Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 503–507, January 1984
Genetics

Correction of complex heteroduplexes made of mouse *H-2* gene sequences in *Escherichia coli* K-12

(repair/gene conversion/genetic diversity/H-2 antigens)

BRIGITTE CAMI, PHILIPPE CHAMBON, AND PHILIPPE KOURILSKY

Unité de Biologie Moléculaire du Gène, Equipe de Recherche no. 201 du Centre National de la Recherche Scientifique et Service Commun no. 20 de l'Institut National de la Santé et de la Recherche Médicale, Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris Cédex 15, France

Communicated by Sydney Brenner, September 19, 1983

ABSTRACT We have prepared heteroduplexes between two plasmids that carry, in the same orientation, two H-2 cDNA inserts, 1.15 and 1.0 kilobase long, respectively. Their sequences encode two distinct class I transplantation antigens of the mouse and differ by 8% of their nucleotides. Molecules with a rearranged array of restriction sites were found after transformation and cloning in an *Escherichia coli recA*$^-$ host. Nucleotide sequences showed that the rearranged molecules derived their nucleotides from the two parental strands. Thus, correction of these complex heteroduplexes takes place in *E. coli* and probably involves repair mechanisms. It provides the basis for a mutational process in which several nucleotides (amino acids) can be altered in a single event. It also offers a practical means of making genetic variants. Several other implications are discussed.

Heteroduplexes can form *in vivo* by DNA strand exchange between partially homologous, but not identical, sequences (reviewed in ref. 1). They can also result from replication mistakes. In *Escherichia coli*, the newly synthesized strand, being transiently undermethylated, is preferentially corrected (reviewed in refs. 2 and 3). *E. coli dam*$^-$ mutants, deficient in a major methylation activity, display high mutation rates, as expected from random correction of either strand (4).

Heteroduplexes can be prepared *in vitro*, transformed into living cells, and their *in vivo* correction can then be studied. Such analyses have been carried out mostly in *E. coli* (5–8; see ref. 1 for review). With heteroduplexes of λ phage DNA carrying up to four nucleotide mismatches, Wagner and Meselson (9) observed independent correction as well as cocorrection of the marker mutations. Heteroduplexes of simian virus 40 (10, 11) and polyoma (12) mutants have been transfected into mammalian cells, where mismatch repair is also believed to take place.

These studies have been carried out with "simple" heteroduplexes, carrying one or a few nucleotide mismatches. Little is known about the correction of more complex structures, involving many noncomplementary nucleotides, which we, and others, suspect to be a mechanism capable of generating considerable diversity (3, 13, 14; see below). Because this idea may explain some of the genetic polymorphism in eukaryotic multigene families (14), particularly in the mouse *H-2* genes studied in our laboratory, we have undertaken an analysis of the fate of complex heteroduplexes. As a first step, these studies have been carried out in *E. coli*.

In the *H-2* multigene family, which encodes the polymorphic class I transplantation antigens, proteins and genes analyzed so far display high homology, with 80–95% of identical residues between any two aligned sequences (see refs. 15–18 for review). We have selected for study two blocks of *H-2* sequences, about 1 kilobase (kb) long, which differ in many positions, prepared heteroduplexes *in vitro*, and transformed them into *E. coli*. We report here that correction takes place, and we discuss several implications.

MATERIAL AND METHODS

Bacterial Strains. The *recA*$^+$ and *recA*$^-$ *E. coli* strains used here were 803 *supE supF* $r^{k-}m^{k-}$ and 803 *supE supF* $r^{k-}m^{k-}$ *recA*$^-$ (19). The *recA*$^-$ strain has been periodically tested in this laboratory for UV sensitivity, formation of small colonies, and inability to support the growth of certain λ mutants. The *dam*$^-$ strain was gM82 *dam*$^-$ (4). Strains harboring pH-2^d-1 and pH-2^d-3 have been described (20, 21).

Enzymes and Isotopes. Restriction enzymes were purchased from New England BioLabs and Bethesda Research Laboratories and were used in the conditions recommended by the manufacturers. Polynucleotide kinase was from Boehringer Mannheim and terminal deoxynucleotidyl transferase from P-L Biochemicals. [γ-^{32}P]ATP and α-^{32}P-labeled cordycepin (specific activity, 3,000 Ci/mmol; 1 Ci = 37 GBq) were purchased from Amersham.

