Mismatch repair at stop codons is directed independent of GATC methylation on the *Escherichia coli* chromosome

Kim Sneppen & Szabolcs Semsey

Center for Models of Life, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark.

The mismatch repair system (MMR) corrects replication errors that escape proofreading. Previous studies on extrachromosomal DNA in *Escherichia coli* suggested that MMR uses hemimethylated GATC sites to identify the newly synthesized strand. In this work we asked how the distance of GATC sites and their methylation status affect the occurrence of single base substitutions on the *E. coli* chromosome. As a reporter system we used a lacZ gene containing an early TAA stop codon. We found that occurrence of point mutations at this stop codon is unaffected by GATC sites located more than 115 base pairs away. However, a GATC site located about 50 base pairs away resulted in a decreased mutation rate. This effect was independent of Dam methylation. The reversion rate of the stop codon increased only slightly in dam mutants compared to mutL and mutS mutants. We suggest that unlike on extrachromosomal DNA, GATC methylation is not the only strand discrimination signal for MMR on the *E. coli* chromosome.

Single base substitutions in DNA typically occur by misincorporation of nucleotides during DNA synthesis. Single base substitutions left behind by the replication complex can be corrected by the mismatch repair (MMR) system. MMR is an evolutionarily conserved mechanism which can function in a relatively short time interval after replication. In case of the *E. coli* MMR system, mismatches are recognized by the MutS protein. After mismatch recognition, the system scans the DNA to find a signal for discrimination of the newly synthesized strand and the template strand. In the widely accepted model, the MMR system uses methylation of adenines at 5′-GATC-3′ sequences as discrimination signals, which is performed by the DNA adenine methylase (Dam) after replication. The MutS protein together with MutL activates the MutH endonuclease, which creates a nick on the unmethylated strand at a hemimethylated GATC located nearby the mismatch. The misincorporated nucleotide is corrected by excision and resynthesis of the DNA strand nicked by MutH. This process requires DNA helicase II (UvrD), single strand DNA binding protein (SSB), and DNA polymerase III. The steps between mismatch recognition and MutH mediated strand incision are much debated. Communication between the mismatch site and the GATC site may involve translocation of MutS along the DNA (cis model) or DNA loop formation between the two sites (trans model).

In the cis model the time frame available for proper action of the MMR system depends on how fast the daughter strand is methylated, on the distance of the discrimination signal from the mismatch, and also on the distance on DNA that the MutS protein needs to scan for locating the discrimination signal.

The reported half-life values for hemimethylated DNA behind the replication fork vary from seconds to several minutes depending on the experimental system used, and also on the action of specific proteins that can hinder methylation of certain DNA regions. Assuming that the migration speed of the replication fork is about 1000 bp/s, there may be a few thousand to several hundred thousand bases of hemimethylated DNA available for the MMR system.

Previous studies demonstrated that GATC sequences affect the repair efficiency of G-T mismatches on artificial bacteriophage heteroduplexes in *E. coli* in a number and distance dependent manner. A single hemimethylated GATC site could serve as a strand discrimination signal for the MMR system as long as its distance from the mismatch was less than 1 kb. The MMR system was not sensitive to the relative orientation of the mismatch to the hemimethylated GATC site, i.e. it could efficiently repair the error on the unmethylated strand from both directions. However, on the *E. coli* chromosome single nucleotide deletion mismatches could be
efficiently repaired even if the closest GATC sequence was 2 kb away, and the chromosomal context had a larger influence on the frame-shift mutation rate than the local GATC content.10

In this work, we investigate the inconsistency of the above results by studying the effect of local chromosomal GATC content on the rate of single base substitutions (SBS). These are recognized by MutS by studying the effect of local chromosomal GATC content on the chromosome. To address this question, we created a chromosomal reporter system in E. coli. First, we created four versions of the lacZ gene, which differed in their GATC content (Figure 1, A–D). Constructs ‘A’ and ‘B’ carry the wild type lacZ gene containing 14 GATC sites. In constructs ‘C’ and ‘D’, all these sites are eliminated by sense mutations. Constructs ‘A’ and ‘C’ contain a 12 bp insertion right upstream of the lacZ gene. This 12 bp sequence contains a GATC site.

