THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 258, No. 16, Issue of August 25, pp. 10007-10012, 1983
Printed in U.S.A.

Codon Usage and Mistranslation

IN VIVO BASAL LEVEL MISREADING OF THE MS2 COAT PROTEIN MESSAGE*

(Received for publication, December 3, 1982)

Jack Parker‡, Timothy C. Johnston‡, Peter T. Borgia§, Greg Holtz‡, Erik Remaut¶, and Walter Fiers¶

From the ‡Department of Microbiology, Southern Illinois University, Carbondale, Illinois 62901, the §Department of Medical Microbiology and Immunology, Southern Illinois University School of Medicine, Springfield, Illinois 62708, and the ¶Laboratory of Molecular Biology, State University of Ghent, Ledenganckstraat, 35 B-9000 Ghent, Belgium

The coat protein of the small RNA virus MS2 shows charge heterogeneity *in vivo*. In most strains there is a basic satellite of the native protein. We have shown that this basic satellite is greatly diminished or absent in strains with the streptomycin-resistant allele, *rpsL*, a mutation which leads to increased translational accuracy. Further, the satellite is present in cells where the coat protein is encoded by duplex DNA. Tryptic digests of the satellite show that it contains new lysine-containing peptides which appear to be the same as those found in derivatives of coat protein which have a lysine for asparagine substitution. Sequencing of the NH_2-terminal 19 amino acids of the satellite protein shows that the asparagine codon AAU at amino acid 12 is misread approximately 8 times more frequently than the AAC at amino acid 3. We conclude that the satellite species is the result of basal level lysine for asparagine substitution. These substitutions are most likely caused by preferential misreading of AAU codons at a frequency of approximately 5×10^{-3}, 10-fold higher than the average error frequency.

The recent explosion of DNA sequence data from extremely diverse organisms indicates that, although the genetic code is universal (or nearly so, see Ref. 1), organisms have profound biases in codon use in degenerate codon families (see Ref. 2). Several hypotheses have been put forward to explain this. These include protection against mutation to deleterious codons (3), dinucleotide preferences (4) or avoidances (5), and optimized translational efficiency (6–8). For the chromosomal genes of *Escherichia coli* it is quite clear that in genes which are highly expressed, *i.e.* many protein monomers per cell, there is a strong preference for codons which are read by the most abundant isoaccepting tRNA (7, 9). However it is not clear if this is related to the cause of codon bias or is an effect of it, since tRNA synthesis may be regulated by codon demand (see Ref. 10).

However there are also strong codon preferences among synonymous codons which are read by the same tRNA such as the histidine, asparagine, and lysine codons of *E. coli*. These codons may be used preferentially because they optimize codon-anticodon interactions (see Refs. 7 and 8). Some *E. coli* chromosomal genes do not show this pattern of codon usage. These genes are usually poorly expressed.

* This research was supported by Grant GM25855 from the National Institute of General Medical Sciences. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We have been interested in studying errors in translation which occur *in vivo*. It has been demonstrated that very high levels of mistranslation (greater than 0.1 error/cognate codon) occurs when cells are starved for asparagine or histidine (see Refs. 11 and 12). The observed mistranslation seems to be a result of misreading of the third position of the codon (13, 14). Because the substituted amino acids differ in charge and the resulting mistranslated protein is stable (15), this type of mistranslation is relatively easy to quantify. Starvation for other amino acids does not always lead to such high levels of mistranslation (see Refs. 14 and 15). Interestingly there is a very strong preferential codon use in the asparagine and histidine codons of highly expressed *E. coli* genes (see "Discussion").

Using *E. coli* infected with the RNA virus MS2 to study mistranslation during asparagine starvation (14), we found that the coat protein shows charge heterogeneity in unstarved cells. This charge heterogeneity results in the appearance of a single, more basic satellite of the coat protein. This satellite can be explained as being the result of basal level misreading of asparagine codons. However, the required mistranslation frequency per asparagine codon (1 to 2×10^{-3}) would be 5 to 10 times the frequency of basal level amino acid substitutions calculated for other systems (16–18). The charge heterogeneity could also be explained by post-translational modification, a high level of misacylation or mutation of the MS2 genomic RNA. In this paper we show that the heterogeneity is the result of basal level codon misreading. We discuss this finding in relation to codon usage pattern and its possible significance to other studies of basal level mistranslation.

EXPERIMENTAL PROCEDURES

Bacterial Strains, Virus, and Plasmids—The bacterial strains used in this study are all derivatives of *E. coli* K-12. The genotype of each strain is shown in Table I. A single lysate of the bacteriophage MS2 was used throughout the course of these studies. The plasmid pPLaACR26 confers kanamycin resistance and carries a cDNA copy of the entire coding region of MS2 which is under the transcriptional control of the bacteriophage λ promoter P_L (19). The plasmid pRK248At2 confers tetracycline resistance and encodes a temperature-sensitive λ repressor protein (20).

