Published in final edited form as:

Nature. 1982 September 23; 299(5881): 365–367.

Random components in mutagenesis

P. L. Foster^{*}, E. Eisenstadt^{†,‡}, and J. Cairns[†]

^{*} Interdisciplinary Programs in Health, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

[†] Department of Microbiology, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

[‡] Laboratory of Toxicology, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

Abstract

The mutability of DNA varies enormously from one base pair to another. Part of this variation is due to the specificity of the reaction between mutagen and base, but much of the variation is due to unknown causes. A genetic system developed by Miller and colleagues¹ allows the mutation frequencies of a large number of different base pairs in the *lacI* gene of *Escherichia coli* to be compared. For example, Coulondre and Miller² found that the sites most readily mutated by UV light are almost 100 times more often mutated than the least susceptible sites. A recently completed study³ of mutagenesis with neocarzinostatin (NCS) in the *lacI* gene has prompted us to re-examine some previous studies^{2,4,5} of mutagenesis in this gene. Our analysis, reported here, suggests that the mutations induced by certain mutagens fall into two classes: mutations in one class are clearly distributed non-randomly, that is, they are very common at some sites and significantly less common at others; mutations in the second class, however, occur at low frequency and appear to be randomly distributed. Both classes of mutations seem to occur only at damaged bases.

One of the largest collections of fully characterized point mutations is a set of about 600 nonsense mutations generated in the *lacI* gene by UV light². There are 64 sites in this gene that can yield amber (TAG) or ochre (TAA) codons by single base changes. Overall, the distribution of UV-induced mutations among these sites is obviously non-random; 3 of the 64 possible nonsense codons were each represented more than 50 times, 3 did not appear in the collection at all and many occurred only once or twice. Because UV-mutagenesis clearly involves non-random elements, it has often been assumed that the entire process is non-random and that each base pair occupies some unique position in a continuum of mutability. However, if we disregard both the base substitutions being monitored and the distribution of events over the genetic map, and instead simply look at the distribution of frequencies, it becomes apparent that the rarer mutations form a Poisson distribution (that is, their frequencies appear to be randomly distributed) with a mean of 2.7 occurrences per site (Fig. 1a). Of the 64 sites, 41 fall into this distribution. (We have included in this set the 3 codons in which no nonsense mutations were detected, because a mean occupancy of 2.7 would be expected to leave 3 out of 41 sites unoccupied.)

Solely on the basis of the frequencies of the different mutations, we therefore conclude that UV-induced mutations arise from two classes of events: apparently random, low-frequency occurrences (LFOs) that account for about one-third of the mutations, and non-random, high-frequency occurrences (HFOs) that account for the rest. This conclusion is greatly strengthened when we consider the base sequences at the 64 sites. As pointed out by Coulondre *et al.*⁵, all of the dozen or so ‘hotspots’ for UV-induced G · C to A · T transitions

found in the *lacI* gene are at sites of adjacent pyrimidines. Based on our analysis, we find that all 23 of the sites for HFOs are base pairs at which the pyrimidine has a pyrimidine next to it; in contrast, 15 of the 41 LFO sites involve base pairs where the pyrimidine has purines on both sides of it. This difference between the sites for HFOs and LFOs is highly significant ($\chi^2 = 9.4$, $P < 0.005$). Thus, it seems likely that UV-induced HFOs require interaction between adjacent pyrimidines (for example, they may be due to pyrimidine dimers⁶ or some other lesion involving a pyrimidine pair⁷) whereas LFOs do not. As the UV-induced LFOs occur at every monitorable site in the *lacI* gene and can generate each of the five base pair changes that yield nonsense codons (G · C to A · T, T · A or C · G; A · T to T · A or C · G), it seems likely that UV light is producing, at random, all possible base substitutions at every base pair.

For a base pair to be a site for UV-induced HFOs, it clearly is not enough that its pyrimidine be adjacent to another pyrimidine. Of the 49 base pairs in the *lacI* gene that involve adjacent pyrimidines and are capable of generating nonsense codons, only 23 were sites for HFOs. For example, of the 14 Py-C-A-G and Py-C-A-A sequences in the *lacI* gene where transition of the C to a T yields a nonsense codon, 10 are sites for HFOs and cover a 10-fold range of mutability; the remaining 4 are sites for LFOs only. So some factor other than neighbouring base sequence must be important in determining the frequency of mutation at these sites. As the loss of the *uvrB*-dependent excision repair pathway has no effect on the relative UV mutability of most sites⁴, accessibility to this repair pathway cannot account for most of the variation in mutation frequency. We conclude, therefore, that the mutability of different pairs of adjacent pyrimidines is probably determined by the effect that some factor, such as DNA secondary structure, has on the frequency of formation of premutational lesions.

Mutagenesis by UV light requires induction of the ‘SOS response’⁸. Because LFOs seem to be random and indiscriminate, it was conceivable that they represent untargeted mutagenesis due to error-prone DNA repair or replication. We therefore extended our analysis to the published results for 4-nitroquinoline-*N*-oxide (NQNO), which is also an SOS-dependent mutagen⁹. Coulondre and Miller² have shown that NQNO is highly specific for G · C base pairs, with a strong preference for producing G · C to A · T transitions. We see here that NQNO, like UV light, produces a non-random distribution of HFOs plus an apparently random distribution of LFOs (Fig. 1*b*). However, in contrast to the UV results, the mutations that are produced by the NQNO-induced HFOs and LFOs are distinguishable—all the G · C transitions arise from HFOs, whereas nearly all the G · C transversions arise from LFOs (any G · C transitions produced by LFOs cannot be detected because the sites for transitions are dominated by HFOs). Since untargeted mutagenesis would be expected to include base substitutions of A · T sites, the failure of NQNO to mutate A · T base pairs implies that both HFOs and LFOs result from targeted events. We can conclude, therefore, that the low level of apparently random mutagenesis following treatment with mutagens such as NQNO or UN light cannot be accounted for by untargeted mutagenesis.