Because appearance and repair of mismatches may depend on their local DNA context and chromosomal location,11,13 all constructs were placed at the same chromosomal location. Mutation rates were calculated from the occurrence of single base substitutions in this stop codon which restored a functional lacZ gene (Table 1). We sequenced such functional lacZ genes in 25 colonies which grew on a minimal lactose plate and found mutations in all three positions of the stop codon (Table 2).

In construct ‘B’ the lacZ gene contained the same GATC sites as found in the wild type lacZ gene, with the closest one being about 115 bp from the stop codon. In construct ‘D’, in which all GATC sites were eliminated, the closest GATC site was located upstream in the ybbP gene 2433 bp from the stop codon (the closest downstream GATC site was 5566 bp away). Importantly, we observed the same rate of reversion of the stop codon in ‘D’ as in the wild type construct (‘B’). These results suggest that within these limits (115 to 2433 bp), the distance between the mismatch and the closest GATC site does not affect MMR efficiency on the chromosome.

Figure 1 | Structure and chromosomal context of the reporter constructs used in this study. The reporter constructs (A–E) contained the zeocin resistance cassette and the lacZ gene which was inactivated by the C20A substitution resulting in a stop codon (red line). The positions of GATC sequences in the different constructs are indicated by vertical lines. Arrowheads indicate the direction of transcription. The replication fork proceeds from left to right in this region. The local sequence context of the stop codon is shown on the top. Measured mutation rates (M) and 95% confidence intervals (95% CI) are shown on the right. The mutation rates and 95% confidence intervals observed upon mutS deletion were 3.1 (2.2–4.1), 13.3 (11.2–15.6), 3.8 (2.8–4.9), and 13.6 (11.1–16.3) × 10⁻⁹/generation for strains ‘A’–‘D’, respectively.

Results

The effect of the distance between the mismatch and GATC sites on MMR activity on the chromosome. The expected average distance between GATC sites on DNA is 256 bp but the actual distances on the E. coli chromosome vary from 4 to 4840 bp. Here we asked whether the occurrence of single base substitutions depend on the GATC context on the E. coli chromosome. To address this question, we created a chromosomal reporter system in E. coli (Figure 1) using a mutant β-galactosidase (lacZ) gene which has an early TAA stop codon (codon 7). The TAA stop codon serves as the most abundant stop codon in the different constructs are placed at the same chromosomal location. Mutation rates were calculated from the occurrence of single base substitutions in the different constructs. Mutation rates and 95% confidence intervals observed upon mutS deletion were 3.1 (2.2–4.1), 13.3 (11.2–15.6), 3.8 (2.8–4.9), and 13.6 (11.1–16.3) × 10⁻⁹/generation for strains ‘A’–‘D’, respectively.

ATP-bound MutS can diffuse along naked DNA at 0.1 μm²/s in vitro, and spends about 10 minutes on the DNA having closed boundaries. Therefore, it could find a site located ~2.5 kb away in about 10 seconds. However, diffusion of MutS along the E. coli chromosome in vivo is most likely obstructed by other DNA binding proteins. For example, about one HU dimer is present per 100 bp on the chromosome on average, and slow dissociation of HU dimers from DNA would be a substantial barrier for MutS diffusion to longer distances. That is, prokaryotic MutS proteins face a similar difficulty in reaching a distant site on DNA as eukaryotic MutS homologues do due to the presence of nucleosomes. Nucleosomes dissociate from DNA at a comparable rate to dissociation of HU from the E. coli chromosome. The eukaryotic mismatch recognition heterodimer hMSH2-hMSH6 is able to facilitate the disassembly of nucleosomes, however, the process requires a relatively long time (τ, of 23 to 117 minutes, depending on the modification of the nucleosome). The E. coli chromosome is covered by DNA binding proteins of diverse nature. Therefore, it is unlikely that MutS could facilitate their disassociation universally.

