is that the bypassing signals in the 3' coding region are not sequence-specific and can be provided by different mRNA sequences, which expose an mRNA hairpin located at a particular distance from the landing codon. A steric hindrance provided by the secondary structure element may stimulate landing, although further experiments will be required to confirm this model.

The present data provide the first insight into the kinetic mechanism of translational bypassing. Synthesis of the peptide up to the take-off glycine (47 amino acids) is completed within less than 1 min (not shown). We did not observe a prolonged pausing at the UAG stop codon, which may explain why under normal conditions the efficiency of bypassing is not defined by the competition between termination and take-off, although at elevated concentrations release factor 1 can reach the UAG codon and termination can compete with bypassing<sup>8</sup>. The efficiencies of both take-off and landing (for those ribosomes that reached the landing site) are close to 1 (Fig. 4). However, a part of the ribosomes (and/or tRNA) may be lost from the mRNA track during sliding. At any time, >97% of ribosomes successfully move by one nucleotide towards the landing site, whereas <3% are lost due to drop-off. The loss of a small portion of active sliding complexes at each nucleotide step explains why the bypassing efficiency decreases with increasing gap length.

From the length of the gp60 fragment synthesized in the present experiments (90 amino acids), and the half-time of



synthesis (150 s), the estimated average translation rate of the byp is about 0.6 amino acid/s. The average translation rate is not altered when the gap length is increased by insertion of a non-structured sequence element (Fig. 4h). This suggests that the time required for traversing the gap is small compared with the time required for the synthesis of the peptide. In contrast, when a short SL structure was inserted, the average rate of synthesis decreased to 0.3 amino acid/s. Thus, even a thermodynamically weak SL structure ( $\Delta G^\circ$  of the inserted helix is  $-5 \text{ kcal mol}^{-1}$ , as calculated by mfold) significantly retards the movement through the gap, thereby affecting the overall synthesis rate. Despite the kinetic effect, inserting a defined SL structure in the 3' gap part reduced the bypassing efficiency by only 15%. Similarly, chemical probing suggest that the 3' end is predominantly unstructured in solution, and weak potential secondary structure elements that may exist are not important for bypassing<sup>10</sup>. These results argue against the existence of a defined mRNA secondary structure in the 3' part of the coding gap region, which is spooled into the A site during the movement of the ribosome over the gap. When moving along the mRNA, the ribosome does not recognize alternative GGA codons within the gap<sup>9</sup>. The SD-like sequence has a moderate effect but may be important for the fidelity of landing site selection, particularly in the context of other mutations<sup>8,9,23</sup>.

Taken together with previous genetic work and mRNA structural probing, the present biochemical and mutational analyses provide a detailed insight into the signals for translational bypassing. Nevertheless, crucial details of the mechanism remain obscure, such as why ribosomes start to slide, or whether the conformation of ribosomes changes during bypassing, or the role of protein L9 and the structural dynamics of tRNA<sub>2</sub><sup>Gly</sup>. Furthermore, because many different signals contribute, the pathway by which efficient bypassing is stimulated may depend on experimental details, for example, the nature of reporter constructs, expression levels or the presence of mutations. Regardless of the exact kinetic pathway, the existence of efficient bypassing demonstrates how the ribosome—guided by elements specified by the mRNA—is able to overcome the fundamental colinearity rule of translation. The ability of the ribosome to glide over the mRNA track must have evolved in response to the needs of normal translation. The conformation of the ribosome, which is stabilized during translational bypassing, may represent one of the intermediates during tRNA–mRNA translocation or upon sliding over the mRNA during translation initiation, for example, on mRNAs lacking an SD sequence or on polycistronic mRNAs. Future progress in understanding the kinetics of translational bypassing will show whether the ribosome's capacity to slide is more commonly used than thought so far and whether it can be utilized for engineering programmed bypassing constructs for the controlled expression of designer proteins.