Media—The minimal medium used was buffered with 4-morpholinepropanesulfonic acid (21) and supplemented with glucose, thiamin, and required amino acids. The medium was slightly modified to increase the $CaCl_2$ concentration when the culture was to be infected with MS2 (14). The rich medium used was L-broth (22). Strains carrying the plasmids pPLaACR26 and pRK248At2 were routinely grown in medium containing kanamycin at 50 μg/ml and tetracycline at 10 μg/ml.

Growth and Labeling of Strains Infected with MS2—Cultures to be infected with MS2 were grown aerobically in minimal medium at 37 °C in Erlenmeyer flasks and with rotary shaking. Growth was

TABLE I
Bacterial strains

Strain	Genotype	Source or reference
JK1	*asnA31, asnB32, thi-1, relA, spoT*	14
JK100	*asnA31, asnB32, thi-1, relA, spoT, fuc, lysA*	13
JK120	as JK100 except *fuc*$^+$, *relA*$^+$	13
JK242	as JK1 except *rpsL125*	14
JK306	as JK100 except *rpsL125*	Lab collection
JK336	JK100/pRK248At2, pPLaACR26	Transformation
JK358	JK306/pRK248At2, pPLaACR26	Transformation
JK380	JK120/pRK248At2, pPLaACR26	Transformation

monitored by absorbance at 420 nm. At a cell density of approximately 3×10^8 cells/ml, 10 ml of the culture were infected with MS2 at a multiplicity of infection of 10. At 5-min postinfection, rifampicin was added to give a final concentration of 200 μg/ml. In some experiments the culture was starved for asparagine at 10-min postinfection by filtration and resuspension in medium lacking asparagine (13). A portion of the infected culture was labeled with either [^{35}S]methionine or [^{14}C]lysine as previously described (14). The labeling was from 20- to 40-min postinfection with unstarved cells and 30 to 70 min with starved cells.

Induction of Strains Carrying Plasmids—Cultures of cells containing pPLaACR26 and pRK248At2 were grown overnight at 28 °C in medium containing antibiotics. These cells were diluted 10-fold into L-broth without antibiotics and were incubated at 28 °C until the cell density was approximately 2×10^8 cells/ml (2–3 generations after dilution). At that point half the culture was transferred to 42 °C to induce λ P_L. Incubation was continued for 30 min at 42 °C. The cultures were then quickly chilled.

Electrophoresis—Extracts for electrophoresis were prepared (23) and two-dimensional electrophoresis was carried out by the nonequilibrium pH gradient electrophoresis method of O'Farrell *et al.* (24) for basic proteins. The first dimension was run for 3.5 h at 400 V. The second dimension was a sodium dodecyl sulfate containing 12% or 15% polyacrylamide slab. Gels containing radioactive proteins were stained with Coomassie brilliant blue R, dried, and autoradiograms made (23). Proteins from unlabeled cultures were visualized using the silver-based color staining method of Sammons *et al.* (25). Color-stained gels were filmed under incandescent light with tungsten-biased color film.

Recovery and Analysis of Labeled Coat Protein—Autoradiograms of gels containing [^{14}C]lysine-labeled coat protein were aligned with the dried polyacrylamide gels and areas containing the protein were excised. The gel fragments were washed extensively with 15% (v/v) methanol to remove stain and sodium dodecyl sulfate. The fragments were then lyophilized. The gel fragments were rehydrated in 50 mM $NaHCO_3$, pH 7.8. The protein in the fragments was then digested *in situ* at 37 °C by the addition of *N*-tosyl-L-phenylalanyl chloromethyl ketone-treated trypsin to 40 μg/ml followed by an equal amount after 4 h of incubation. After a total of 8 h of incubation, the liquid was then removed and lyophilized. Alternatively, in some experiments, protein was electroeluted (75 V for 12 h) from the gel pieces into a dialysis bag. The protein was then dialyzed against distilled water and dried. The protein was then treated with trypsin as above.

The lyophilized material was dissolved in 0.2 M pyridine acetate buffer, pH 3.1, and the peptides separated on a column of Aminex 50 W-X4 (0.9 × 40 cm). The peptides were eluted by a 500-ml gradient of 0.2 M pyridine acetate, pH 3.1, to 2.0 M pyridine acetate, pH 5.0, at 50 °C (26). Fractions were collected and dried. Scintillation fluid was added and the samples were counted. In order to more easily compare experiments having different total counts, the percentage of total counts in each fraction was calculated and plotted.

Reverse Phase Chromatography of the Peptides in Peaks A and B—Coat protein satellite labeled with [^{14}C]lysine was electroeluted, digested with trypsin, and peaks A and B (see Fig. 3) isolated as described above. Positions of the peaks were determined by counting the radioactivity in 5% of each fraction. The fractions of peaks A and B were pooled and dried. The peptides of each peak were separated on a Waters HPLC system using a C18 column. The mobile phase used was a linear gradient of 100% 5 mM ammonium acetate, pH 6, to 30% 5 mM ammonium acetate, 70% acetonitrile over 90 min. The positions of the peptides were determined by counting the radioactivity in each fraction.