Figure 1 | Structure and chromosomal context of the reporter constructs used in this study. The reporter constructs (A–E) contained the zeocin resistance cassette and the lacZ gene which was inactivated by the C20A substitution resulting in a stop codon (red line). The positions of GATC sequences in the different constructs are indicated by vertical lines. Arrowheads indicate the direction of transcription. The replication fork proceeds from left to right in this region. The local sequence context of the stop codon is shown on the top. Measured mutation rates (M) and 95% confidence intervals (95% CI) are shown on the right. The mutation rates and 95% confidence intervals observed upon mutS deletion were 3.1 (2.2–4.1), 13.3 (11.2–15.6), 3.8 (2.8–4.9), and 13.6 (11.1–16.3) × 10⁻⁹/generation for strains ‘A’–‘D’, respectively.

In construct ‘B’ the lacZ gene contained the same GATC sites as found in the wild type lacZ gene, with the closest one being about 115 bp from the stop codon. In construct ‘D’, in which all GATC sites were eliminated, the closest GATC site was located upstream in the ybbP gene 2433 bp from the stop codon (the closest downstream GATC site was 5566 bp away). Importantly, we observed the same rate of reversion of the stop codon in ‘D’ as in the wild type construct (‘B’). These results suggest that within these limits (115 to 2433 bp), the distance between the mismatch and the closest GATC site does not affect MMR efficiency on the chromosome.

If MMR fully depends on the availability of hemimethylated GATC sites, then either the MMR system acts faster or hemimethylated GATC sites are available for a longer time on the chromosome than on bacteriophage heteroduplexes. Hemimethylated GATC sites are typically available for about 1–2 minutes on the chromosome after replication. The time required for recognition of mismatches by MutS is probably not a limiting factor because MutS is associated with the replication complex.

Results

The effect of the distance between the mismatch and GATC sites on MMR activity on the chromosome. The expected average distance between GATC sites on DNA is 256 bp but the actual distances on the E. coli chromosome vary from 4 to 4840 bp. Here we asked whether the occurrence of single base substitutions depend on the GATC context on the E. coli chromosome. To address this question, we created a chromosomal reporter system in E. coli (Figure 1) using a mutant β-galactosidase (lacZ) gene which has an early TAA stop codon (codon 7). The TAA stop codon serves as the most abundant stop codon in the different constructs are placed at the same chromosomal location. Mutation rates were calculated from the occurrence of single base substitutions in the different constructs. Mutation rates and 95% confidence intervals observed upon mutS deletion were 3.1 (2.2–4.1), 13.3 (11.2–15.6), 3.8 (2.8–4.9), and 13.6 (11.1–16.3) × 10⁻⁹/generation for strains ‘A’–‘D’, respectively.

ATP-bound MutS can diffuse along naked DNA at 0.1 μm²/s in vitro, and spends about 10 minutes on the DNA having closed boundaries. Therefore, it could find a site located ~2.5 kb away in about 10 seconds. However, diffusion of MutS along the E. coli chromosome in vivo is most likely obstructed by other DNA binding proteins. For example, about one HU dimer is present per 100 bp on the chromosome on average, and slow dissociation of HU dimers from DNA would be a substantial barrier for MutS diffusion to longer distances. That is, prokaryotic MutS proteins face a similar difficulty in reaching a distant site on DNA as eukaryotic MutS homologues do due to the presence of nucleosomes. Nucleosomes dissociate from DNA at a comparable rate to dissociation of HU from the E. coli chromosome. The eukaryotic mismatch recognition heterodimer hMSH2-hMSH6 is able to facilitate the disassembly of nucleosomes, however, the process requires a relatively long time (τ, of 23 to 117 minutes, depending on the modification of the nucleosome). The E. coli chromosome is covered by DNA binding proteins of diverse nature. Therefore, it is unlikely that MutS could facilitate their disassembly universally.
DNA bound proteins obstruct MutS diffusion depending on their dissociation rate from DNA. We have simulated the potential effect of slower MutS diffusion rate on MMR efficiency (Figure 2). Using different MutS diffusion rates, we computed the probability that MutS reaches a site located 2400 bp away in any of the directions within 90 seconds (typical lifetime of a hemimethylated GATC site). We found that already at a 5-fold slower diffusion rate MutS would not reach the distant GATC site in 10% of the cases, which would result in a detectable increase in the observed mutation rate. For example, assuming that only 1% of mismatches are left uncorrected by the MMR system in the close vicinity of GATC sites, the above 10% increase in the uncorrected errors would result in an about 10-fold increase in the observed mutation rate.