## Methods

**Materials.** Biochemicals were from Merck, nucleotide triphosphates from Jena Bioscience. The following oligonucleotides were used in the experiment of Fig. 3: Oligo 1, 5'-CATTGCATAGTTCATTGAAGAGGATTTAATAGAACG-3'; Oligo 2, 5'-ATAGTTCATTGAAGAGGATTTAATAGAACG-3'. [<sup>3</sup>H]methionine and [<sup>14</sup>C]leucine were from Hartmann Analytics. Bodipy-FL-succinimidyl ester and Lumio Green Detection Kit were from Invitrogen. Total *E. coli* tRNA was from Roche and oligonucleotides were from IBA. 70S ribosomes, EF-Tu, EF-G, IF1, IF2, IF3, [<sup>3</sup>H]Met-tRNA<sup>Met</sup> and [<sup>3</sup>H]Met-tRNA<sup>Met</sup> were prepared from *E. coli*<sup>30–32</sup>. Site-directed mutagenesis of the WT gene 60 construct<sup>9</sup> was performed using the Quickchange PCR protocol with the appropriate oligonucleotide primers (Supplementary Table 1). mRNAs were produced by T7 RNA-polymerase *in-vitro* transcription and purified by ion-exchange chromatography on HiTrap Q HP 5 ml column (GE Healthcare). To prepare fluorescence-labelled [<sup>3</sup>H]Met-tRNA<sup>Met</sup> (Bpy-Met-tRNA<sup>Met</sup>), HPLC-purified [<sup>3</sup>H]Met-tRNA<sup>Met</sup> (30  $\mu\text{M}$ ) was incubated with BODIPY-FL sulfosuccinimidyl ester (Bpy-SSE; Invitrogen, D6140; 4 mM) in 20 mM HEPES buffer (pH 8.5) for 4 min at 0 °C. The reaction was stopped by adding 0.2 M potassium acetate (pH 5), and the tRNA was precipitated by ethanol.

Excess dye was removed by four additional ethanol precipitation steps<sup>33,34</sup>. The degree of labelling was >97% according to HPLC analysis on Biosuite 250 HR 30 cm SEC column (Waters).

**Translation.** Translation was carried out in buffer A (50 mM Tris–HCl, pH 7.5, 70 mM NH<sub>4</sub>Cl, 30 mM KCl and 7 mM MgCl<sub>2</sub>) at 37 °C, unless stated otherwise. Initiation complexes were formed by incubating ribosomes (0.5  $\mu\text{M}$ ), mRNA (1.5  $\mu\text{M}$ ), IF1, IF2 and IF3 (0.75  $\mu\text{M}$  each), GTP (1 mM) and either Bpy-<sup>[3</sup>H]Met-tRNA<sup>Met</sup> or initiator [<sup>3</sup>H]Met-tRNA<sup>Met</sup> (0.75  $\mu\text{M}$ ), in buffer A at 37 °C for 30 min. Ternary complex EF-Tu–GTP–aminoacyl-tRNA was prepared by incubating EF-Tu (58  $\mu\text{M}$ ) with GTP (1 mM), phosphoenol pyruvate (3 mM) and pyruvate kinase (0.1 mg ml<sup>-1</sup>) for 15 min at 37 °C, then adding purified total aa-tRNA (about 60  $\mu\text{M}$ ) and EF-G (1  $\mu\text{M}$ ) and incubating for 1 min at 37 °C. *In vitro* translation was started by mixing initiation ribosomal complexes (0.08  $\mu\text{M}$ ) with the ternary complexes (24  $\mu\text{M}$ ) total aa-tRNA (containing [<sup>14</sup>C]Leu-tRNA<sup>Leu</sup>) and incubated at 37 °C. Reactions were terminated by either shock-freezing aliquots in liquid nitrogen or by adding 2 × loading buffer for SDS–polyacrylamide gel electrophoresis. Samples were treated with RNase A for 30 min and the translation products were separated by Tris-Tricine gel electrophoresis<sup>35,36</sup>. Fluorescent and radioactive peptides were detected after gel electrophoresis using FLA-7000 scanner (FujiFilm) and quantified using the Multi-Gauge software. Unless stated otherwise, bypassing efficiency was calculated as the intensity of the N-terminal Bpy label in the byp divided by the sum of intensities from bypassing (byp) and truncated (stop) gp60 bands. Similarly, the efficiency of read-through (Fig. 1e) was calculated from the Bpy signal in the read-through band (rt) divided by the sum of intensities of byp, stop and rt bands. When [<sup>14</sup>C]Leu was used to visualize the byp and stop products (Fig. 1d), bypassing efficiency was calculated in the same way, taking into account that the byp product contains five times as many Leu residues as the stop product. To calculate bypassing efficiency from Lumio detection, the intensities of the byp and  $\Delta\text{gap}$  products were compared (Fig. 1c).