Sequencing Protein—Automated Edman degradation of the protein in the basic satellite spot was performed using a Beckman 890C sequencer, modified by the addition of a vacuum cold trap, and using 0.1 M Quadrol Beckman program (number 121078) and Beckman reagents. The protein was added in 67% acetic acid. Polybrene (3 mg; Aldrich) was added to prevent loss during degradation. The product of each cycle was placed in a scintillation vial and dried. The radioactivity in each cycle was counted in a scintillation counter.

Chemicals—Reagent grade chemicals were purchased from Fisher or Sigma. Chemicals for electrophoresis are as given in O'Farrell (23). Trypsin was purchased from Worthington. Pyridine was redistilled before use. Aminex 50W-X4 was purchased from Bio-Rad. Film and developing solutions were from Kodak.

RESULTS

Charge Heterogeneity of Coat Protein in MS2-infected Cells—One of the four genes of the RNA bacteriophage MS2 encodes the catalytic subunit of the phage replicase. The replicase is not inhibited by the antibiotic rifampicin. Therefore if infected cells are treated with rifampicin, phage RNA continues to be synthesized and phage-encoded proteins are produced. Late in infection such cells produce primarily MS2 coat protein (14). Fig. 1*A* shows an autoradiogram of a portion of a two-dimensional gel containing [^{35}S]methionine-labeled protein synthesized in the presence of rifampicin 20 min after infection of strain JK1. There is a clearly visible basic satellite spot which contains 2% of the counts found in normal coat protein. This pattern is also observed in other *relA* strains of *E. coli* (results not shown).

Fig. 1*B* shows an autoradiogram from a similar experiment, but using an *rpsL* strain (JK242) congenic to JK1. The satellite is not visible. The level of reduction of the satellite is difficult to quantitate in such experiments because there is typically some streaking of the coat protein. In all experiments

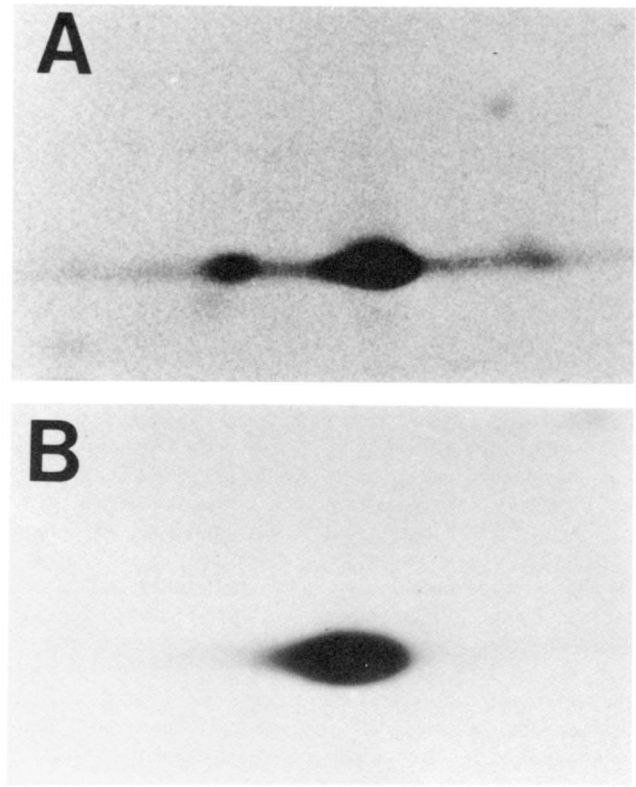

FIG. 1. **Autoradiograms of two-dimensional gels containing coat protein made in MS2-infected, rifampicin-treated cells.** Cells were grown, infected, treated with rifampicin, and labeled as described under "Experimental Procedures." Each *panel* shows only that portion of the autoradiogram showing coat protein. The more acidic portion of the gel is toward the *right*. *A* shows coat protein from JK1 (*rpsL*$^+$). *B* shows coat protein from JK242 (*rpsL*).

with strains carrying the *rpsL 125* allele, the relative level of the basic satellite was reduced a minimum of 2-fold. Thus the charge heterogeneity of coat protein observed in such experiments is reduced or abolished in *E. coli* mutants with ribosomes which read the message with greater fidelity than those of wild type strains (27). This would indicate that most or all of the heterogeneity is a result of misreading, not misacylation.

Charge Heterogeneity of Coat Protein in Cells with the Coat Protein Gene on a Plasmid—Although the above results quite clearly indicate the satellite is caused by codon misreading, the experimental conditions used do not allow one to extrapolate easily to the normal *in vivo* condition. Each cell is synthesizing essentially a single protein. Rifampicin could have effects other than its role in stopping the initiation of RNA synthesis. Further the gene examined is part of an RNA genome. For these reasons we performed experiments utilizing a plasmid which carries the MS2 coding region under the control of λP_L.

When the promoter P_L is induced by shifting the strains carrying pRK248At2 and pPLaACR26 to 42 °C, the MS2 coat protein becomes the major protein species synthesized (19). However other proteins are synthesized and at 30-min postinduction, coat protein synthesis is about 10% of total protein synthesis.