Formation and Transformation of Heteroduplexes. One plasmid (several micrograms) was digested by *Eco*RI and *Hin*dIII; the other was cut by *Bam*HI and *Sph* I. The plasmids were then extracted once by chloroform/isoamyl alcohol, precipitated by ethanol, and resuspended in 10 mM Tris·HCl, pH 7.5/1 mM EDTA, at a concentration of 250 μg/ml; 500 ng of each plasmid was mixed in a final volume of 10 μl of the same buffer. The mixture was denatured by boiling for 3 min in water. Annealing was for 4 hr at 63°C (22). The sample was then diluted 1:10 in 0.1 M Tris·HCl (pH 7.1) and aliquots containing 10–50 ng of DNA were transformed into *E. coli* (23).

DNA Sequence Analysis. Nucleotide sequences were determined as described by Maxam and Gilbert (24) using DNA fragments labeled by [γ-^{32}P]ATP and polynucleotide kinase, or ^{32}P-labeled cordycepin and terminal deoxynucleotidyl transferase (25).

RESULTS

Choice of Sequences. In the mouse, most somatic cells display at their surface three types of class I molecules, coded by distinct loci (*H-2D*, *K*, and *L*) of chromosome 17 (reviewed in refs. 15–18). We have isolated previously two cDNA clones, pH-2^d-1 and pH-2^d-3, that probably encode the *H-2D* and *L* products, respectively, in the *d* haplotype (20, 21). These cDNAs were cloned in the bacterial plasmid pBR322. The inserts are 1.15 and 1 kb long and represent incomplete copies of the 1,800-nucleotide long H-2 mRNAs, starting from poly(A) in 3′. They all encompass the third extracellular domain, the membrane spanning region, and the cytoplasmic COOH terminus of H-2 heavy chain, as well

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); bp, base pair(s).

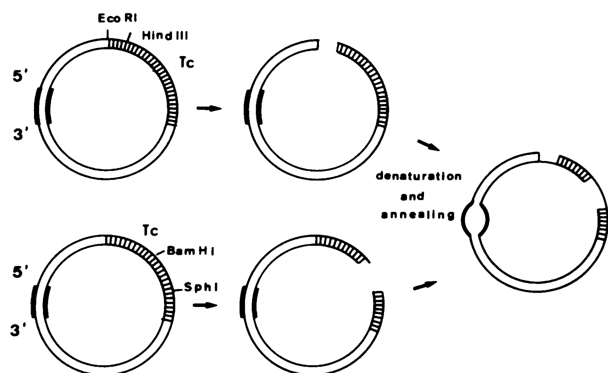

FIG. 1. Preparation of heteroduplexes. The figure depicts the formation of heteroduplexes with gaps in the *Tc* gene. Other molecules are formed in the annealing reaction, particularly homoduplexes of truncated plasmids and concatenates of heteroduplex molecules. Examination of the annealed mixture by electrophoresis indicated that, in our experimental conditions, the latter were much less abundant (≈20% or less) than circular heteroduplexes.

as 480 base pairs (bp) of noncoding sequence downstream from the stop codon. We have selected for study pH-2^d-1 and pH-2^d-3 in which the inserts have the same orientation with regard to pBR322. Their nucleotide sequences can readily be aligned. They differ in 86 positions including a 3-bp deletion in pH-2^d-1 and a 9-bp deletion in pH-2^d-3. In addition, the pH-2^d-1 insert extends 142 bp further at the 5′ end. It also carries a longer poly(A) tract in the 3′ end (40 residues versus 30 in pH-2^d-3). The lengths of the G·C homopolymeric tails are roughly similar but have not been precisely determined.

Preparation and Transformation of Heteroduplexes. pH-2^d-1 and pH-2^d-3 were digested to completion with two sets of restriction enzymes inactivating the tetracycline resistance (Tc^R) gene. Neither molecule alone could, in principle, confer Tc^R upon transformation, but heteroduplexes could, provided that the two single strand gaps are repaired (Fig. 1).

