Premature termination codons signal targeted gene repair by nonsense-mediated gene editing in *E. coli*

Xiaolong Wang1*, Xuxiang Wang1, Haibo Peng1, Chunyan Li1, Gang Chen1, Jianye Zhang1, Michael Lynch2

1. College of Life Sciences, Ocean University of China, Qingdao, 266003, P. R. China
2. Center for Mechanisms of Evolution, BioDesign Institute, Arizona State University, Tempe, AZ, 85287-7701

* To whom correspondence should be addressed: Xiaolong Wang, Department of Biotechnology, Ocean University of China, No. 5 Yushan Road, Qingdao, 266003, Shandong, P. R. China, Tel: 0086-139-6969-3150, E-mail: Xiaolong@ouc.edu.cn.
Abstract

Frameshift mutations yield truncated proteins, leading to loss-of-function, genetic disorders or even death. Reverse mutations, which restore the wild-type phenotype of a mutant, have been assumed to be far rarer than forward mutations. However, in this study, screening tests showed that the revertants of a frameshift mutation are detected more frequently than expected in *E. coli*. Sanger sequencing of the revertants revealed that the reversion of the frameshift mutation is not caused by random mutagenesis but by active targeted gene repair. Molecular studies suggest that premature termination codons (PTCs) in the nonsense mRNAs signal the repair of the frameshift mutation. Genome survey indicate that the genome sequence of a revertant is stable, confirming that the DNA replication proofreading/mismatch repair system of the revertant is not defective. Transcriptome profiling identified dozens of upregulated genes/pathways that possibly involve in frameshift repair, including DNA replication, RNA processing, RNA editing, mismatch repair and homologous recombination. Introducing synthetic DNA or RNA oligonucleotides into the frameshift can promote the gene repair. Based on these data and an in-depth review of previous studies, we hypothesized a molecular model for frameshift repair referred to as *nonsense-mediated gene editing* (NMGE): nonsense mRNAs are recognized by mRNA surveillance by PTC signaling, edited by RNA editing and then used to direct the repair of their defective coding gene through mismatch repair and homologous recombination. Moreover, NMGE may also serve as a driving force for molecular evolution and a new source of genetic diversity, leads to a widespread existence of frameshift homologs.
1. Introduction

DNA replication and DNA repair happen daily in cell division and multiplication. DNA damage, point mutations, insertions and deletions (indels) can be induced by chemical and physical mutagens, such as radiation, pollutants, or toxins. In addition, because of the imperfect nature of DNA replication, mutations and indels also occur spontaneously as replication errors or slipped-strand mispairing. In a protein-coding gene, if the size of an indel is not a multiple-of-three, it causes a frameshift mutation, which results in a substantial change of its encoded amino acid sequence downstream; in addition, it often produces a number of nonsense codons and yields truncated and dysfunctional proteins [1], which may lead to a genetic disorder or even death [2-5].

The molecular mechanisms/pathways involved in the repair of DNA damages and point mutations have been intensively studied [6], including mismatch repair, base- or nucleotide-excision repair, homologous recombination and non-homologous end joining. The maintenance of reading frames of protein-coding genes is fundamentally important for cells. A reverse mutation phenomenon was discovered as early as in the 1960s [7]. Point mutations have been explained by random mutagenesis, and it has been assumed that reverse mutations occur at a lower rate than forward mutations. But, it remains unclear whether a molecular mechanism exists for the repair of frameshift mutations (hereafter frameshift repair). However, here we report that the reversion of a frameshift mutation occurs far more frequently than expected in E. coli, and that they are not randomly produced but involve active targeted gene repair.

2. Materials and Methods

2.1 Frameshift mutagenesis and back mutation screening

A plasmid containing a frameshifted β-lactamase gene (bla), pBR322-(bla-), was constructed by site-directed mutagenesis (Fig 1): in plasmid pBR322, a G:C base pair, located in the upstream (+136) of the wild-type bla gene (bla+), was deleted by using overlapping extension polymerase chain reaction (OE-PCR). Competent cells of Escherichia coli strain DH5α were transformed with pBR322-(bla-), propagated in tetracycline broth (TCB), and then diluted cultures were plated on a tetracycline plate...
(TCP) to count the total number of bacteria \((N_0)\) and on ampicillin plates (ACPs) to screen for and count the number of revertants \((N_1)\). The recovery rates were calculated as \(N_1/N_0\). The screened revertants were picked and propagated in ampicillin broth (ACB) at 37°C with 200 rpm shaking and with 1 mL of overnight seed culture in 10 mL ACB. The growth rates of the revertants were evaluated by the doubling time during their exponential growth phase. Their plasmids DNA were extracted and their \(bla\) genes were Sanger sequenced.

### 2.2 Construction and expression of a PTC-free frameshifted \(bla\)

A PTC-free frameshifted \(bla\), denoted as \(bla\#\), was derived from \(bla\)- by replacing each nonsense codon with a sense codon according to the readthrough rules (Table 1), and by adding a stop codon (TAA) in the 3’-end. The \(bla\#\) gene was chemically synthesized by Sangon Biotech, Co. Ltd (Shanghai), inserted into an expression vector, pET28a, and transformed into competent cells of \(E. coli\) strain BL21. Plasmid pET28a contains a kanamycin-resistance gene, therefore the transformed bacteria were plated on a kanamycin plate (KCP), the transformants were propagated in kanamycin broth (KCB), plated on ACPs to screen for revertants, and then their \(bla\) genes were Sanger sequenced. The expression of the frameshifted BLA was induced by 0.1 mM IPTG. Their total protein samples were extracted, frameshifted BLA was purified by nickel column chromatography and analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The activity of BLA was measured by iodometry.

### 2.3 Introduction of DNA or RNA oligonucleotides into the frameshift mutant

Competent cells of the frameshift mutant, which contains the defective plasmid pBR322-(bla-), were transformed with a 39-nt synthetic DNA or RNA oligonucleotide, which has a wild-type or mutated sequence restoring the base pair which was deleted in OE-PCR. The transformed cells were cultured in TCB at 37°C for 6 hours with 200 rpm shaking, and then diluted cultures were plated on TCPs to count the total number of bacteria \((N_0)\), and on ACPs to screen for and count the number of revertants \((N_1)\). The recovery rates were calculated by \(N_1/N_0\). The revertants were subcultured in ACB to test their growth rates, and their \(bla\) were Sanger sequenced. The growth rates were
evaluated by the doubling time during the exponential growth phase by culturing at 37°C with 200 rpm shaking and with 1 mL of overnight seed culture in 10 mL ACB.

2.4 Genome resequencing and structure/variation analysis of the E. coli strains

Genomic DNA samples were extracted from wild-type (bla+) and revertant (bla*) strains. Library preparation and genome sequencing were conducted by a commercial service provided by Novogene Co. Ltd. The library was sequenced on an Illumina HiSeq 250PE platform and paired-end reads were obtained. Raw reads in fastq format were processed by removing the adapter sequence and low-quality reads. The clean reads with high quality were mapped onto the reference genome E. coli K12 MG1655 (NC_000913.3) to identify single nucleotide polymorphisms (SNPs), indels and structure variations (SVs) of each sample. Samtools v0.1.18 was used to sort the reads and reorder the bam alignments [8]. SNP typing was performed using the genome analysis tool kit (GATK) [9]. Circos was used to display the coverage of the reads and the distribution of SNPs and indels on the ring diagram [10]. BreakDancer was used to detect structure variations [11, 12].

2.5 Transcriptome analysis of the E. coli strains

Total RNA samples were extracted from the wild type (bla+), the frameshift (bla-) and the revertants (bla*) E. coli strains. Library preparation and RNA sequencing were conducted by a commercial service provided by Novogene Co. Ltd. After library preparation and cluster generation, the library was sequenced on an Illumina HiSeq platform. The paired-end raw reads were processed by removing the adapter sequence and low-quality reads. The high-quality clean reads were mapped onto the annotated reference genome, E. coli K12 MG1655 (NC_000913.3), and the expression levels of each gene were calculated for each sample and compared to each other to identify differentially expressed genes (DEGs), enriched Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

(1) Quantification of gene expression level

Bowtie (v2-2.2.3) was used for index building and aligning the clean reads to the reference genome [13]. HTSeq (v0.6.1) was used to count the numbers of clean reads mapped to each gene [14]. And then expected number of Fragments PerKilobase of
transcript sequence per Millions base pairs sequenced (FPKM) of each gene was calculated based on their lengths and read counts.