Fig. 2*A* shows a portion of a gel containing total protein from a culture of JK336 harvested 30 min after induction. The *arrow* marks coat protein. The satellite spot is clearly visible. Fig. 2*B* shows a similar gel but from an experiment using JK358, a streptomycin-resistant derivative. No satellite is visible, indicating that, as in the experiments with infected cells, the satellite is the product of codon misreading. A similar experiment was performed with a streptomycin-sensitive (*rpsL*$^+$) stringent strain (*relA*$^+$). These cells also show a satellite spot (results not shown).

Peptide Analysis of Mistranslated Coat Protein—The above experiments clearly indicate that the basic satellite is the result of codon misreading. These results, however, do not allow one to make any meaningful estimates of the error frequency. There are a large number of amino acid substitutions which could result from a single base misreading which would give a more basic protein. Although some of these are unlikely (see Ref. 28), many are not. Therefore we analyzed the mistranslated protein by peptide analysis.

We have already shown that starvation for asparagine greatly increases the mistranslation frequency in bacterial and mammalian cells (11, 13). Further the mistranslated protein made during these conditions is more basic because it contains lysine for asparagine substitutions (14). The MS2 coat protein message is mistranslated at very high levels during asparagine starvation (approximately 0.3 substitutions/asparagine codon). Mistranslated derivatives of coat protein having one extra lysine, comprising the +1 spot on the gel, co-migrate with the extra satellite spot (14).

Peptides from labeled coat protein were isolated from gels as described under "Experimental Procedures." The peptides were then run under conditions where 9 of the 11 native tryptic peptides are resolved. Fig. 3*N* shows the pattern obtained with [^{14}C]lysine-labeled native coat protein. The native lysine-containing peptides are identified using the nomenclature of Konigsberg *et al.* (26). The lysine-containing peptide T10 is not soluble under these conditions. The peptide T7/T9 is a partial digestion product. The peptides were identified by co-migration with chemical amounts of native coat protein and subsequent amino acid analysis.[1] The *+1 panel* of Fig. 3 shows the pattern obtained with [^{14}C]lysine-labeled coat protein made during asparagine starvation. We have previously shown that the +1 spot results from a single lysine for asparagine substitution (14). Two new peaks appear in the chromatogram. These peaks, *A* and *B*, are always found in digestions of the +1 protein and are also present in digestion of coat protein derivatives with as many as seven substitutions. The unidentified peak at fraction 10 is probably a partial digestion product and is not always present in the +1 digest. The satellite spot (*EBS*) shows essentially the same pattern as seen with the +1 spot from starved cells. There are two new lysine-labeled peaks at the same elution position as *A* and *B*.

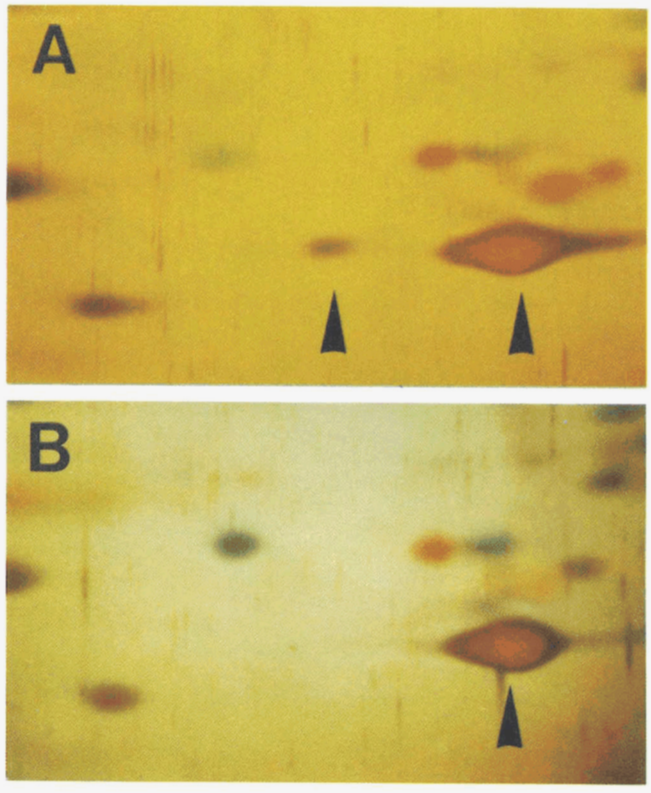

FIG. 2. **Silver-stained two-dimensional gels containing coat protein made from plasmid-encoded genes.** Cells containing the plasmids pRK248At and pPLaACR26 were grown and induced as described under "Experimental Procedures." The cells were harvested 30-min postinduction and two-dimensional gels of total protein were run and stained as described under "Experimental Procedures." The *panels* show only that portion of the gels containing coat protein. The gels are oriented with the more acidic proteins to the *right* and in each case the heavily stained protein marked with an *arrow* is coat protein. *A* shows JK336 (*rpsL*$^+$) and *B* shows JK358 (*rpsL*).