The reversion rate of the stop codon is affected by a GATC site located 50 base pairs away. Constructs ‘A’ and ‘C’ contain a 12 bp GATC containing sequence about 50 bp upstream of the lacZ gene. Construct ‘A’ had a lacZ gene with the wild type GATC pattern, while in ‘C’ all GATC sites were eliminated from lacZ, corresponding to ‘B’ and ‘D’, respectively. We found similar mutation rates of the stop codon in case of ‘A’ and ‘C’, however, these rates were about 4-fold lower than the rates observed in case of constructs ‘B’ and ‘D’.

To determine whether the observed 4-fold difference in the mutation rate is specific to the stop codon used in our reporter system, we compared the single nucleotide substitution rate of the endogenous rpoB gene in strains A–D. We found that mutations in the rpoB gene causing rifampicin resistance appeared with similar probabilities in gene in strains A–D. We found that mutations in the rpoB gene compared the single nucleotide substitution rate of the endogenous lower than the rates observed in case of constructs ‘B’ and ‘D’.

Effect of mutS and mutL deletions on the reversion rate of the stop codon. To test whether the protection provided by the 12 bp insertion resulted from an increased MMR efficiency, we created mutS deletion derivatives of the four strains carrying the constructs placed in the ‘direct’ orientation (Figure 1, A–D). We observed about 20-fold increase in the reversion rate of the stop codon in all the four cases as a result of mutS deletion (see Figure 1 legend and Table 1). Similar results were obtained when we compared mutL deletion derivatives of strains carrying constructs ‘C’ and ‘D’ (Table 1).

![Figure 2](https://www.nature.com/scientificreports)

**Figure 2** Efficiency of mismatch repair as function of MutS diffusion rate along the DNA. At each value of the diffusion rate, MutS located at the mismatch is released to perform one-dimensional random walks along the DNA. The simulation was repeated 10000 times. The efficiency of the mismatch repair is scored as the fraction of the released MutS that reach either a site 2400 bp downstream of the mismatch or a site 5550 bp upstream of the mismatch within 90 seconds. These 90 seconds correspond to the average lifetime of a hemi-methylated GATC. The shown behavior was reproduced (within a factor 2 in diffusion constant) in a more elaborate model where many independently methylated GATC sites (methylated in 90 seconds on average) are placed at 256 bp intervals outside the −2400 to 5550 bp region. In that more complicated model the repair efficiency was scored as the probability that MutS reached any of these sites in a hemimethylated state.

---

**Table 1** The effect of mutS, mutL, and dam deletions on the reversion rate of the stop codon in constructs C and D. 95% confidence intervals are shown in parentheses, while fold changes relative to the corresponding wild type constructs are underlined. Reversion rates observed in the mutL deletion strains reflect that at the last two positions of the stop codon reversion can arise only by transversions, which are less enriched than transitions in the absence of MutL.