(2) Differential expression analysis

Prior to differential expression gene analysis, for each strain the read counts were adjusted by the edgeR program [15]. Differential expression of the read counts of two conditions was performed using the DEGSeq R package (1.20.0) [16]. The P values were adjusted by the Benjamini & Hochberg method. The threshold for significantly differential expression were set as: corrected P-value (qvalue) < 0.005 and log2 (Fold change) ≥ 1.

(3) GO and KEGG enrichment analysis of differentially expressed genes

GO enrichment analysis was implemented by GOseq R package [17], in which gene length bias was corrected. GO terms with corrected P-value less than 0.05 were considered significantly enriched by DEGs. KOBAS 2.0 [18] was used to test the statistical enrichment of DEGs in KEGG pathways (http://www.genome.jp/kegg/).

3. Results and Analysis

3.1 The growth of the frameshift mutant and the revertants

The plasmid pBR322 contains two wild-type resistance genes, a β-lactamase gene (bla+) and a tetracycline resistance gene (tcr+). Using OE-PCR, one single G:C base pair in the upstream of bla+ was deleted (Fig 1). In the resulting frameshift bla (bla-), seventeen nonsense codons appeared. Because the active sites of BLA, including the substrate binding site, all locate in the region downstream the deletion, this frameshift mutation is expected to be a loss-of-function. The plasmid containing a wild-type or a frameshifted bla gene was transformed into E. coli (strain DH5α and BL21), then the transformants were propagated in tetracycline broth (TCB), dilution series were plated on tetracycline plate (TCPs) to estimate the total number of viable bacteria and plated on ampicillin plates (ACPs) to test their ampicillin resistance. The wild-type (bla+) bacterial grew well on both ACP and TCP (Fig 2A, left); and it was expected that the deficiency-type (bla-) bacterial would not grow on ACP. However, repeatedly a few ampicillin-resistant colonies (bla*) were observed (Fig 2A, middle). The possibility of
Premature termination codons signal frameshift repair

carryover- or cross-contamination was ruled out by blank controls (Fig 2A, right). So, these ampicillin-resistant colonies were considered as revertants whose \textit{bla} genes were repaired by reverse mutations.

In addition, although most of the revertants can grow in ACB, the growth rates of the revertants are often very much slower when compared to the wild-type strain: by culturing at 37°C with 200 rpm shaking, the wild type took 12 to 24 hrs. to reach the late log phase, while the revertants took up to 36 to 48 hrs., and many of the revertants failed to grow when they were subcultured in ACB, even without adding ampicillin.

Hitherto, it seems that there is nothing unusual, because the reverse mutations are fairly common phenomenon, arising from simple random mutations (Fig 3): in DNA replication, natural or induced mutations occur randomly, a forward mutation causes a defective mutant or a loss-of-function, and a reverse mutation restores the wild-type phenotype. Under this view, these revertants survived only because they were ‘lucky’: their defective \textit{bla} gene restored coincidently by a reverse mutation, while most of the other bacteria died without a reverse mutation.

### 3.2 frameshift repair is not a random mutation but a targeted gene repair

The classical random mutagenesis model for reverse mutation sounds faultless in explaining the reversions of point mutations. However, in a ‘thought experiment’, we noticed that the model is insufficient to explain the frameshift repair observed herein: (1) in the model, a bacterium survives if its frameshifted \textit{bla} gene is restored through random mutagenesis. However, a reverse mutation must occur in the process of DNA replication or repair in a living cell. If the mutant itself could not live in ampicillin, the reverse mutations could not occur after they were cultured in ampicillin but must have occurred before they were cultured in ampicillin; (2) a frameshift mutation can only be repaired by an appropriate insertion or deletion restoring the reading frame of the frameshifted gene. However, even if relatively small, a bacterial genome consists of millions of base pairs over thousands of genes. If the cells did not ‘know’ which gene is frameshifted, it will be almost impossible to repair it through random insertion or deletion; (3) In \textit{E. coli}, the rate of point mutation is very low. As determined by whole-genome sequencing [19], the baseline mutation rate of the wild-type \textit{E. coli} is
~10^{-3} per genome (or ~10^{-9} per nucleotide) per generation. The frameshift mutation rate measured are even about ten-times lower than that of point mutations [19]. It is generally assumed that reverse mutations occur at a lower rate than forward mutations. For a certain frameshift mutation, the reverse mutation rate should be extremely low if it depends on random mutagenesis, because insertions/deletions are far rarer than substitutions. Therefore, the revertants of a frameshift should be very difficult to be detected. However, through repeated screening tests, we found that the revertants of \textit{bla}- can always be detected easily, and it seems that the reversions of the frameshift mutation occur much more frequently than expected. The measured average recovery rate of \textit{bla}- was 2.61\times10^{-8} per cell division in DH5\alpha and reached up to 1.19\times10^{-7} per cell division in BL21, both significantly higher than the baseline.

Therefore, it is possible that an efficient mechanism specific for frameshift repair has been developed in the history of life. However, how a frameshifted gene is being identified and repaired at the molecular level is unknown.

3.3 \textit{Sanger sequencing of the \textit{bla} genes of the revertants}

To investigate how the frameshift gene is repaired, the revertants were picked and subcultured in ACB. The initial revertants usually grew very slowly, and many of the subcultures failed to grow, but the surviving subcultures grow faster and faster. Their plasmid DNA was extracted, and their \textit{bla} were Sanger sequenced. Unexpectedly, in these initial revertants and subcultures, the \textit{bla} genes are often neither a wild type nor a revertant, but still frameshifted (Fig 2B, top). It is puzzling how these frameshifted \textit{bla} could be functional. To clarify this, the subcultures were re-subcultured and their \textit{bla} were re-sequenced for over 300 replications.

In the later subcultures, the sequencing results of the \textit{bla} genes are also often still frameshifted, but their Sanger diagrams often contain two sets of superimposed peaks (Fig 2B, middle), indicating that they consist of two types of \textit{bla} sequences: one is the original frameshifted \textit{bla}, and the other is the repaired one. By visual inspection of the Sanger diagrams, the repaired sequences of the revertants (\textit{bla}*) were read from the secondary peaks. Some sequencing diagrams even contain three or more sets of peaks, but the sequences are nonreadable. In the final fast-growing subcultures, most of their
bla copies were repaired, and their Sanger diagrams contain few superimposed peaks (Fig 2B, bottom).

As shown in Fig 2C, by aligning bla* sequences against bla+ and bla- sequences, it was found that they contain a variety of indels and substitutions, especially G:C←→A:T transitions (Fig 2C, circles). Previously, we developed a novel tool for multiple sequence alignment, codon and aa unified alignment (CAUSA) [20], which can align and visualize coding sequences and their encoded protein sequences in the same user interface. As shown in Fig 2D, in CAUSA, it is clearly demonstrated that these bla* coding genes often contain some stop codons, and the translated protein sequences are different from, but highly similar to, that of the wild-type BLA if the stop signals were ignored. In other words, the frameshifted genes are restored far from completely or immediately, but partially and progressively, resulting in a variety of frameshift variants, called frameshift homologs.

3.4 The genome of the revertant is not highly variable but very stable

As is well known, an extraordinary high rate of spontaneous mutagenesis occurs over its whole genome in a mutator, such as E. coli strain mutD5, due to its defective proofreading and mismatch repair during DNA replication [21]. The E. coli strain we used in this study, either DH5α or BL21, is not a mutator, but the frameshift/revertants obtained may change into mutators if their proofreading and mismatch repair system is damaged during culturing, so that high-rate of mutagenesis can occur in their genomes, not only bla gene but also other genes.

To clarify whether the other genes had also changed, the tetracycline-resistant (tcr) gene in the same plasmid of the frameshift/revertants was also sequenced by the Sanger method, but almost no overlapping peaks were observed in their sequencing diagrams, suggesting that the high-level variations observed in bla* are not a result of random mutation, but more probably an active gene repair targeting specifically to the frameshifted bla gene, but not to other normal genes.