The peptides, whether from native or mistranslated coat protein, are not recovered in equimolar amounts in this system. However, the relative peak heights are extremely reproducible. All experimental evidence indicates that during asparagine starvation the most prevalent error is a lysine for asparagine substitution (13, 14). Further, this error occurs at such a high frequency (0.3 error/codon) that if the +1 spot contained protein species with any other basal level error, it would not be detected. The fact that the satellite peptide map is so similar to the +1 map would argue that the same type of error is occurring, a lysine for asparagine substitution. No soluble native peptide which terminates with lysine contains an asparagine. Therefore the fact that no native lysine-labeled peptide is significantly reduced in the peptide map of the +1 or the satellite spot is also consistent with this type of substitution.

[1] T. C. Johnston and P. T. Borgia, unpublished results.

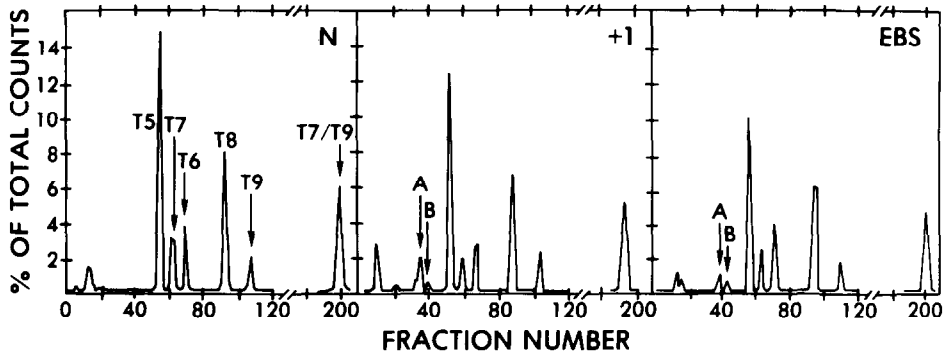

FIG. 3. **Chromatograms of soluble lysine-containing tryptic peptides of MS2 coat protein.** Strain JK1 was grown, infected with MS2, labeled with [^{14}C]lysine, and harvested as described under "Experimental Procedures." Radioactive protein was identified on two-dimensional gels, digested with trypsin, and analyzed by column chromatography also as described under "Experimental Procedures." Fractions of 2 ml were collected and counted by liquid scintillation counting. Counts were converted to percentage of total counts recovered. *N* is a chromatogram of the lysine-labeled tryptic peptides from native coat protein. The nomenclature used is that of Konigsberg *et al.* (26). *+1* shows the lysine-labeled tryptic peptides from the coat protein derivatives made during asparagine starvation which have 1 lysine for asparagine substitution (14). *Peaks A* and *B* are unidentified new peptides which result from the substitution. *EBS* shows the profile of lysine-labeled peptides from the extra basic satellite of coat protein found in unstarved cells.

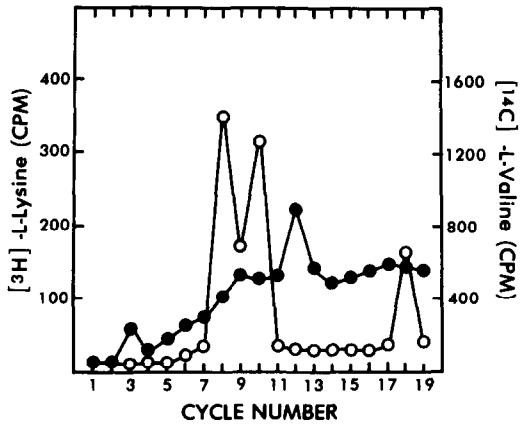

FIG. 4. **Edman degradation cycles of radioactive protein from the basic satellite spot.** Strain JK1 was infected with MS2, treated with rifampicin, labeled with [^{3}H]lysine and [^{14}C]valine from 20 to 40 min, and the basic satellite spot was purified on two-dimensional gels as described under "Experimental Procedures." Approximately 80,000 cpm of ^{3}H and 100,000 cpm of ^{14}C from the basic satellite spot and 100 nmol of cold coat protein were subjected to automated Edman degradation. ○, [^{14}C]valine in each cycle; ●, [^{3}H] lysine. Backgrounds are not subtracted.

Peaks A and *B* each contain at least two peptides. These peptides are resolved by reverse phase chromatography (results not shown); preliminary amino acid composition studies are consistent with their being the product of lysine for asparagine substitutions at AAU codons.[2] However, proof that this is so necessitates sequencing mistranslated protein.