Another SOS-dependent mutagen that we have examined is NCS (Fig. 1*c*). Like UV light and NQNO, it produces both HFOs and LFOs; the LFOs include examples of each of the five kinds of base changes that can be detected as nonsense mutations. However, NCS mutates only 37 of the 64 sites in the *lacI* gene; the sites it fails to mutate include examples of all monitorable base changes³. As in the case of NQNO, the failure of NCS to mutate nearly half the monitorable sites in the *lacI* gene would be difficult to explain if untargeted mutagenesis were making a major contribution to SOS-dependent mutagenesis.

Although we have looked at only a few mutagens, we now realize that the available data are less informative than they seem at first sight. For example, because of the nature of the genetic code, an analysis of nonsense mutations cannot distinguish which of the four

pyrimidine pairs (T-T, T-C, C-T, or C-C) is the best target for UV mutagenesis; any effect attributable to the pyrimidine pair is confounded by the constraints on both the event being monitored (which can be transitions or transversions at cytosines, but only transversions at thymines) and the reading frame (for example, substitutions for thymine cannot be monitored at the first position of a codon). To separate these factors would require an analysis of missense mutations, which are not subject to the same base sequence restrictions. But there is yet another problem. In any mutant collection, the rarer events (LFOs) will be present in rather small numbers and, because of this, their frequency distribution will be dominated by the variance of the Poisson distribution. Thus, although we have said that LFOs appear to be randomly distributed, this suggestion needs to be confirmed by a much larger collection of LFOs.

We can, however, draw some important conclusions about LFOs. First, like HFOs, they are likely to be targeted, that is, to occur only at damaged bases. Second, they can account for a significant proportion of the mutations arising after exposure to a mutagen (for example, one-third of the UV-induced mutations). Last, they seem to be able to induce all base changes—the LFO lesion seems to be non-informational and to allow the insertion of a base at random. These conclusions suggest that, unlike HFO lesions, which are clearly specific to each mutagen, the LFO lesion may be the same for a number of mutagens; one example of such a common lesion would be an apurinic or apyrimidinic site.

P.L.F. was a fellow of the Interdisciplinary Programs in Health and was supported by grant CR807809-01-1 from the US Environmental Protection Agency. E.E. was supported in part by US NIH grant CA26135.

References

1. Miller, JH. The Operon. Miller, JH.; Reznikoff, WS., editors. Cold Spring Harbor Laboratory; New York: 1978. p. 31-88.
2. Coulondre C, Miller JH. *J molec Biol* 1977;117:557–606.
3. Foster PL, Eisenstadt E. *J Bact.* (submitted).
4. Todd PA, Glickman BW. *Proc natn Acad Sci USA* 1982;79:4123–4127.
5. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. *Nature* 1978;274:775–780. [PubMed: 355893]
6. Fisher, GJ.; Johns, HE. *Photochemistry and Photobiology of Nucleic Acids*. Wang, S-Y., editor. Vol. 2. Academic; New York: 1976. p. 225-295.
7. Lippke JA, Gordon LK, Brash DE, Haseltine WA. *Proc natn Acad Sci USA* 1981;78:3388–3392.
8. Witkin EM. *Bact Rev* 1976;40:869–907. [PubMed: 795416]
9. Kondo S, Ichikawa H, Iwa K, Kato T. *Genetics* 1970;66:187–217. [PubMed: 4934196]

Fig. 1.

The frequency distributions of approximately 600 UV-induced² (a), 1,000 NQNO-induced² (b) and 200 NCS-induced³ (c) amber and ochre mutations in the *lacI* gene of *E. coli*. The histograms show the number of sites at which k mutations were found. The inset histograms show the lower end of the distributions for individual values of k from 0 to 9. These have been matched to Poisson distributions, whose means were estimated using the relationship:

$$m = (k+1) \frac{N_{k+1}}{N_k}$$

where N_k and N_{k+1} are the number of sites at which occurred k and $k+1$ mutations, respectively. The observed values of N_k for several consecutive classes were used to calculate m . The total number of sites (N_T) expected to fall into the Poisson distribution was calculated from the observed values of N_k using the relationship:

$$N_T = \frac{\sum_{k=1}^5 N_k}{e^{-m} \sum_{k=1}^5 m^k / k!}$$

The total number of mutations (T) in the Poisson distribution was then calculated as:

$$T = mN_T$$

a, UV-induced mutations at all sites are shown; sites where the pyrimidine of the target base pair is not adjacent to another pyrimidine are shaded. The Poisson distribution includes 111 mutations at 41 sites ($N_k \leq 6$); $m = 2.7$ was estimated from N_0 to N_2 . *b*, NQNO-induced mutations at G-C sites only are shown; sites where G-C transversions occur are stippled. The Poisson distribution included 63 mutations at 18 sites ($N_k < 8$); $m = 3.3$ was estimated from N_1 and N_2 . *c*, NCS-induced mutations at all sites are shown. The Poisson distribution included 72 mutations at 29 sites ($N_k \leq 6$); $m = 2.5$ was estimated from N_1 and N_2 .