To clarify further whether the frameshift/revertants had changed into mutators, the genome sequences of a frameshift (bla-) and a revertant (bla*) were sequenced by
high-throughput NGS sequencing. The obtained clean reads were mapped onto and
compared with the reference genome sequence of the *E. coli* strain K12 (Fig 4A). As
shown in Fig 4B and Table 2, the SNP/indel levels of the revertant and frameshift
tested are both very low, suggesting that their proofreading/mismatch repair system is
not defective. Therefore, they are not mutators, but adopt a stringent proofreading and
mismatch repair strategy during DNA replication. In addition, if the variations of *bla*
are the result of whole-genome random mutagenesis, the number of variations in *bla*
should be proportional to the number of genomic variations, and therefore the genome
sequences of the revertants should contain more variations when compared to that of
the frameshift. But in fact, as shown in Fig 4B and Table 2, the SNP/indel levels of
the revertant are not higher but even lower than those of the frameshift. The number
of variations of *bla* are independent on that of genome variations, suggesting that the
high-level of variations observed in their *bla* are not caused by whole-genome random
mutagenesis, but more probably through an active mechanism specifically targeting
the frameshifted *bla* gene.

In addition, genome structure analysis shows that there are quite some structural
variations (SVs) in each of the two genomes tested. However, as shown in Fig 4C and
Table 3, most of the SV sizes are 100–200 bp, approximate to the length of the reads.
The lengths of the SVs should follow a normal or any other skewed distribution, these
equal-length SVs are therefore considered as mostly falsely-mapped reads rather than
ture SVs. As is known, NGS sequencing technologies have several major limitations,
especially short read lengths and amplification bias [22], posing serious problems in
sequence assembly, genome mapping and structures analysis. However, thanks to the
extremely high sequencing depth, the SNP/indel data of NGS is considered as valid
and reliable.

### 3.5 PTCs signal the repair of the frameshift mutation

We speculated that it was the PTCs presented in the nonsense mRNAs that signal
the repair of frameshift mutations. To validate this, a *PTC-free bla* (denoted as *bla#*)
was derived from *bla* by replacing each nonsense codon with an appropriate sense
codon in accordance with the *in vivo readthrough rules* (Table 1). The *bla#* gene was
synthesized chemically, cloned in the plasmid vector pET-28a and expressed in *E. coli* strain BL21. The *bla*# transformants were plated on ACPs, but no revertant was found. The *bla*# gene of the transformants were Sanger sequenced, while few variations were observed, suggesting that the *bla*# gene is not repaired as efficiently as the *bla*- gene. Analysis of total and purified protein by SDS-PAGE revealed a band corresponding to the predicted molecular mass (34-kDa), representing the product expressed from the PTC-free frameshifted *bla* (Fig 5A). No lactamase activity was detected by iodimetry test of the product (the frameshifted BLA) (Fig 5B), suggesting that it is indeed a loss-of-function.

In summary, the *bla*# gene was kept unchanged and a dysfunctional product was expressed in its wrong reading frame in *E. coli*. The only difference between *bla*# and *bla*- is that *bla*- consists of a certain number of PTCs but *bla*# does not. Therefore, it must be the PTCs that signaled the frameshift repair in the frameshifts. We postulated that the transcripts of the frameshifted gene (the nonsense mRNAs) must be involved, because presumably these PTCs can only be recognized in a nonsense mRNA in the process of translation but not likely possible in a DNA double helix.

### 3.6 Identifying genes/proteins involved in frameshift repair

We speculated that frameshift repair requires the upregulation of relevant genes, proteins and pathways. By transcriptome sequencing and profiling, as shown in Fig 6A, the gene expression patterns of these strains are significantly different from each other. When compared to the wild-type (*W*), respectively 292 or 247 genes displayed upregulated transcription (Fig 6A, Supplementary dataset 1-2) in the frameshift (*FS*) or revertant (*RE*). In the pathway enrichment analysis, respectively 65 or 58 pathways were identified as upregulated (Supplementary dataset 3-4). However, the upregulated genes/pathways are mostly nutrition-related, which do not likely involve in frameshift repair but more probably due to the different growth conditions of these different cells. Therefore, we selected the upregulated genes/pathways that most probably involve in frameshift repair manually. The criterion of selection is whether the gene/pathway is involved in the processing of the storage of genetic information (DNA or RNA). As shown in Fig 6B-6C and Table 4-5, the selected upregulated genes/pathways include
DNA replication (Fig 6D), RNA degradation/processing (Fig 6E), mismatch repair and homologous recombination (Fig 6F).

In addition, the selenocysteinyl-tRNA-specific translation factor (selB) gene was upregulated in one of the two revertants tested (Supplementary dataset 5). SelB is a specialized translation factor that takes the place of elongation factor EF-Tu, in the case of selenocysteines being inserted into a peptide chain at the site of a UGA codon [23, 24]. The upregulation of selB suggested that the nonsense codons remained in the mRNAs might be readthrough in the revertants.

By carefully inspecting the list of DEGs (Supplementary dataset 6), we noticed that the expression levels of the cytosine or isoguanine deaminase gene (codA) and the adenosine deaminase gene (add) were upregulated in the frameshift, and those of the cytidine/deoxycytidine deaminase gene (cdd) and the tRNA-specific adenosine deaminase gene (tadA) were also upregulated in the revertant:

(1) In *E. coli*, cytosine or isoguanine deaminase (CodA) catalyzes the hydrolytic deamination of cytosine to uracil, and also the deamination of isoguanine, a mutagenic oxidation product of adenine in DNA, and of isocytosine [25, 26]. In eukaryotes, cytosine deaminase is involved in RNA editing [27].

(2) In *E. coli*, cytidine/deoxycytidine deaminase (Cdd) scavenges exogenous and endogenous cytidine and 2'-deoxycytidine for uridine monophosphate (UMP) synthesis [28-33]. In eukaryotes, cytidine/deoxycytidine deaminase is also involved in RNA editing [34, 35].

(3) In *E. coli*, adenosine deaminase (Add) is an inducible enzyme that converts adenine, adenosine, and deoxyadenosine to guanine nucleotides [36-38]. In eukaryotes, adenosine deaminase (ADAR2) is an important enzyme which is responsible for RNA editing [39-51].

(4) The tRNA-specific adenosine deaminase (TadA) catalyzes the deamination of adenosine to inosine at the wobble position 34 of tRNA-Arg2. TadA was first identified in yeast [52] and the first prokaryotic RNA editing enzyme to be identified in *E. coli* [53] and found widely exist in bacterial [53-56]. TadA
is related to the eukaryotic RNA editing enzymes, ADAR1 and ADAR2 [52], which widely exist in human [57] mouse [58] and fruit fly [59].

As above mentioned, Sanger sequencing showed that the base substitutions in the repaired bla are mostly G:C→A:T transitions, suggesting that they are derived from deamination of DNA bases. As is known, DNA deamination generates deoxyinosine from deoxyadenosine, which links DNA mismatch repair and RNA editing [60]. Here we show that DNA deamination and mismatch repair pathways are upregulated in the frameshift, suggesting that RNA-editing and mismatch-repair are probably both involved in frameshift repair. These genes are also upregulated in the revertant, suggesting that gene repair is still ongoing in the revertants. As mentioned above, the genome sequence of the revertant is not highly variable but stable, suggesting that these upregulated genes are tightly controlled, their function is restricted specifically to the target gene, and thus do not cause many off-targeted variations in their genomes. However, further experimental investigation is needed to validate the regulation and the function of many of these genes in the repair of frameshift mutations.