Sequence of the NH_2-terminal Portion of the Basic Satellite Protein—The first 19 amino acids of native coat protein contain 2 asparagines, positions 3 and 12. These are encoded by an AAC and an AAU, respectively (29). Satellite protein labeled with [^{3}H]lysine and [^{14}C]valine was electroeluted from gels and subjected to automated, sequential Edman degradation. The results are shown in Fig. 4. As can be seen there is lysine incorporated at position 12 and, to a lesser extent, at position 3. The recovery of valine at positions 8, 10, and 18 allows one to correct the values for repetitive yield. With this correction and the subtraction of background, there is 7.2 times more lysine at position 12 (encoded by AAU) than at position 3 (encoded by AAC). Since each species of protein in the satellite spot can contain only 1 lysine for asparagine substitution and the relative amount of [^{14}C]lysine to [^{3}H] valine in the sample is known, one can calculate the fraction of molecules in the satellite spot which contain such errors. Approximately 7% of the molecules in the satellite contain a lysine at position 12. Since the satellite spot in this experiment was approximately 2% of the total coat protein, the substitution frequency at that particular codon is 1.5×10^{-3}. Similar calculations yield a substitution frequency of 1.8×10^{-4} for position 3.

[2] T. C. Johnston, unpublished results.

DISCUSSION

The coat protein of the small RNA bacteriophage MS2 shows charge heterogeneity when synthesized *in vivo*. In addition to the predominant "native" species there is a more basic species which can be resolved by isoelectric focusing. This basic satellite is synthesized from both the RNA genome (Fig. 1*A*) and message transcribed from a DNA copy (Fig. 2*A*). In both cases the relative proportion of the satellite can be greatly reduced by using a streptomycin-resistant host (Figs. 1*B* and 2*B*). This indicates that the satellite results from *in vivo* misreading of a codon or codons in the coat protein message. When [^{14}C]lysine-labeled satellite is digested with trypsin at least four new "non-native" peptides are produced (*peaks A* and *B*). The chromatographic fingerprint of the satellite peptides is virtually indistinguishable from that of a coat protein species with the same charge, +1, synthesized in asparagine-starved cells. The satellite then, like this +1 species, is almost certainly the result of a lysine for asparagine substitution. This result is confirmed by the sequencing of the NH_2 terminus of the satellite.

The amount of protein in this satellite is 1–3% of the total coat protein in *relA* strains (Figs. 1 and 2) and less than that in *relA*$^+$ (results not shown; see also Ref. 30). If misreading of the 10 asparagine codons in MS2 were random then one can calculate (31) that the mistranslation frequency per asparagine codon would be approximately 2×10^{-3} (for satellite 2% of total coat protein). This is almost an order of magnitude above frequencies calculated for valine for isoleucine substitutions (16, 17) and cysteine for arginine substitutions (18). Furthermore, if this were the average error frequency at all

asparagine codons in *E. coli* then most proteins would show one or two basic satellite spots when separated on isoelectric focusing gels. This is quite obviously not so for most *E. coli* proteins identified by the two-dimensional gel system of O'Farrell (see Refs. 13, 30, and 32). However the asparagine codon usage pattern in the coat protein message is very different from that in many *E. coli* messages.

Table II shows the codon usage patterns in the MS2 coat protein gene and strongly and weakly expressed *E. coli* chromosomal genes. Only the CAN and AAN codon families are shown. For the purposes of this table, strongly expressed genes are those whose protein product exists in more than 1000 copies/cell and weakly expressed genes are those where evidence exists that fewer than 100 copies of protein are found per cell. Interestingly, although the MS2 coat protein gene is very highly expressed, approximately 10^6 monomers/infected cell, the asparagine codon usage does not show the strong bias of such chromosomal genes. Because the protein products of the strongly expressed genes are abundant, many have been identified using the two-dimensional gel system (31). Very few, if any, of these abundant proteins have basic satellite spots. Since several of these proteins contain over 10 asparagine residues and a satellite spot containing 1% of the total protein would be observable, the average lysine for asparagine substitution frequency must be $<3 \times 10^{-4}$. For the AAC codon at amino acid 3 of coat protein we found a frequency of 1.8×10^{-4}. The only gene product from a weakly expressed gene which has been identified is the *araC* product (53). This protein, the arabinose repressor, does show a basic satellite spot. The *araC* gene has 14 asparagine residues; 10 are encoded by AAU. One can explain the charge heterogeneity of the arabinose repressor and that of coat protein, and also the lack of such heterogeneity in proteins from highly expressed chromosomal genes by hypothesizing that AAU codons are more error prone than the AAC codons. In coat protein, if all the heterogeneity was caused by misreading at the four AAU codons the error frequency at those codons would be 5×10^{-3}. A similar frequency would yield a basic satellite of the arabinose repressor containing 5% of the protein, approximately the amount observed. However the mistranslation frequency we measured for AAU was only approximately 2×10^{-3}. There are several possible explanations for this. One is that the low number of counts and the relatively high background led to an underestimate of the lysine at position 12. This explanation seems unlikely to account for the entire difference however. A more likely explanation is that there is a difference in the misreading frequency based on codon context and that some AAU codons have a misreading frequency above 5×10^{-3}. Proof of this will require measurements at all the asparagine codons.