| Construct | WT × 10⁻⁹/generation | ΔmutS × 10⁻⁹/generation | ΔmutL × 10⁻⁹/generation | Δdam × 10⁻⁹/generation |
|-----------|-----------------------|-------------------------|--------------------------|------------------------|
| C         | 0.19 (0.1–0.29)       | 3.8 (2.8–4.9); **19.9** | 4.5 (3.5–5.6); **23.6** | 0.63 (0.37–0.93); **3.3** |
| D         | 0.64 (0.43–0.88)      | 13.6 (11.1–16.3); **21.2** | 14.2 (11.8–16.6); **22** | 3.0 (2.2–3.8); **4.6** |

**Table 2** Possible point mutations at the TAA stop codon and their occurrences in 25 revertants

| Sequence | Occurrence | Coded amino acid |
|----------|------------|------------------|
| TAG      | 12         | Leu              |
| AAA      | 5          | Lys              |
| TCA      | 3          | Ser (WT)         |
| TAC      | 3          | Tyr              |
| CAA      | 2          | Gln              |
| GAA      | 0          | Glu              |
| TGA      | 0          | Stop             |
| TAG      | 0          | Stop             |

**Figure 2** Efficiency of mismatch repair as function of MutS diffusion rate along the DNA. At each value of the diffusion rate, MutS located at the mismatch is released to perform one-dimensional random walks along the DNA. The simulation was repeated 10000 times. The efficiency of the mismatch repair is scored as the fraction of the released MutS that reach either a site 2400 bp downstream of the mismatch or a site 5550 bp upstream of the mismatch within 90 seconds. These 90 seconds correspond to the average lifetime of a hemi-methylated GATC. The shown behavior was reproduced (within a factor 2 in diffusion constant) in a more elaborate model where many independently methylated GATC sites (methylated in 90 seconds on average) are placed at 256 bp intervals outside the −2400 to 5550 bp region. In that more complicated model the repair efficiency was scored as the probability that MutS reached any of these sites in a hemimethylated state.
Table 3 | Mutation rates of sequence variants of construct ‘C’. The 12 bp sequence insertion (bold-faced) and its sequence context in construct ‘C’ are shown on the top. Sequence changes in the two sequence variants are marked red. Measured mutation rates (M) and 95% confidence intervals (95% CI) are shown on the right.

| Sequence                        | Mutation Rate x 10⁻⁹/generation |
|--------------------------------|---------------------------------|
| GAATTCGGGGGATGCTCTAGA           | 0.19 (0.1–0.29)                 |
| GAATTCGGGGGTTCCTCTAGA           | 0.99 (0.68–1.3)                 |
| GAATTCGGGITTGATCTAGA            | 0.13 (0.06–0.21)                |

Mismatch repair is directed in the absence of Dam methylation.

The 12-bp insertion present in constructs 'A' and 'C' contains a GATC site, which can be methylated by the Dam methylase, and which can serve as a strand discrimination site. To test the role of GATC methylation in mismatch repair occurring at the stop codon, we created dam deletion derivatives of cells carrying constructs 'C', which has a single GATC sequence in the reporter region, and 'D', which has none. Elimination of GATC methylation resulted in a relatively small increase (~4-fold) in the reversion rate of the stop codon compared to the 20-fold increase observed in the cases of mutS and mutL deletions (Table 1).

Discussion

In this work we studied the effect of local chromosomal GATC content on the rate of single base substitutions (SBS). We found that within the limits of 115 to 2433 bp, the distance between the mismatch and the closest GATC site does not affect MMR efficiency on the chromosome. This observation is in agreement with the findings of Martina et al., who reported that frameshift mutation rate on the E. coli chromosome is independent of the distance to GATC sites located about 200 to 2000 bp away. However, we found that a GATC site located about 50 base pairs from the stop codon could provide a short range protection from single base substitutions. This protection was independent of Dam methylation, MutS, and MutL.

Although we do not yet understand the mechanism underlying this observation, our results suggest that the GATC content on the chromosome may influence the mutation rate at different locations. Such regulation could become important under conditions where MutS is depleted and therefore the point mutation rate is higher. However, the potential protective function of GATC sites is counteracted in E. coli because methylated bases are mutational hotspots on the chromosome.