### 3.7 Both DNA and RNA can direct the repair of frameshift mutation

Synthetic DNA oligonucleotides have been widely used for targeted gene repair [61-63]. Here, competent cells of the frameshift mutants (E. coli strain DH5α or BL21) were transformed by heat-shocking with a 39-nt synthetic sense (ss-), antisense (as-) or double-stranded (ds-) DNA, having a wild-type (WT) or mutated (MT) sequence to recover the base which was deleted in the OE-PCR. The recovery rate (fr) was very low and unstable in DH5α, but much higher and more stable in BL21. We performed statistical analysis on the data of BL21. As shown in Fig 7A-7B, in the blank or heat-shock-only control groups, the average fr is 1.19×10⁻⁷ and 1.30×10⁻⁷, respectively, suggesting that heat-shock does not bring significant difference to the recovery rate by itself. Comparing with the blank control group, on average fr increased to 1.93-fold in the ss-DNA group, 6.72-fold in the as-DNA group and 2.94-fold in the ds-DNA group, respectively, of which differences are all significant (F-test P<0.01). These data are
consistent with other’s previous studies that both sense- and antisense-DNA can
induce targeted gene repair [64].

It has also been reported that synthetic RNA oligos, DNA/RNA complexes and
endogenous transcripts can direct the repair of DSBs of DNA [65-70]. We postulated
that synthetic RNAs can also promote frameshift repair if the NMGE model is correct.

To validate this, the frameshift mutants (strain DH5α or BL21) were transformed with
a 39-nt WT or MT RNA to recover the base deleted. The MT RNA caused an increase
of \(f_r\) to 3.58-fold, however, it is surprising that the WT RNA caused a decrease of \(f_r\) to
0.51-fold (Fig 7B). The differences of the recovery rates among different groups are
all significant by F-test, unfortunately, however, are not significant by t-test, because
their standard deviations are very large (Fig 7B).

For each group, the bla genes were sequenced to test whether they were repaired
or not. The Sanger diagram showed that a designed base was inserted into the repaired
bla gene in both the ssDNA (Fig 7C) and the ssRNA group (Fig 7D). In the control
group, the type and location of the inserted base are both unpredictable; in the ssDNA
or ssRNA group, however, the type of the insertional base is often consistent with the
designed type, and the location of the inserted base is exactly at the site of deletion,
suggesting that, like DNA, RNA can also direct targeted gene repair by itself.

Homologous recombination must play an important role in frameshift repair, as
the average recovery rate measured in BL21 (recA+) is significantly higher than that
of DH5α (recA-). However, homologous recombination may not be the only pathway
responsible for frameshift repair. In all DNA, RNA and control groups, G:C→A:T
substitutions were often observed in the repaired bla (Fig 7C). In the ssRNA group,
G:C→A:T substitutions were also often observed. In addition, MT-ssRNA-136-A, a
ssRNA designed with an G→A substitution, not only resulted in an increased average
recovery rate in BL21 (Fig 7B), but also efficiently introduced the substitution into
the repaired bla (Fig 7D), suggesting that mismatch repair and/or RNA editing might
be involved in the repair.


4. Discussion

4.1 A new model for the repair of frameshift mutations

The mechanisms for the repair of DNA damages and point mutations have been well-understood. However, the molecular mechanism for the recognition and repair of frameshift mutations is unknown at present. To gain further insight, we conducted an in-depth literature survey of previous studies that covered a wide range of prokaryotic and eukaryotic organisms. We put many of these fragmentary evidences together by presuming that the molecular mechanism of frameshift repair exists and is conserved among prokaryotic and eukaryotic species. Based on our studies and many others, we hypothesize a molecular model for frameshift repair, referred to as nonsense-mediated gene editing (NMGE), consisting of six main steps as follows (Fig 8):

Step 1. Nonsense mRNA recognition: in principle, a frameshift coding gene cannot be recognized at DNA level, and therefore must be first transcribed before it can be recognized. When a frameshift gene is transcribed, a number of nonsense mRNAs are produced. The frameshifted transcripts are recognized as nonsense mRNAs by RNA surveillance [71-77]. The stop codons emerge inside a frameshifted gene or mRNAs are called premature termination codons (PTCs) [78-80], also known as hidden stop codons (HSCs) [1]. When an nonsense mRNA is translated, the ribosomes are blocked when they encounter the PTCs [81]. Nonsense mRNAs are identified through mRNA surveillance, and usually causes mRNA decay [82-85].

Step 2. Nonsense mRNA processing: In eukaryotes, nonsense-mediated mRNA decay (NMD) is the best-studied nonsense mRNA processing pathway [86-88]. NMD targeting mRNAs harboring PTCs for degradation enhances their decay rate and reduces the abundance of the mRNA transcribed from that gene. In bacteria, it is not yet clear how nonsense mRNAs are processed, but it is known that the quality of their mRNAs is also tightly controlled [89]. Therefore, we speculate that NMD may also exist in bacteria, although it has not been characterized. Nevertheless, in both prokaryotes and eukaryotes, nonsense
mRNAs must be processed by one of the pathways, either degradation (step 2), translation (step 3), or editing (step 4).

**Step 3. Nonsense mRNA translation:** Nonsense mRNAs are not always fully degraded but sometimes translated. Frameshifted or nonsense-mutated coding genes can be expressed through some special mechanisms, including programmed translational frameshifting [90-97] and translational readthrough [98-101]. Especially, if the target gene is essential, a cell survives only if it produces functional proteins by translating a nonsense transcript/mRNA and dies if no protein is produced or the products are truncated or nonfunctional.

**Step 4. RNA editing:** nonsense mRNAs must be involved in the recognition and repair of the frameshift mutation, but a nonsense mRNA is defective by itself, and thus, cannot template the repair of its coding gene directly, therefore, it is more likely that the nonsense mRNAs are modified by RNA editing prior to templating the repair. Nonsense mRNAs can be recognized by PTC signaling, but it might be impossible for cells to locate the frameshift-causing indel exactly, either in the genomic DNA or transcripts, and thus, not always possible to restore the original coding sequence. In addition, a cell cannot distinguish whether a PTC is caused by an indel or by a point mutation. Therefore, the cells can only insert, delete and substitute nucleotides in the nonsense mRNAs within or nearby a PTC, resulting in a variety of variants. After editing, the repaired mRNAs are translated again, and the remaining PTCs are translated each into an appropriate amino acid by translational readthrough.

**Step 5. RNA-directed DNA repair:** to repair the frameshifted coding gene, mRNAs are transported back to chromosomal DNA to localize its defective coding sequence. the defective coding gene is repaired by using an edited mRNA as the template. A bacterium recovers if functional mRNAs and proteins are produced by a normal transcription and translation process.

As abovementioned, substantial evidence suggests a direct link between RNA editing and DNA repair, which can also support this NMGE model. However, further
systematic experimental investigations are needed to elucidate this process in detail and validate it in other species, especially the eukaryotes.

4.2 The NMGE model better explains the variations of the bla genes

As shown by a question mark in Fig 8, there exist two possible strategies: (1) direct NMGE: using an original nonsense mRNA to identify the defective coding sequence, and then editing it by DNA-level mutagenesis; however, whether or not the edited CDS is repaired, it must be transcribed and translated again to be functional; (2) indirect NMGE: the nonsense mRNAs are first edited, and then transported back to DNA to identify the CDS and to template the gene repair, i.e., a nonsense mRNA is functionalized through RNA editing prior to directing the repair of its own CDS. Compared to the direct strategy, indirect NMGE may be more efficient, because DNA repair only happens in a living cell gaining a functional mRNA by RNA editing.

As described above, distinct types of bases were substituted, inserted or deleted in the bla genes of the revertants. These editing events are mostly located near to or in the downstream of the frameshift-causing indel, but not in the distant upstream (Fig 2B-2C). This phenomenon can only be explained by the PTC-signaled RNA editing in the nonsense mRNAs: the PTCs presented downstream the indel may not only signal the recognition of nonsense mRNA but serve as flags for RNA editing. Because the PTCs appear in nonsense mRNAs only downstream of the indel, PTC-signaled RNA editing are restricted near to or in the downstream of the indel, potentially also restrict DNA editing to the corresponding region in the coding gene. Therefore, the frameshift mutation might be repaired by editing the nonsense mRNAs rather than the CDS directly. It is also possible that sometimes editing may cause even more problems in a nonsense mRNA, while these non-repaired mRNAs can be recycled or degraded. This process is repeated in cells until the PTCs in their bla genes are completely eliminated. When the cells are cultured in ampicillin media, the fate of a cell depends on whether it can produce a sufficient amount of functional proteins. By transcribing, each copy of a target gene can produce many copies of nonsense mRNAs, each of which might be edited in a different way, without changing the coding DNA sequences. Therefore,
the probability of producing a functional mRNA by editing nonsense mRNAs can be much higher than that of random mutagenesis in their coding sequences.