We do not believe that this high mistranslation frequency is confined only to AAU codons. We originally reported that, during histidine starvation, elongation factor G is mistranslated at a higher frequency than elongation factor Tu (11). Although not yet completely sequenced, the gene for elongation factor G (9) clearly has a higher ratio of CAU to CAC than do the genes for elongation factor Tu (45, 46). Indeed elongation factor G often shows acidic charge heterogeneity on two-dimensional gels as if a histidine codon had been misread as a less basic amino acid (13), the same type of error as is found in histidine starvation (11, 12). In depth analysis has not been performed using histidine codons since our model protein, coat protein, contains no histidine residues and histidine is a relatively rare amino acid.

Since certain codons for a given amino acid may be mistranslated at a higher frequency than other codons, calculation of average mistranslation frequencies can be very misleading. It is not clear how many other codons may be misread at high frequencies or in which organisms. Although not universally true (54), many highly expressed genes in other organisms have a clear preference for AAC over AAU (2, 55). It is possible that this preference exists because of selection pressure to avoid a high frequency of errors. It is not clear why weakly expressed genes do not show the same preference. It has been argued that this is because of lack of selection pressure (56). However some codon preferences are nearly identical in strongly and weakly expressed *E. coli* chromosomal genes; this is true in the lysine codons AAA and AAG (see Table II). It is even conceivable that there is some biological advantage in having some proteins with a higher than average frequency of some specific errors. It will be necessary to identify other error-prone codons however before any meaningful analysis can be done on this problem.

Orgel originally proposed that error feedback caused by mistranslated proteins could lead to error catastrophe (57, 58), one model of aging. It seems clear from our work that experiments attempting to prove or disprove this hypothesis must be based on actual measurements of error frequencies in particular proteins. Further it suggests that it is possible that the target proteins for error catastrophe may not be abundant proteins.

TABLE II

Codon usage in the CAN and AAN families

Amino acid	Codon	No. in MS2 coat protein gene	No. in *E. coli* genes encoding	
			Abundant protein[a]	Low level protein[b]
Histidine	CAU	0	15	35
	CAC	0	81	24
Glutamine	CAA	1	39	56
	CAG	5	209	67
Asparagine	AAU	4	28	54
	AAC	6	201	45
Lysine	AAA	5	323	64
	AAG	1	103	24

[a] *lpp* (33), *ompA* (34), *pa* 4, *C, D, E, F, H* (see Refs. 35 and 36), *recA* (37), *rplA, J, K, L* (38), *rpoB* (39), *rpmB, G* (40), *rpsA* (41), *rpsJ, L* (9, 42), *rpsT* (43), *ssb* (44), *tuf A, B* (45, 46).

[b] *ampC* (47), *araC* (48), *dnaG* (49), *galR* (50), *lacI* (51), *lexA* (52), *trmD* (G. Björk, personal communication).

Acknowledgments—We thank Tin Cao for contributions to the protein sequence analysis. Jack Parker would like to thank C. Kurland and J. Gallant for stimulating discussions and advice. We thank Elizabeth Parker for typing the manuscript.

REFERENCES

1. Gray, M. W., and Doolittle, W. F. (1982) *Microbiol. Rev.* **46,** 1–42
2. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. (1981) *Nucleic Acids Res.* **9,** r43–r74
3. Modiano, G., Battistuzzi, G., and Motulsky, A. G. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 1110–1114
4. Nussinov, R. (1981) *J. Mol. Biol.* **149,** 125–131
5. Heindell, H. C., Liu, A., Paddock, G. V., Studnicka, G. M., and Salser, W. A. (1978) *Cell* **15,** 43–54
6. Grosjean, H., Sankoff, D., Min Jou, W., Fiers, W., and Cedergren, R. J. (1978) *J. Mol. Evol.* **12,** 113–119
7. Ikemura, T. (1981) *J. Mol. Biol.* **151,** 389–409
8. Grosjean, H., and Fiers, W. (1982) *Gene* **18,** 199–209
9. Post, L. E., and Nomura, M. (1980) *J. Biol. Chem.* **255,** 4660–4666
10. Chavancy, G., and Garel, J.-P. (1981) *Biochimie* **63,** 187–195
11. Parker, J., Pollard, J. W., Friesen, J. D., and Stanners, C. P. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 1091–1095