In agreement with previous reports, we observed that dam mutants have weaker mutator phenotype than mutL and mutS mutants. If GATC methylation is the only strand discrimination signal for MMR, then the dam mutation is expected to have the same effect on the mutation rate as mutS or mutL mutations. There are two possible explanations for this discrepancy. One is that some of the dam mutant cells are lost due to cell death, and the other is that the MMR system is able to correct the majority of replication errors in the absence of Dam methylation.

Mismatch repair creates double strand breaks in dam mutant cells due to MutH mediated cleavage of both unmethylated strands at GATC sequences located nearby mismatches, which may result in cell death. To account for the lower mutation frequency of Δdam cells compared to ΔmutS cells, about 4 out of 5 mutants (~80%) must be lost due to the above process. However, the misincorporation rate at replication is roughly one per replicated genome.

Therefore, loss of 80% of mismatches in the dam mutant would result in a substantial increase in the population doubling time, which was not observed. Also, double strand breaks are most likely repaired efficiently and do not persist for a long time in dam mutants.

Although previous experiments showed that the old and new strands can be discriminated by the MMR system based on their methylation status, this does not mean that it is the only signal used to discriminate the strands. Claverys and Mejean demonstrated that in a GATC free plasmid system, 50–70% of replication errors occurring at a TAG stop codon can be corrected in a mutSL dependent way. Our experimental results and theoretical simulations support this model, i.e. that the MMR system can correct replication errors in the absence of Dam methylation. In our system, we observed only about 4-fold higher reversion rate at the stop codon in the dam mutants compared to wild type cells, while mutS deletion gave a 20-fold increase in reversion. The effect of dam deletion could be explained in part by the random replication initiation process in the dam mutants. However, we observed no differences in the reversion rate at the 20-fold increase observed in the cases of mutS and mutL deletions (Table 1).

Methods

Plasmid and strain construction.

The kanamycin resistance gene from plasmid pCM1013 was PCR amplified using primers 5'-GGGTAACGGCGAGGTG-3' (CTCAGTCACGACGTTGTAAAAC GACGGCCAGTGAATCAGAGTCCCG-3') and 5'-ACATAATGAGGTTGTTCTAAGGAGGAA-TACGGCACAGATGCGTCCGCTGTGGTTATATTTATAGTCTCCGGCTCTAGAAGAAGCTGTCG-3' and inserted into the plasmid pEM7/zeo (Invitrogen). All GATC sites were eliminated by silent substitutions. The kanamycin resistance gene was transcribed constitutively from the lac promoter.

The dam deletion was created in plasmid pEM7/zeo (Invitrogen). We have made a C→T substitution at position 2299 to eliminate the GATC site from the lac operon. The different versions of the lacZ gene, differing in their GATC content, were placed downstream of the lacI gene and inserted into the E. coli chromosome by recombinogenic (Xie et al. 2003), between positions 523237 and 523641. The sequences of the chromosomal constructs and their ~300 bp flanking regions (see Supplementary Material) were verified (Eurofins MWG Operon).

In the ΔmutS strain the lacI gene was replaced by a chloramphenicol resistance gene (Cat). The CmR cassette in plasmid pBBI32 was PCR amplified using primers 5'-ATGAGCTCATGAAATATTTGCAGCCCGATCAGTTATCTCAGGATGACTAACACCCGACATCCCGTGGCATTAGGCAC-3' and 5'-TTA CACCCGCCCTTCTACAGGCTATCATA-3' and inserted into the rpsL gene on the E. coli chromosome by recombination, between positions 523237 and 523641. The sequences of the chromosomal constructs and their ~300 bp flanking regions (see Supplementary Material) were verified (Eurofins MWG Operon).