4.3 The NMGE model better explains the growth of the revertants

As abovementioned, frameshift repair is not genially a sudden + complete event but a progressive process. Interestingly, the sizes of wild-type colonies are uniform (Fig 2A), showing that they are uniform in growth rates; while the sizes of revertant colonies varied greatly (Fig 7A), indicating varied growth rates associate with the diversified activities of their BLA. Compared with random mutagenesis, NMGE not only better explains the high-level variations of the bla genes, but also the diversified growth rates of the revertants:

(1). In the initial revertants, whose bla genes were still frameshifted, survival can be explained in three possible ways: (a) the frameshifted bla is translated by *ribosomal frameshifting*, producing a wild-type BLA; (b) the frameshifted bla is translated by *translational readthrough*, producing a functional, partially frameshifted BLA; (c) the initial slowly-growing revertants may also contain a few copies of repaired bla, with the proportion of the repaired bla being too low to be detected by Sanger sequencing. As pBR322 is a multi-copy plasmid, it might be sufficient to support the survival of the host in ampicillin even when only a small proportion of its copies are repaired. The host bacterial grow slowly, and many of them failed to grow, probably due to these abnormal processes are very time- and resource-consuming, as cell energy and resources would be depleted for gene repair if the defective bla is not repaired soon.

(2). By NMGE, the bla genes were repaired progressively, not only before but also after culturing in the ampicillin media. As the substrate of BLA, ampicillin is neither the cause nor the terminator of the gene repair, but a selection pressure to screen and display the revertants, which is the result of gene repair.

(3). NMGE is triggered by the PTCs presented in the bla mRNAs, and the aim of NMGE is to eliminate these PTCs. In the initial revertants, there are so many PTCs in the bla mRNAs that the probability of producing a functional product
by mRNA editing is very low, so their growth rates and survival rates are very low, as most revertants died before their bla were sufficiently repaired.

(4). Later, in revertants whose bla are being repaired, with less and less remaining PTCs, energy and resources are used less and less for gene repair, more and more for growth, their growth rates and survival rates recover gradually.

(5). Finally, gene repair complete if all nonsense codons are eliminated. The bla coding genes are recovered, not necessarily the same as the original wild-type, but consisting of a variety of variants, whose growth rates may become normal, slower, or sometimes even faster.

Previously, we have reported that the genetic code, the protein-coding genes and genomes of all species are nearly optimal for frameshift tolerance [102]. Widespread premature termination codons were recently observed in human [78] and fruit fly [103] populations, suggest the ongoing evolutionary plasticity at the genetic level in modern human and other species. In this study, we further demonstrate that PTCs signal the frameshift repair. Frameshift mutations in protein-coding genes are repaired through NMGE, resulting in a variety of frameshift homologs, many of which are functional and preserved in nature. Therefore, the widespread existence of frameshift homologs among species may be considered as natural evidence for this model. In short, NMGE is not only a molecular mechanism for frameshift repair, but also be a driving force for molecular evolution and a source of genetic diversity.

4.4 Previous studies are supportive to this NMGE model

In the above, we proposed that NMGE depends on RNA-directed DNA repair but is triggered by PTCs, and the nonsense mRNAs are edited prior to directing the repair of the coding DNA. When a frameshift mutation occurs, the defective gene is repaired not by direct gene editing but by NMGE. Although it may seem surprising, but in fact it is more reasonable than direct gene editing. The above proposed NMGE model well explains our observations and is supported by numerous previous studies. In fact, the model is an integration of several key links, which have already been quite clear:

(1). mRNA decay is linked to other nonsense mRNA processing pathways
Nonsense-mediated mRNA decay (NMD) is the best-studied nonsense mRNA processing pathway [86-88]. In eukaryotes, NMD uses the presence of an exon junction complex (EJC) downstream of PTCs as a second signal to distinguish a PTC from a true stop codon [104]. In addition to NMD, the nonsense mRNAs may also be subject to other nonsense-mediated pathways, such as translational repression [105], transcriptional silencing [106], and alternative splicing [107], all of which are connected to NMD [105-107]. In eukaryotes, the NMD regulating factors, including three interacting proteins, UPF1, UPF2 (also known as NMD2) and UPF3, are encoded by highly conserved genes originally identified in yeast [108].

Frameshifted or nonsense-mutated coding genes can be expressed through some special mechanisms, including ribosomal frameshifting [90-97] and translational readthrough [98-101]. Ribosomal frameshifting occurs in all known organisms, from E. coli to mammals [109-112], whereby the ribosome is guided toward a triplet codon that is shifted one base position to the upstream (+1) or downstream (−1). In yeast, telomere maintenance is globally controlled by the ribosomal frameshifting process and NMD [97]. Translational readthrough is a process wherein a nonsense mRNA is translated by translating PTCs each into a specific amino acid [113-117]. Mutations in upf genes caused nonsense suppression and subsequent readthrough in yeast and 5-20 fold decreases in the rates of NMD [118], suggested a direct link between NMD, nonsense suppression and readthrough. The three Upf proteins all interact with Sup35, and Upf1 interact further with a second release factor, Sup45 (eRF1), and binding of a SMG-1-Upf1-eRF1-eRF3 (SURF) complex to EJC triggers Upf1 phosphorylation and NMD [119]. In addition, the nonsense suppression caused by upf mutations is additive to those by sup35 and sup45 mutations [120], suggesting that nonsense suppression is a nonsense-mRNA processing pathway alternative to NMD. In summary, although nonsense mRNAs usually lead to translation errors and NMD, when NMD is inhibited, they can also be translated by translational frameshifting, nonsense suppression or translational readthrough.

(2). RNA can direct DNA repair
We here demonstrated that many G:C→A:T substitutions were observed in the repaired bla in the revertants; in addition, four deaminase genes (codA, cdd, add and tadA) and the mismatch repair genes/pathways were upregulated in the frameshift and/or the revertant. We therefore speculate that mismatch repair and RNA editing are both involved in frameshift repair. The details of the mechanism of nonsense mRNA editing and how it connects to DNA repair has not been made clear.

All types of RNAs are subject to translation, degradation or processing, and a variety of molecular/cellular mechanisms are involved in RNA processing [121]. In the last few decades, studies have established a link between DNA repair and RNA processing. For example, several proteins respond to DNA damage and base-excision repair enzymes, such as SMUG1, APE1 and PARP1, have been shown to participate in RNA surveillance and processing [122].

As is known, RNA is transcribed in the nucleus and then transported to the cytoplasm in eukaryotes [123]. Recently, there is growing evidence that RNAs can also be transported back to the nucleus to repair DNA double strand breaks (DSBs), a process known as RNA-directed DNA Repair [65-70]. In yeast [69, 70] and human embryonic kidney (HEK-293) cells [124], not only a synthetic DNA/RNA hybrid, but also a RNA-only oligonucleotide can serve as a template for DNA synthesis for the repair of a DSB in a homologous DNA. In E. coli, short RNA patches can also direct DNA modification [124, 125]. However, this is the first report that RNA also can direct precise and targeted repair of a frameshift mutation by itself. Although this process is not yet clear in detail at the molecular level, mismatch repair and homologous recombination are likely to be involved, as they are both upregulated in the frameshift and the revertant tested (Fig 6F).

(3) RNA editing is linked to DNA repair:

RNA editing has been studied extensively and intensively [126-130], suggesting that RNA editing exists in many species, including E. coli [34], kinetoplastids [131], trypanosomes [132], plants [133, 134], mammals [135] and human [136-143]. Although the are several different mechanisms of RNA editing, such as A-to-I editing...
Premature termination codons signal frameshift repair

[144-147], C-to-U editing [148-152], and the insertion/deletion of uridine [153-155], the basic schematism are conserved among species. In particular, the nonsense apoB mRNA editing complex suppresses its own decay by NMD [156], suggesting that nonsense mRNAs can be subjected to RNA editing, which is linked to NMD.