**Statistics.** The s.d. were analysed from a large collection of technical and biological replicates. The magnitude of standard deviation depended on the band efficiency: when the bypassing efficiency was low, <10%, the standard deviation was about 30% of the given value, for example, 10 ± 3%. For the higher bypassing efficiencies, standard deviation was lower, about 10%, for example, absolute bypassing efficiency for the WT construct at standard conditions was 22 ± 2%.

## References

- Lang, B. F. *et al.* Massive programmed translational jumping in mitochondria. *Proc. Natl Acad. Sci. USA* **111**, 5926–5931 (2014).
- Smith, M. C. *et al.* Evolutionary relationships among actinophages and a putative adaptation for growth in *Streptomyces* spp. *J. Bacteriol.* **195**, 4924–4935 (2013).
- Huang, W. M. *et al.* A persistent untranslated sequence within bacteriophage T4 DNA topoisomerase gene 60. *Science* **239**, 1005–1012 (1988).
- Weiss, R. B., Huang, W. M. & Dunn, D. M. A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell* **62**, 117–126 (1990).
- Huang, W. M., Wei, L. S. & Casjens, S. Relationship between bacteriophage T4 and T6 DNA topoisomerases. T6 39-protein subunit is equivalent to the combined T4 39- and 60-protein subunits. *J. Biol. Chem.* **260**, 8973–8977 (1985).
- Bonocora, R. P., Zeng, Q., Abel, E. V. & Shub, D. A. A homing endonuclease and the 50-nt ribosomal bypass sequence of phage T4 constitute a mobile DNA cassette. *Proc. Natl Acad. Sci. USA* **108**, 16351–16356 (2011).
- Herr, A. J., Wills, N. M., Nelson, C. C., Gesteland, R. F. & Atkins, J. F. Drop-off during ribosome hopping. *J. Mol. Biol.* **311**, 445–452 (2001).
- Herr, A. J., Gesteland, R. F. & Atkins, J. F. One protein from two open reading frames: mechanism of a 50 nt translational bypass. *EMBO J.* **19**, 2671–2680 (2000).
- Wills, N. M. *et al.* Translational bypassing without peptidyl-tRNA anticodon scanning of coding gap mRNA. *EMBO J.* **27**, 2533–2544 (2008).
- Todd, G. C. & Walter, N. G. Secondary structure of bacteriophage T4 gene 60 mRNA: implications for translational bypassing. *RNA* **19**, 685–700 (2013).
- 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. *J. Mol. Biol.* **309**, 1029–1048 (2001).
- Bucklin, D. J., Wills, N. M., Gesteland, R. F. & Atkins, J. F. P-site pairing subtleties revealed by the effects of different tRNAs on programmed translational bypassing where anticodon re-pairing to mRNA is separated from dissociation. *J. Mol. Biol.* **345**, 39–49 (2005).
- Wills, N. M. in *Recoding. Expansion of Decoding Rules Enriches Gene Expression* (eds. Atkins, J. F. & Gesteland, R. F.) 365–381 (Springer, 2010).
- Herbst, K. L., Nichols, L. M., Gesteland, R. F. & Weiss, R. B. A mutation in ribosomal protein L9 affects ribosomal hopping during translation of gene 60 from bacteriophage T4. *Proc. Natl Acad. Sci. USA* **91**, 12525–12529 (1994).