In the ΔmutSΔdam strain the lacI gene was replaced by a chloramphenicol resistance gene (Cat). The CmR cassette in plasmid pBBI32 was PCR amplified using primers 5'-ATGAGCTCATGAAATATTTGCAGCCCGATCAGTTATCTCAGGATGACTAACACCCGACATCCCGTGGCATTAGGCAC-3' and 5'-TTA CACCCGCCCTTCTACAGGCTATCATA-3' and inserted into the rpsL gene on the E. coli chromosome by recombination, between positions 523237 and 523641. The sequences of the chromosomal constructs and their ~300 bp flanking regions (see Supplementary Material) were verified (Eurofins MWG Operon).

In the strain in which ΔmutSΔdam strain the lacI gene was replaced by a chloramphenicol resistance gene (Cat). The CmR cassette in plasmid pBBI32 was PCR amplified using primers 5'-ATGAGCTCATGAAATATTTGCAGCCCGATCAGTTATCTCAGGATGACTAACACCCGACATCCCGTGGCATTAGGCAC-3' and 5'-TTA CACCCGCCCTTCTACAGGCTATCATA-3' and inserted into the rpsL gene on the E. coli chromosome by recombination, between positions 523237 and 523641. The sequences of the chromosomal constructs and their ~300 bp flanking regions (see Supplementary Material) were verified (Eurofins MWG Operon).

Deletions were confirmed by PCR, and elimination of GATC methylation was further confirmed by digestion of genomic DNA extracts by Ksp22I (SibEnzyme), which is marker of the lac operon. In this strain (ΔmutSΔdam) the kanamycin resistance gene and the lacI gene are transcribed constitutively from the lac promoter.

The reporter construct was first assembled in plasmid pEM7/zeo (Invitrogen). We have made a C→T substitution at position 2299 to eliminate the GATC site from the lac operon. The different versions of the lacZ gene, differing in their GATC content, were placed downstream of the lacI gene on the E. coli chromosome by recombination, between positions 523237 and 523641. The sequences of the chromosomal constructs and their ~300 bp flanking regions (see Supplementary Material) were verified (Eurofins MWG Operon).
overnight at 37°C using rifampicin and counting the colonies which appeared after incubation of plates

0.004% (w/v) vitamin B1. Plates were incubated at 37°C for 3 days and blue colonies were counted. The occurrence of rifampicin resistant cells was determined by placing 10 μl of cell suspension on LB plates containing 30 μg/ml kanamycin and 30 μg/ml rifampicin and counting the colonies which appeared after incubation of plates overnight at 37°C. Mutation rates were determined from the distribution of the number of mutants in the cultures by the MISS-Maximum Likelihood Estimator Method, using the FALCOR web tool.

Model of MutS diffusion along the DNA. To simulate the mismatch repair efficiency through MutS mediated mismatch localization, we constructed a model where MutS is released from the mismatch site at time zero, and moves along the DNA by random diffusion. We assume that MutS locates the mismatch immediately after the replication fork has left the site. For a given assumed diffusion constant D, the MutS molecule subsequently step a distance l = 15 nm right or left for each time step dt = 0.002D. The nearest GATC upstream is 2433 bp and the nearest GATC downstream is 5566 bp. This correspond to position −55 and +126 in units of the lattice size of l = 15 nm. The GATC site is reached if and only if the released MutS passes either of these two positions within the time frame of 90 seconds. The 90 seconds mimics the average lifetime of hemi-methylated GATC sites on the DNA. For each value of the diffusion constant D, 1000 MutS releases were examined. Figure 2 shows the probabilities of reaching a GATC site calculated from the simulations.

Importantly, we also examined an extended model that includes all GATC sites beyond the −2433 bp and the +5566 bp positions, and in addition, also includes the fact that each hemi-methylated site becomes fully methylated with a rate of 1/900 seconds. The resulting curve for probability to reach a GATC is similar to the one shown in Figure 2, apart from being displaced by about a factor 2 to the left (GATC localization can be done with half of the shown diffusion constant).