In control experiments, pBR322 was cut by one or the other pair of enzymes (*Eco*RI/*Hin*dIII or *Bam*HI/*Sph* I). When digested molecules of only one type were denatured and reannealed, none or few Tc^R transformants were obtained. When heteroduplexes of pBR322 digested by one and the other set of enzymes were made, about 1–2 × 10^5 Tc^R transformants per μg of DNA were obtained—i.e., 1/10th to 1/20th the number obtained with undigested pBR322 in a $recA^-$ or the isogenic $recA^+$ host (Table 1).

Heteroduplexes of pH-2^d-1 and pH-2^d-3 were prepared. The transformation efficiency was further decreased (Table 1). However, the number of Tc^R transformants was much

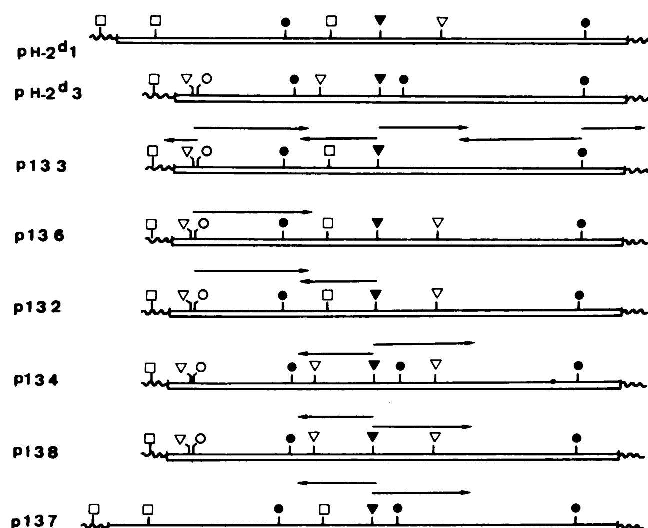

FIG. 2. Restriction maps of rearranged clones. Plasmid DNA, digested by one or several restriction enzymes, was subjected to electrophoresis in a 1% agarose gel. The inserts are shown as bars and plasmid DNA as a wavy line. The unique *Sac* I site present in all plasmids is shown as ▼. Other enzymes used are indicated as follows: *Bgl* II (○), *Hin*fI (▽), *Hpa* II (□), *Rsa* I (●). The regions in which sequences were determined are indicated by arrows, the origins of which correspond to the restriction sites used for terminal labeling.

higher than (10-fold or more) the backgrounds obtained with self-annealed pH-2^d-1 or pH-2^d-3, or a mixture of nondenatured, or separately self-annealed plasmids (see legend to Table 1).

Restriction Analysis of Tc^R Transformants. Sixty $recA^+$ and 48 $recA^-$ Tc^R transformants were reisolated and plasmid DNA was characterized by restriction mapping using *Bgl* II, *Hin*fI, *Hpa* II, and *Rsa* I, which readily discriminate between the parental molecules (Fig. 2). By this test, the 60 Tc^R $recA^+$ transformants distributed about equally between the two parental types, and no other kind of molecule was found. In contrast, 5 of the 48 plasmids isolated in the $recA^-$ host displayed a novel combination of restriction sites; the remaining 43 clones were of the two parental types (Table 1).

To examine the possible involvement of *dam* methylation (2, 3, 7, 8) we introduced pH-2^d-1 and pH-2^d-3 into a dam^- host and prepared DNA. The extent of methylation of the G-A-T-C sites was monitored with *Mbo* I, *Sau*3A, and *Dpn* I, which recognize the G-A-T-C sequence in different methylation contexts (26, 27). These controls (not shown) indicated that the sites were essentially all methylated or unmethy-

Table 1. Analysis of Tc^R clones obtained on transformation by heteroduplex DNA

Parental molecules	*dam* methylation	*E. coli* host	No. of Tc^R transformants*	No. of clones analyzed	Clone type† pH-2^d-1	pH-2^d-3	Rearranged	Clones studied
pBR322/pBR322	+/+	$recA^+$	1.7 × 10^5					
	+/+	$recA^-$	1 × 10^5					
pH-2^d-1/pH-2^d-3	+/+	$recA^+$	7.4 × 10^3	60	30	29	0	
	+/+	$recA^-$	3.8 × 10^3	48	19	24	5	p132, p133 p137, p138
	−/+	$recA^-$	5 × 10^3	24	4	18	2	p134
	+/−	$recA^-$	8 × 10^3	24	16	7	1	
	−/−	$recA^-$	1 × 10^3	20	7	12	1	p136
pH-2^d-3/p133	+/+	$recA^-$	8 × 10^3					