Both DNA mismatch repair and RNA editing involve deamination, the removal of an amine group from a DNA/RNA molecule. Enzymes that catalyze this reaction are called DNA/RNA deaminases. Deamination of DNA deoxyadenosine bases generates deoxyinosine, create missense mutations predisposing humans to cancer, and interfere with other molecular genetic processes. Cytidine deamination was discovered first in wheat as a mechanism of mRNA editing [157, 158]. The characterization of the tRNA-specific adenosine deaminases (ADATs) has led to the suggestion that these enzymes, as well as the cytidine/adenosine deaminases (CDARs/ADARs) acting on pre-mRNAs, belong to a superfamily of RNA dependent deaminases [159]. The first example of C→U editing is the apolipoprotein B (apoB) mRNA, in which a single, site-specific cytidine deamination introduces a stop codon (UAA) into the reading frame of the edited transcript [160]. It is well known that deamination based RNA editing plays many important roles, and aberrant or dysregulated RNA editing might contribute to genomic instability in cancer [161]. In eukaryotes, several studies have provided evidence that deamination is related to both DNA repair and RNA editing. For example, endonuclease V, which is highly conserved from E. coli to human [60], is involved in excision repair that removes deoxyinosine from DNA.

TadA, the first prokaryotic deaminase for RNA editing identified in E. coli, displays sequence similarity to the yeast tRNA deaminase subunit Tad2p [53]. In addition, cytidine deaminase (CodA) has been used as a molecular model for RNA editing and a mechanism for RNA substrate recognition [34]. These studies suggest that RNA editing is conserved between prokaryotes and eukaryotes, and a direct link between the regulation of mismatch repair and RNA editing, and therefore support our speculation that mismatch repair and RNA editing may both involve in frameshift
repair. However, the details of the mechanism of nonsense mRNA editing and how it connects to DNA repair has not been made clear.

4.5 The possibility of RNA-only gene/genome editing

CRISPR/Cas9 for RNA-guided genome editing [162-164] has been widely used for the introduction or correction of specific mutations, which can be used to repair disease-causing frameshift/point mutations [165, 166]. However, present RNA-guided genome editing requires the transferring of foreign genes or proteins in the host cells. Deoxyoligonucleotide-based targeted gene repair [61, 62] could be more favorable and convenient due to its non-transgenic properties, but suffers from inefficiency because it relies on the induction of the endogenous DNA repair system while in the meantime being suppressed by mismatch repair [63].

In addition, synthetic DNA oligonucleotides have been widely used for targeted gene repair [61-63]. And it has also been reported that synthetic RNA or DNA/RNA complexes can direct the repair of double-strand breaks of DNA [65-70], but it has never been reported that endogenous mRNA can be edited and used to direct targeted gene repair. By transforming short synthetic RNAs, here we demonstrated that RNA can direct targeted gene repair by itself alone. It seems that RNA-directed gene repair is very inefficient and unstable, probably owing to the degradation of the naked RNAs by intra- or extra-cellular RNases. If the short RNA strands were protected by certain artificial packaging materials, it would probably result in increased efficiencies for RNA gene editing. In eukaryotes, liposome, polyethylenimine, and nano materials has been widely used for carrier agents to improve the transportation, distribution and retention of therapeutic nucleic acids within host cells, but methods for transferring of DNA/RNA into bacterial are restricted to heat shock or electroporation. Nevertheless, the components of NMGE, including RNA editing, mismatch repair and homologous recombination, are essential pathways highly conserved among various species from bacteria to human. Therefore, if NMGE is confirmed to be conserved, it is potentially possible that genes and genomes can be edited by using RNA-only molecules without introducing any exogenous gene or protein.
Author Contributions

Xiaolong Wang conceived the study, designed the experiments, analyzed the data and wrote the paper. Xuxiang Wang, Chunyan Li, Haibo Peng and Yalei Wang performed the experiments. Gang Chen and Jianye Zhang provided some equipment and materials, Michael Lynch supported this study, gave conceptual suggestions, and revised the manuscript.

Acknowledgements:

This study was supported by The National Natural Science Foundation of China (Grant No. 31571369).

Figure legends

Fig 1. The introduction of a frameshift mutation in the upstream of the bla gene in the plasmid pBR322. Sanger sequencing diagram of: (A) the wild type (bla+); (B) the frameshift mutant (bla-). (C) Alignment of the nucleotide sequence of bla+ and bla-.

Fig 2. Growth of the wild-type/frameshift E. coli and sanger sequencing of the bla genes:
(A) Growth of the wild-type and frameshift E. coli on ampicillin- or tetracycline plates: bla+: wild type; bla-: frameshift; bla*: revertants; blank: blank control
(B) The Sanger diagram of the bla genes of the revertants: Top: in the initial revertants, most of the bla genes are still frameshifted; Middle: in the later subcultures, sequencing diagrams contains two sets of overlapping peaks, the main peaks are the frameshifted bla, the second peaks are the repaired bla; Bottom: in the final fast-growing subcultures, sequencing diagrams contains very few overlapping peaks.
(C) The ClustalW alignment of the wild-type, frameshifted and repaired bla gene sequences: bla+: wild-type; bla-: frameshift; bla*A-E: different (independent) revertants; box: the base G deleted in the OE-PCR; underlined: the bases (in blue) inserted or deleted in the revertants; circles: base substitutions (G:C/\rightarrow/A:T transitions).
(D) The codon and amino acid unified alignment of bla and their translated protein sequences: underlined: the base G deleted in codon 46; red box: stop codons; red star: stop signals.

Fig 3. The traditional model for frameshfr1t repair: random mutagenesis: Step 1: a frameshift mutation occurs in a coding gene, many hidden stop codons (red bars) emerge in the CDS, and in the nonsense mRNAs, cause translational termination and resulted in truncated products; Step 2: the coding gene is repaired by a reverse mutation through random mutagenesis, the reading frame is restored immediately, stop codons are hidden (green bars), and translation proceeds, producing functional mRNAs and proteins, which are identical and same to the original genes and proteins.

Fig 4. NGS resequencing and variation analysis the genomes of the frameshift/revertant:
(A) The genome-wide variation map of the frameshift /revertant: to visualize the levels of read coverages and the distribution of mutations, Circos software was used to show the clean reads of each sample, the depth of coverage (number of clean reads covering a given base position of the reference sequence), and the SNP/indel, using the reference genome sequence as a ruler. The outermost circle: the coordinate position of the reference genome, and then, outside-in, the distribution of indel/SNPs (red dots), the depth of coverage (grey), the GC content (green) and the GC skew value (pink) in the reference genome, respectively. There are two samples, the frameshift
(bla-) and the revertant (bla*), so there are two sets of indel/SNP, and reads coverage depth mapped onto the reference genome sequences.

(B) The distribution of the numbers of SNPs/indels of the frameshift/revertant on the reference genome: The abscissa axis shows the chromosome coordination of the reference genome sequence, and the ordinate axis represents the SNP/indel density on the genome sequence in per 10 kb. The left scale: SNP (red peaks); the right scale: indel (blue peaks). See Table 2 for data.

(C) The types and lengths of the reported “structure variations” of the frameshift/revertant: The horizontal axis represents the range of SV lengths, and the vertical axis represents the number of SVs. See Table 3 for data.

Fig 5. Expression of the PTC-free frameshifted bla in E. coli BL21:

(A) SDS-PAGE of the lysates of pET28a-bla# recombinant E. coli cells; Lane 1: protein marker; Lane 2 and 3: whole-cell lysates of uninduced and induced cells; lane 4 and 5: the supernatant of uninduced and induced cells; (B) Detection of β-lactamase activity by iodimetry: (1) the revertants of E. coli BL21/pET28a-bla- caused the fading of the iodine/starch solution, suggesting that their bla- was repaired, produced active β-lactamases, and transformed ampicillin into penicillium thiazole acid, which bind with starch and competitively inhibit the binding of iodine to starch; (2) E. coli BL21/pET28a-bla# did not cause fading, suggesting that it was not repaired but expressed a frameshifted BLA (BLA#), which is an inactive product; (3) the negative control (E. coli BL21), no β-lactamase, no fading.