1. Kunkel, T. A. & Erie, D. A. DNA mismatch repair. Annu Rev Biochem 74, 681–710, doi:10.1146/annurev.biochem.74.082803.133243 (2005).
2. Modrich, P. Mechanisms and biological effects of mismatch repair. Annu Rev Genet 25, 229–253, doi:10.1146/annurev.ge.25.120190.001305 (1991).
3. Kolodner, R. D., Mendillo, M. L. & Putnam, C. D. Coupling distant sites in DNA during DNA mismatch repair. Proc Natl Acad Sci U S A 104, 12953–12954, doi:10.1073/pnas.0705691104 (2007).
4. Campbell, J. L. & Kleckner, N. E. coli orC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromatid replication fork. Cell 62, 967–979 (1990).
5. Sanchez-Romero, M. A. et al. Dynamic distribution of seqA protein across the chromosome of escherichia coli K-12. MboI 1, doi:10.1186/mbio.00121-10 (2010).
6. Marians, K. J. Prokaryotic DNA replication. Annu Rev Biochem 61, 673–719, doi:10.1146/annurev.bi.61.070192.033245 (1992).
7. Bruni, R., Martin, D. & Juricj, I. d(GATC) sequences influence Escherichia coli mismatch repair in a distance-dependent manner from positions both upstream and downstream of the mismatch. Nucleic Acids Res 16, 4873–4890 (1988).
8. Lahue, R. S., Su, S. S. & Modrich, P. Requirement for d(GATC) sequences in mismatch repair. Proc Natl Acad Sci U S A 94, 1482–1486 (1997).
9. Cooper, D. L., Lahue, R. S. & Modrich, P. Methyl-directed mismatch repair is bidirectional. J Biol Chem 268, 1823–1829 (1993).
10. Martina, M. A., Correa, E. M., Argarana, C. E. & Barra, J. L. Plasmid expression of mutS, -L and/or -H gene in Escherichia coli dam cells results in strains that display reduced mutation frequency. Mutat Res 637, 197–204, doi:10.1016/j.mrrev.2007.07.006 (2008).
11. Robbins-Manke, J. L., Zdarsky, Z. Z., Marinus, M. & Essigmann, J. M. Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient Escherichia coli. J Bacteriol 187, 7027–7037, doi:10.1128/JB.187.20.7027-7037.2005 (2005).
12. Lahue, R. S., Su, K. G. & Modrich, P. DNA mismatch correction in a defined system. Science 245, 160–164 (1989).
13. Fekete, R. A. & Chattoraj, D. K. A cis-acting sequence involved in chromosome segregation in Escherichia coli. Mol Microbiol 55, 175–183, doi:10.1111/j.1365-2958.2004.03492.x (2005).
14. Danesko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97, 6640–6645, doi:10.1073/pnas.120163297 (2000).
15. Rosche, W. A. & Foster, P. L. Determining mutation rates in bacterial populations. Methods 20, 4–17, doi:10.1016/meth.1999.09.001 (2000).
16. Hall, B. M., Ma, C. X., Liang, P. & Singh, K. K. Fluctuation analysis Calculator: a web tool for the determination of mutation rate using Luria-Delbrück fluctuation analysis. Bioinformatics 25, 1564–1565, doi:10.1093/bioinformatics/btp233 (2009).

Acknowledgments
This research was supported by the Danish National Research Foundation. We thank Sine Løvssen and Jan B. Dodd for critically reading the manuscript.

Author contributions
S.S. and K.S. designed the experiments. S.S. performed in vivo experiments, K.S. performed in vitro experiments. S.S. and K.S. wrote the paper. S.S. and K.S. designed the experiments. S.S. performed in vivo experiments, K.S. performed in vitro experiments. S.S. and K.S. wrote the paper.

Additional information
Supplementary information accompanies this paper at http://www.nature.com/scientificreports

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

How to cite this article: Sneppen, K. & Semsey, S. Mismatch repair at stop codons is directed independent of GATC methylation on the Escherichia coli chromosome. Sci. Rep. 4, 7346; DOI:10.1038/1057346 (2014).