12. O'Farrell, P. H. (1978) *Cell* **14,** 545–557
13. Parker, J., and Friesen, J. D. (1980) *Mol. Gen. Genet.* **177,** 439–445
14. Parker, J., Johnston, T. C., and Borgia, P. T. (1980) *Mol. Gen. Genet.* **180,** 275–281
15. Parker, J. (1981) *J. Biol. Chem.* **256,** 9770–9773
16. Loftfield, R. B. (1963) *Biochem. J.* **89,** 82–92
17. Loftfield, R. B., and Vanderjagt, D. (1972) *Biochem. J.* **128,** 1353–1356
18. Edelmann, P., and Gallant, J. (1977) *Cell* **10,** 131–137
19. Remaut, E., De Waele, P., Marmenout, A., Stanssens, P., and Fiers, W. (1982) *Eur. Mol. Biol.* **1,** 205–209
20. Bernard, H.-U., Remaut, E., Hershfield, M. V., Das, H. K., Helinski, D. R., Yanofsky, C., and Franklin, N. (1979) *Gene* **5,** 59–76
21. Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974) *J. Bacteriol.* **119,** 736–747
22. Lennox, E. S. (1955) *Virology* **1,** 190–201
23. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250,** 2007–4021
24. O'Farrell, P. Z., Goodman, H. M., and O'Farrell, P. H. (1977) *Cell* **12,** 1133–1142
25. Sammons, D. W., Adams, L. D., and Nishizawa, E. E. (1981) *Electrophoresis* **2,** 135–141
26. Konigsberg, W., Weber, K., Notani, G., and Zinder, N. (1966) *J. Biol. Chem.* **241,** 2570–2588
27. Gorini, L. (1974) in *Ribosomes* (Nomura, M., Tissières, A., and Lengyel, P., eds) pp. 791–803, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
28. Ellis, N., and Gallant, J. (1982) *Mol. Gen. Genet.* **188,** 169–172
29. Min Jou, W., Haegeman, G., Ysebaert, M., and Fiers, W. (1972) *Nature (Lond.)* **237,** 82–88
30. Gallant, J. A. (1979) *Annu. Rev. Genet.* **13,** 393–415
31. Harley, C. B., Pollard, J. W., Chamberlain, J. W., Stanners, C. P., and Goldstein, S. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 1885–1889
32. Bloch, P. L., Phillips, T. A., and Neidhardt, F. C. (1980) *J. Bacteriol.* **141,** 1409–1420
33. Nakamura, K., Pirtle, R. M., Pirtle, I. L., Takeishi, K., and Inouye, M. (1980) *J. Biol. Chem.* **255,** 210–216
34. Beck, E., and Bremer, E. (1980) *Nucleic Acids Res.* **8,** 3011–3027
35. Kanazawa, H., Kayano, T., Mabuchi, K., and Futai, M. (1981) *Biochem. Biophys. Res. Commun.* **103,** 604–612
36. Kanazawa, H., Mabuchi, K., Kayano, T., Noumi, T., Sekiya, T., and Futai, M. (1981) *Biochem. Biophys. Res. Commun.* **103,** 613–620
37. Horii, T., Ogawa, T., and Ogawa, H. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 313–317
38. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H., and Dennis, P. P. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76,** 1697–1701
39. Ovchinnikov, Yu. A., Monastyrskaya, G. S., Gubanov, V. V., Gur'ev, S. O., Chertov, Yu., Modyanov, N. N., Grinkevich, V. A., Makarova, I. A., Marchenko, T. V., Polovnikova, I. N., Lipkin, V. M., and Sverdlov, E. D. (1980) *Dokl. Akad. Nauk SSSR* **253,** 994–998
40. Lee, J. S., An, G., Friesen, J. D., and Isono, K. (1981) *Mol. Gen. Genet.* **184,** 218–223
41. Schnier, J., Kimura, M., Foulaki, K., Subramanian, A.-R., Isono, K., and Wittmann-Liebold, B. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 1008–1011
42. Olins, P. O., and Nomura, M. (1981) *Cell* **26,** 205–211
43. Mackie, G. A. (1981) *J. Biol. Chem.* **256,** 8177–8182
44. Sancar, A., Williams, K. R., Chase, J. W., and Rupp, W. D. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4274–4278
45. Yokota, T., Sugisaki, H., Takanami, M., and Kaziro, Y. (1980) *Gene* **12,** 25–31
46. An, G., and Friesen, J. D. (1980) *Gene* **12,** 33–39
47. Jaurin, B., and Grundström, T. (1981) *Proc. Natl. Acad. Sci. U. S. A.* **78,** 4897–4901
48. Wallace, R. G., Lee, N., and Fowler, A. V. (1980) *Gene* **12,** 179–190
49. Smiley, B. L., Lupski, J. R., Svec, P. S., McMacken, R., and Godson, G. N. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 4550–4554
50. von Wilcken-Bergmann, B., and Müller-Hill, B. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 2427–2431
51. Farabaugh, P. (1978) *Nature (Lond.)* **274,** 765–769
52. Horii, T., Ogawa, T., and Ogawa, H. (1981) *Cell* **23,** 689–697
53. Kolodrubetz, D., and Schleif, R. (1981) *J. Mol. Biol.* **149,** 133–139
54. Dugaiczyk, A., Law, S. W., and Dennison, O. E. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79,** 71–75
55. Holland, M. J., Holland, J. P., and McAllister, L. (1982) in *Genetic Engineering of Microorganisms for Chemicals* (Hollaender, A., ed) pp. 291–303, Plenum Press, New York
56. Nichols, B. P., van Cleemput, M., and Yanofsky, C. (1981) *J. Mol. Biol.* **146,** 45–54
57. Orgel, L. E. (1963) *Proc. Natl. Acad. Sci. U. S. A.* **49,** 517–521
58. Orgel, L. E. (1970) *Proc. Natl. Acad. Sci. U. S. A.* **67,** 1476