*Per microgram of DNA; not normalized with respect to the transformation efficiencies measured with intact pBR322. The latter were about 4 × 10^6 transformants per μg of DNA in the $recA^+$ host, and 1.5 × 10^6 in $recA^-$, but varied 2- to 3-fold in different experiments. Backgrounds (often 0 and always less than 10^3) have not been subtracted.

†Determined by restriction mapping only.

```
                    Third domain                                       T.M.                        C

         pH-2d-1   C-C-G-C-G-T-A-T-G-A-G-A-     -GGCAAGGAG-T-A-CA-G  A-A-CA-A-A-C-T-C-G-T-C-A  G-T-C-T-
         pH-2d-3   T-A-C-T-A-A-C-A-T-G-T-C-CTG-          -C-G-TG-C  T-T-TG-G-G-T-G-T-A-C-A-G  A-A-G-C-
CODING   p133      T-A-C-T-A-A-C-T-G-A-G-A-     -GGCAAGGAG-T-A-CA-G  A-A-CA-A-A-C-T-C-G-T-C-A  G-T-C-T-
         p136        -C-T-A-A-C-T-G-A-G-A
         p132        -C-T-A-T-A-T-G-A-G-A-     -GGCAAGGAG-T-A-CA-G  A-A-CA-A-A-C-T-C-G-T-C-A  G-T-C-T-
         p134                        -CTG-          -C-G-TG-C  T-T-TG-G-G-T-G-T-A-C-A-G  A-A-G-C-
         p138                        -CTG-          -C-G-TG-C  T-T-TG-G-G-T-G-T-A-C-A-G  A-A-G-C-
         p137                                  -GGCAAGGAG-T-A-CA-G  A-A-CA-A-A-C-T-C-G-T-C-G  A-A-G-C-


           pH-2d-1   A-TG-G-TG-A-C-CACA-A-GT-GT    TC-GC-C-AA-T-AC-C-C-T-T-G-C-C-T-A-A-G-A-
           pH-2d-3   G-AC-T-GA-G-G-GGTC-C-CA-CATT-CT-   -T-GT-  -TG-T-T-C-G-T-A-G-G-G-  -T-
NON CODING p133      A-TG-G-TG-A-C-CACA-A-GT-GT   -CT-   -T-GT-  -TG-T-T-C-G-T-A-G-G-G-  -T-
           p134      G-AC-T-GA-G-G-GGTC-C-GT-GT   -TC-GC-C-AA-T-TG-
           p138      G-TG-G-TG-A-C-CACA-A-GT-GT   -TC-GC-C-AA-T-AC-
           p137      G-AC-T-GA-G-G-GGTC-C-CA-CATT-CT-   -T-GT-  -TG-
```

FIG. 3. Nucleotide sequences of rearranged regions. The sequences are arranged with regard to the published sequences of pH-2^d-1 and pH-2^d-3 (20, 21), the latter being corrected for three printing mistakes. For simplicity, the only nucleotides shown are those different in the two plasmids. They are separated by bars, indicating one or several identical nucleotides, or a blank, indicating a deletion with regard to the other aligned sequence. Regions coding for the third extracellular domain, transmembrane (TM), and cytoplasmic (C) parts of the molecule are on top, and the noncoding ones on the bottom. Sequence strategies are as described in Fig. 2.

lated on both strands in plasmids grown in the *dam*$^+$ or *dam*$^-$ host, respectively. Heteroduplexes were then prepared with one methylated and one unmethylated parent, and transformed into *E. coli recA*$^-$. Clones were analyzed as described above. Most were of the methylated parental type (Table 1), indicating that *dam* methylation plays a role. When both parents were unmethylated, fewer transformants were obtained. Several rearranged plasmids emerged from these experiments, two of which were further analyzed (Table 1).

The restriction maps of six plasmids with a rearranged *H-2* sequence are shown in Fig. 2.