Fig 6. The transcriptome analysis of different E. coli strains: W-1: wildtype; FS-2: frameshift; R-3: a revertant; RE-4: a subculture of the revertant; (A) The number of differentially expressed genes (DEGs) and the cluster analysis of DEGs; (B) the most enriched GO terms by comparing the frameshift (FS_2) with the wild type (W_1); (C) the most enriched GO terms by comparing the revertant (R_3) with the frameshift (FS_2); (D) DNA replication pathway: (E) genes upregulated in the RNA degradation/processing pathway; (F) genes upregulated in the mismatch repair and homologous recombination pathways. Underlined in B, C: GO terms involve in the processing of DNA or RNA, which probably also involve in frameshift repair; in D, E, F: green box: genes involve in the pathways; rex box: genes upregulated in the frameshift or revertant.

Fig 7. Targeted repair of the frameshift mutation induced by DNA or RNA oligonucleotides.

(A) Both sense- and antisense-strand of DNA and RNA oligonucleotides can promote the repair of the frameshift mutation: Control: no oligo; ssDNA: sense-strand DNA; asDNA: antisense DNA; dsDNA: double-stranded DNA; ssRNA: sense-strand RNA; asRNA: antisense RNA; dsRNA: double-stranded RNA; each group were transformed with 10.0 nM of the corresponding oligo. (B) The statistical chart of the average recovery rates, which are the numbers of revertants found in each of the ampicillin plates divided by the numbers of colonies in the parallel tetracycline plates; (C) Sanger sequencing diagram of the DNA-induced revertants: top: induced by a wild-type ssDNA; bottom: induced by a mutant ssDNA with a C-to-A substitution; (D) Sanger sequencing diagram of the RNA-induced revertants: top: induced by a wild-type ssRNA; bottom: induced by a mutant ssRNA with a G-to-A substitution.

Fig 8. The new model for frameshift repair: nonsense-mediated gene editing (NMGE): Step 1: when the frameshifted gene is transcribed and translated, the nonsense mRNAs containing many PTCs (yellow bars) are identified by mRNA surveillance through PTCs signaling; Step 2: some of the nonsense mRNAs causing translational termination are degraded by NMD; Step 3: some of the nonsense mRNAs are translated by translational frameshifting or readthrough; Step 4: some of
the nonsense mRNAs are repaired by RNA editing, PTCs are replaced by sense codons (green bars) through base insertion, deletion and substitution; Step 5: some of the original or repaired nonsense mRNAs are transported to identify and localize the defective coding gene; Step 6: the coding gene is repaired by RNA-directed DNA repair, the reading frame is restored partially, and the translation proceed, the remaining stop codons were readthrough, producing functional mRNAs and proteins with many variations, and so that the host recovered gradually.
Table 1. The natural suppressor tRNA for nonsense mutations (the readthrough rules).

<table>
<thead>
<tr>
<th>Site</th>
<th>tRNA (AA)</th>
<th>Wild type Code</th>
<th>Anti-code</th>
<th>Correction Code</th>
<th>Anti-code</th>
</tr>
</thead>
<tbody>
<tr>
<td>supD</td>
<td>Ser (S)</td>
<td>→ UCG</td>
<td>CGA←</td>
<td>→ UAG</td>
<td>CU←</td>
</tr>
<tr>
<td>supE</td>
<td>Gln (Q)</td>
<td>→ CAG</td>
<td>CUG←</td>
<td>→ UAG</td>
<td>CU←</td>
</tr>
<tr>
<td>supF</td>
<td>Tyr (Y)</td>
<td>→ UAC</td>
<td>GUA←</td>
<td>→ UAG</td>
<td>CU←</td>
</tr>
<tr>
<td>supG</td>
<td>Lys (K)</td>
<td>→ AAA</td>
<td>UU←</td>
<td>→ UAA</td>
<td>UU←</td>
</tr>
<tr>
<td>supU</td>
<td>Trp (W)</td>
<td>→ UGG</td>
<td>CCA←</td>
<td>→ UGA</td>
<td>UCA←</td>
</tr>
</tbody>
</table>

Table 2. The summary of SNP/indel variations in the bla*/- genome

<table>
<thead>
<tr>
<th>Sample</th>
<th>Indels</th>
<th>SNPss</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INS</td>
<td>DEL</td>
</tr>
<tr>
<td>bla-</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>bla*</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Note: INS: insertion; DEL: deletion; ts: transitions; tv: transversions;

Table 3. The types and lengths of the “structure variations”

<table>
<thead>
<tr>
<th>Sample</th>
<th>Type</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INS</td>
<td>DEL</td>
</tr>
<tr>
<td>Frameshift</td>
<td>77</td>
<td>65</td>
</tr>
<tr>
<td>Revertant</td>
<td>60</td>
<td>50</td>
</tr>
</tbody>
</table>

Note: INS: insertion; DEL: deletion; INV: inversion.
Table 4. Possible NMGE-relevant genes and pathways that were upregulated in the frameshift repair.

<table>
<thead>
<tr>
<th>Pathway Term</th>
<th>Database</th>
<th>ID</th>
<th>Input number</th>
<th>Background number</th>
<th>P-Value</th>
<th>Corrected P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC transporters</td>
<td>KEGG</td>
<td>eco02010</td>
<td>32</td>
<td>171</td>
<td>0.0001</td>
<td>0.0090</td>
</tr>
<tr>
<td>RNA degradation</td>
<td>KEGG</td>
<td>eco03018</td>
<td>1</td>
<td>15</td>
<td>0.7419</td>
<td>0.9762</td>
</tr>
<tr>
<td>DNA replication</td>
<td>KEGG</td>
<td>eco03030</td>
<td>1</td>
<td>17</td>
<td>0.7822</td>
<td>0.9762</td>
</tr>
<tr>
<td>Mismatch repair</td>
<td>KEGG</td>
<td>eco03430</td>
<td>1</td>
<td>22</td>
<td>0.8575</td>
<td>0.9796</td>
</tr>
<tr>
<td>Homologous recombination</td>
<td>KEGG</td>
<td>eco03440</td>
<td>1</td>
<td>27</td>
<td>0.9069</td>
<td>0.9796</td>
</tr>
</tbody>
</table>

Table 5. Possible NMGE-relevant genes and pathways that were upregulated in the revertant.

<table>
<thead>
<tr>
<th>Pathway Term</th>
<th>Database</th>
<th>ID</th>
<th>Input number</th>
<th>Background number</th>
<th>P-Value</th>
<th>Corrected P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC transporters</td>
<td>KEGG</td>
<td>eco02010</td>
<td>32</td>
<td>171</td>
<td>3.53E-05</td>
<td>0.0018</td>
</tr>
<tr>
<td>Homologous recombination</td>
<td>KEGG</td>
<td>eco03440</td>
<td>2</td>
<td>27</td>
<td>0.6539</td>
<td>0.9519</td>
</tr>
<tr>
<td>RNA degradation</td>
<td>KEGG</td>
<td>eco03018</td>
<td>3</td>
<td>15</td>
<td>0.1017</td>
<td>0.5470</td>
</tr>
<tr>
<td>DNA replication</td>
<td>KEGG</td>
<td>eco03030</td>
<td>1</td>
<td>17</td>
<td>0.7564</td>
<td>0.9519</td>
</tr>
<tr>
<td>Mismatch repair</td>
<td>KEGG</td>
<td>eco03430</td>
<td>1</td>
<td>22</td>
<td>0.8357</td>
<td>0.9519</td>
</tr>
</tbody>
</table>
References


7. Streisinger, G., et al., Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. Cold Spring Harb Symp Quant Biol, 1966. 31: p. 77-84.