15. Herr, A. J., Atkins, J. F. & Gesteland, R. F. Coupling of open reading frames by translational bypassing. *Annu. Rev. Biochem.* **69**, 343–372 (2000).
16. Wohlgenuth, I., Pohl, C. & Rodnina, M. V. Optimization of speed and accuracy of decoding in translation. *EMBO J.* **29**, 3701–3709 (2010).
17. Herr, A. J., Wills, N. M., Nelson, C. C., Gesteland, R. F. & Atkins, J. F. Factors that influence selection of coding resumption sites in translational bypassing: minimal conventional peptidyl-tRNA:mRNA pairing can suffice. *J. Biol. Chem.* **279**, 11081–11087 (2004).
18. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
19. Lu, J. & Deutsch, C. Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **12**, 1123–1129 (2005).
20. Bornemann, T., Jockel, J., Rodnina, M. V. & Wintermeyer, W. Signal sequence-independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. *Nat. Struct. Mol. Biol.* **15**, 494–499 (2008).
21. Takyar, S., Hickerson, R. P. & Noller, H. F. mRNA helicase activity of the ribosome. *Cell* **120**, 49–58 (2005).
22. Maldonado, R. & Herr, A. J. Efficiency of T4 gene 60 translational bypassing. *J. Bacteriol.* **180**, 1822–1830 (1998).
23. Herr, A. J., Atkins, J. F. & Gesteland, R. F. Mutations which alter the elbow region of tRNA<sup>2Gly</sup> reduce T4 gene 60 translational bypassing efficiency. *EMBO J.* **18**, 2886–2896 (1999).
24. Pulk, A. & Cate, J. H. Control of ribosomal subunit rotation by elongation factor G. *Science* **340**, 1235970 (2013).
25. Tourigny, D. S., Fernández, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340**, 1235490 (2013).
26. Zhou, J., Lancaster, L., Donohue, J. P. & Noller, H. F. Crystal structures of EF-G-ribosome complexes trapped in intermediate states of translocation. *Science* **340**, 1236086 (2013).
27. Qin, P., Yu, D., Zuo, X. & Cornish, P. V. Structured mRNA induces the ribosome into a hyper-rotated state. *EMBO Rep.* **15**, 185–190 (2014).
28. Milon, P., Maracci, C., Filonava, L., Gualerzi, C. O. & Rodnina, M. V. Real-time assembly landscape of bacterial 30S translation initiation complex. *Nat. Struct. Mol. Biol.* **19**, 609–615 (2012).
29. Bock, L. V. *et al.* Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat. Struct. Mol. Biol.* **20**, 1390–1396 (2013).
30. Milon, P. *et al.* Transient kinetics, fluorescence, and FRET in studies of initiation of translation in bacteria. *Methods Enzymol.* **430**, 1–30 (2007).
31. Rodnina, M. V. & Wintermeyer, W. GTP consumption of elongation factor Tu during translation of heteropolymeric mRNAs. *Proc. Natl Acad. Sci. USA* **92**, 1945–1949 (1995).
32. Rodnina, M. V. *et al.* Thiostrepton inhibits the turnover but not the GTPase of elongation factor G on the ribosome. *Proc. Natl Acad. Sci. USA* **96**, 9586–9590 (1999).
33. Gite, S., Mamaev, S., Olejnik, J. & Rothschild, K. Ultrasensitive fluorescence-based detection of nascent proteins in gels. *Anal. Biochem.* **279**, 218–225 (2000).
34. Mittelstaet, J., Konevega, A. L. & Rodnina, M. V. A kinetic safety gate controlling the delivery of unnatural amino acids to the ribosome. *J. Am. Chem. Soc.* **135**, 17031–17038 (2013).
35. Doerfel, L. K. *et al.* EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* **339**, 85–88 (2013).
36. Schagger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166**, 368–379 (1987).

### Acknowledgements

We thank Thomas Bornemann and Wolf Holtkamp for providing ribosomes with mutations in L4 and L22 before publication; Wolfgang Wintermeyer for critical comments on the manuscript; Anna Bursy, Olaf Geintzer, Sandra Kappler, Theresia Uhlendorf, Tanja Wiles and Michael Zimmermann for expert technical assistance. The work was supported by a grant of the Deutsche Forschungsgemeinschaft (M.V.R.) and by RFBR, research project No. 13-04-40212-H comfi (to A.L.K.). J.F.A. was personally supported by Science Foundation Ireland.

### Author contributions

E.S. and A.L.K. prepared materials and conducted experiments. All authors conceived the research, designed experiments, analysed the data and wrote the paper.

### Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

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

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

**How to cite this article:** Samatova, E. *et al.* High-efficiency translational bypassing of non-coding nucleotides specified by mRNA structure and nascent peptide. *Nat. Commun.* **5**:4459 doi: 10.1038/ncomms5459 (2014).