Partial Nucleotide Sequence of Rearranged Clones. For these six clones, nucleotide sequences were determined in the region of heterologous restriction sites. In a search for variations undetected by restriction mapping, the entire sequence of one clone (p133) was determined. Data are shown in Fig. 3. Sequences of rearranged clones match that of the two parents without involving any new nucleotide.

DISCUSSION

We have prepared *in vitro* complex heteroduplexes from two sequences differing by more than 8% of their nucleotides and shown that, on transformation and cloning in *E. coli*, rearranged sequences can be obtained. This observation can be accounted for either by two recombination events (between truncated plasmid molecules present in the transformation mixture, or between plasmids generated by replicational segregation of the heteroduplexes) or by heteroduplex repair. In other systems so far studied, the efficiency of repair prevailed over recombination (5, 9). Furthermore, the rearrangements observed here occur apparently at random in a *bona fide recA*$^-$ host. We, therefore, favor heteroduplex repair as the most simple and likely explanation.

Results in Table 1 indicate a bias in favor of the methylated parental sequence when the other one is undermethylated. This could mean that segregants of the parental types are generated through *dam*-directed repair, but it might also reflect increased sensitivity of unmethylated strand in the heteroduplex to nucleolytic action. The latter hypothesis may be more likely, because Pukkila *et al.* (8) have shown that fully methylated heteroduplexes of λ DNA are poorly repaired. The rescue of complex heteroduplexes as parental or rearranged sequences may or may not use presently known repair mechanisms. A variety of *E. coli* mutants deficient in replication, recombination, and repair activities will have to be studied to clarify this question.

In 60 clones isolated on transformation of *recA*$^+$ by heteroduplex DNA, no plasmid displayed a rearranged array of restriction sites. However, further experiments (unpublished observations) show that rearranged clones can be found, but they are 2 to 4 times rarer than in *recA*$^-$. Tc^R transformants are 2 to 3 times more abundant, however. Conceivably, part of the Tc^R transformants obtained in *recA*$^+$ arise by recombination between overlapping truncated plasmids, increasing the background of nonrearranged clones. The figure of ≈10% rearranged clones isolated in *recA*$^-$ (5 out of 48) may be an underestimate, because they were identified by restriction mapping, which may leave alterations undetected.

The summary of our present analysis of six rearranged clones, together with the presumed structure of the initial heteroduplex, is shown in Fig. 4. The structure of the six clones can be interpreted as resulting from a single correction (repair) event in a region either internal to the H-2 cDNA sequences (p133, p134) or overlapping an unknown length of adjacent pBR322 sequence (p132, p136, p137, p138).

The 142 bp present in pH-2^d-1 and absent in pH-2^d-3 must create a large loop in the heteroduplexes. Its correction does not appear to be severely limiting in the production of viable transformants because heteroduplexes that lack it (made of pH-2^d-3 and p133) do not yield many more Tc^R transformants (Table 1). The sequence corresponding to the loop is, however, absent in five of the six clones and may, therefore, be preferentially eliminated. This may not hold for smaller loops because the 9- and 3-nucleotide insertions of pH-2^d-1 and pH-2^d-3 are retained in four and two clones, respectively, out of six.

Nucleotide sequences fitting one or the other parent are shown in Fig. 4. The overlaps correspond to sequences identical in pH-2^d-1 and pH-2^d-3. The lengths of the corrected