Premature termination codons


Fig 1. The frameshift mutagenesis of the *bla* gene in the plasmid pBR322

Wild-type (*bla +*):  ...A  T  C  G  A  A  C...

|   |   |   |   |   |   |   |

Frameshift (*bla -*):  ...A  T  C  -  A  A  C...
Fig 2A

**ACP**
- No dilution

**TCP**
- 10^{-7} dilution

**bla+**

**bla-**

**blank**
Fig 2B

Initial revertants (grow Slowly)

Frameshift: AGTGGGTTACATC AACTGGATCTCAAGCGGTAAAG
Wild-type: AGTGGGTTACATC AACTGGATCTCAACAGCGGTAAAG

Later subculture (grow faster)

Frameshift: AGTGGGTTACATC AACTGGATCTCAAGCGGTAAAG
Wild-type: AGTGGGTTACATC AACTGGATCTCAACAGCGGTAAAG

Final subculture (grow fast)

Revertants: AGTGGGTTACATC AACTGGATCTCAAGCGGTAAAG
Wild-type: AGTGGGTTACATC AACTGGATCTCAACAGCGGTAAAG
### Fig 2D

| Seq | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| bta | atg | att | caa | cat | ttc | ggt | tgc | gcc | ctt | att | ccc | ttt | gtt | gct | cca | gaa | acg | ctg | tgt | ggt | gca | cga | gtt | gtt | ggt | gga | ata | aat | gaa | ggt | gat | cag | tgt | cac | caa | gat | tgt | aag | ctg |
| bta | atg | att | caa | cat | ttc | ggt | tgc | gcc | ctt | att | ccc | ttt | gtt | gct | cca | gaa | acg | ctg | tgt | ggt | gca | cga | gtt | gtt | ggt | gga | ata | aat | gaa | ggt | gat | cag | tgt | cac | caa | gat | tgt | aag | ctg |
| bta | atg | att | caa | cat | ttc | ggt | tgc | gcc | ctt | att | ccc | ttt | gtt | gct | cca | gaa | acg | ctg | tgt | ggt | gca | cga | gtt | gtt | ggt | gga | ata | aat | gaa | ggt | gat | cag | tgt | cac | caa | gat | tgt | aag | ctg |
| bta | atg | att | caa | cat | ttc | ggt | tgc | gcc | ctt | att | ccc | ttt | gtt | gct | cca | gaa | acg | ctg | tgt | ggt | gca | cga | gtt | gtt | ggt | gga | ata | aat | gaa | ggt | gat | cag | tgt | cac | caa | gat | tgt | aag | ctg |
| bta | atg | att | caa | cat | ttc | ggt | tgc | gcc | ctt | att | ccc | ttt | gtt | gct | cca | gaa | acg | ctg | tgt | ggt | gca | cga | gtt | gtt | ggt | gga | ata | aat | gaa | ggt | gat | cag | tgt | cac | caa | gat | tgt | aag | ctg |
| bta | atg | att | caa | cat | ttc | ggt | tgc | gcc | ctt | att | ccc | ttt | gtt | gct | cca | gaa | acg | ctg | tgt | ggt | gca | cga | gtt | gtt | ggt | gga | ata | aat | gaa | ggt | gat | cag | tgt | cac | caa | gat | tgt | aag | ctg |
| bta | atg | att | caa | cat | ttc | ggt | tgc | gcc | ctt | att | ccc | ttt | gtt | gct | cca | gaa | acg | ctg | tgt | ggt | gca | cga | gtt | gtt | ggt | gga | ata | aat | gaa | ggt | gat | cag | tgt | cac | caa | gat | tgt | aag | ctg |
| bta | atg | att | caa | cat | ttc | ggt | tgc | gcc | ctt | att | ccc | ttt | gtt | gct | cca | gaa | acg | ctg | tgt | ggt | gca | cga | gtt | gtt | ggt | gga | ata | aat | gaa | ggt | gat | cag | tgt | cac | caa | gat | tgt | aag | ctg |

### Protein Sequences

| Seq | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| bta | LNS | GKI | LES | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE |
| bta | STAV | RSL | RLF | VAP | KNP | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN |
| bta | STAV | RSL | RLF | VAP | KNP | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN |
| bta | STAV | RSL | RLF | VAP | KNP | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN |
| bta | STAV | RSL | RLF | VAP | KNP | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN | VFN |

### Protein Sequences

| Seq | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 |
|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| bta | LNS | GKI | LES | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE | FRP | ERE |
Random mutagenesis

**Frameshift**

1. **Step 1**
   - Transcription
   - mRNA
   - Translational termination
     - Truncated protein
     - Die

2. **Step 2**
   - Reverse mutation
   - DNA
   - Transcription
   - mRNA
   - Translation
     - Functional protein
     - Restore
Fig 5. *Expression of the PTC-free frameshifted bla (bla#)*
Fig 6. Transcriptome Analysis

6A Numbers of differentially expressed genes (DGE)
The Most Enriched GO Terms (FS_2vsW_1)

- quorum sensing
- homeostasis of number of cells
- establishment of localization
- xenobiotic metabolic process
- response to xenobiotic stimulus
- phenylacetate catabolic process
- benzene-containing compound metabolic process
- cellular response to xenobiotic stimulus
- multi-organism cellular process
- oligosaccharide metabolic process
- metal ion transport
- oligosaccharide biosynthetic process
- multi-organism process
- trehalose metabolic process
- cell division
- ion transport
- single-organism carbohydrate metabolism
- viral nucleocapsid
- virion part
- ATP-binding cassette (ABC) transport activity
- DNA polymerase activity
- phosphogluconate dehydrogenase activity
- DNA-directed DNA polymerase activity
- transition metal ion transmembrane transport activity

Number of genes

- biological_process
- cellular_component
- molecular_function
The Most Enriched GO Terms (R_3 vs FS_2)

- oxidation-reduction process
- respiratory electron transport...
- electron transport chain
- organ morphogenesis
- odontogenesis of dentin-contai...
- odontogenesis
- single-organism metabolic proc...
- lipid metabolic process
- copper ion transport
- copper ion transmembrane transp...
- cellular lipid metabolic proces...
- mismatch repair
- organ development
- chloroplast
- plastid
- oxidoreductase activity
- threonine-phosphate decarboxyl...
- extracellular matrix structura...
- structural constituent of toot...
- aspartic-type endopeptidase ac...
- aspartic-type peptidase activi...
- cytokine binding
- chemokine binding
- oxidoreductase activity, actin...
- copper ion transmembrane trans...
- double-stranded DNA binding
- mismatched DNA binding
- ubiquinol-cytochrome-c reductas...
- oxidoreductase activity, acting...
- extracellular matrix structural...

Number of genes

Type:
- biological_process
- cellular_component
- molecular_function
6D genes upregulated in DNA replication
6E genes upregulated in RNA degradation/processing
6F genes upregulated in mismatch repair/homologous recombination
Fig 7A. Frameshift repair induced by DNA or RNA oligos (10 nM)
Fig 7B. The rate of reversions in different group
**Fig 7C.** Sanger sequencing diagram of the ssDNA-induced revertants

**Main Peaks:**

```
TGCACTGGATCTAACAGCGGTA
```

**2^nd Peaks:**

```
TGACGAGTGGTACATCAGG
```

**WT DNA oligo:**

```
TGACGAGTGGTACATCAGG
```

**Wild-type:**

```
TGACGAGTGGTACATCAGG
```
**Fig 7D.** Sanger sequencing diagram of the RNA-induced revertants

**Main Peaks:**


**2nd Peaks:**


**WT RNA oligo:**


**Wild-type:**

Fig 8  Nonsense-mediated gene editing (NMGE)

**Frameshift**

**Step 1** Transcription

DNA

Transcription

**mRNA surveillance**

Nonsense mRNA

**Step 2** Translational Termination

RNA

Translational Termination

NMD

**Step 3** Translation

Protein

No protein

**Step 4** RNA editing

**Step 5** RNA-directed DNA repair (MMR+HR)

CDS

Transcription

Direct NMGE

Indirect NMGE

Translational Frameshifting or readthrough

**Step 6** Translational Termination

**Step 7** Translation

**Step 1** Non-functional

**Step 2** Non-functional

**Step 3** Functional

**Step 4** Functional

**Step 5** Functional

**Step 6** Functional

**Step 7** Functional

**Step 1** Die

**Step 2** Die

**Step 3** Survive

**Step 4** Restore

**Step 5** Restore

**Step 6** Restore

**Step 7** Restore