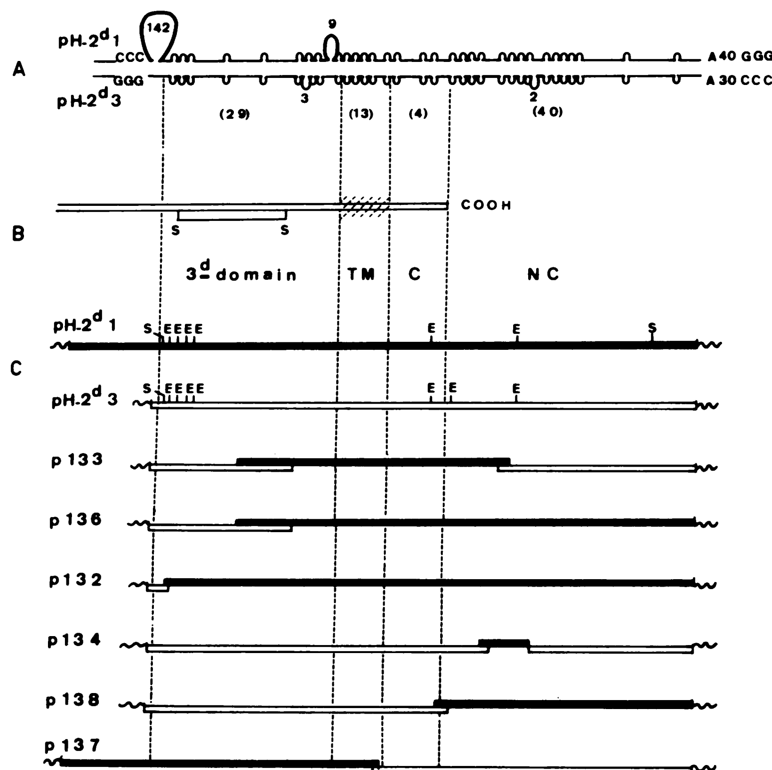

FIG. 4. Structure of rearranged clones. (*A* and *B*) Presumed structure of the starting heteroduplex, showing the 142-nucleotide loop in the 5′ region of pH-2^d-1; three smaller loops of 3, 9, and 2 nucleotides; and a number of nucleotide mismatches depicted as small bubbles. Numbers in parentheses indicate the total number of different nucleotides within various regions of the insert illustrated by the COOH-terminal moiety of a H-2 heavy chain (*B*). As in Fig. 3, the third domain, transmembrane (TM), and cytoplasmic (C) coding regions as well as the noncoding region (NC) are indicated. (*C*) The pH-2^d-1 and pH-2^d-3 inserts are shown as filled and open bars, and the *Sau*3A (G-A-T-C) and *Eco*RII (CCA/TGG) sites are indicated as S and E, respectively. The rearranged clones are depicted as filled and open bars as explained in the text.

regions with borders in the overlaps are thus somewhat ambiguous but vary in the approximate range of 75 bp (in p134) to 400 bp (in p133) to 0.5 kb, 0.9 kb, or more in the others. Earlier estimates of the average repair tracts with λ heteroduplexes were in the 2- to 3-kb range (5, 9).

The G-A-T-C and CCA/TGG sequences that undergo methylation in *E. coli* (28) are indicated in Fig. 4. We have found so far no obvious correlation between their location and that of the corrected areas, nor have we identified any evident bias in the choice of substituted bases. Finally, all sequences determined so far (a total of ≈3 kb) fit exactly one or the other parental sequence. In this sense, the correction process, apart from shuffling sequences, does not appear to be grossly mutagenic.

Correction of complex heteroduplexes may be used as a practical means of engineering genetic variants. One of its interesting characteristics is that all features of the primary structure common to both parents are conserved in the variants. Thus, plasmids p132, p133, and p136, which display rearrangements in the coding region, keep the appropriate reading frame and represent mutants of the COOH-terminal half of H-2 molecules. They, of course, retain all usual traits of heavy chains (17) (glycosylation and phosphorylation sites, cysteins in the appropriate position to make a disulfide bridge, etc.).

Many sequences, particularly in the higher eukaryotes, are only partially homologous and differ in many nucleotides. In spite of this, it has often been postulated (3, 16, 29–33) that they can undergo crossing-over and gene conversion on the basis of their (partial) homology. As was emphasized earlier (14), if hybrid DNA is involved in any of these genetic exchanges, it must be in the form of complex heteroduplexes between the partially homologous sequences. Beyond a possible important evolutionary significance (14) the resolution of these complex heteroduplexes into homoduplexes has at least two interesting implications. (*i*) It offers a mutational mechanism capable of altering several nucleotides (amino acids) in a single step, a process that has often been postulated on the basis of amino acid sequence comparisons (34–36). (*ii*) It may be the source of considerable genetic diversity (3, 13, 14) especially if independent correction events generate patchworks (see ref. 14 for elementary calculations on the number of variants generated). In this regard, we proposed in a variant of the mosaic gene model (16) that it might account for at least some of the variations currently attributed to gene conversion (16, 29, 37, 38), which underlie the polymorphism of class I histocompatibility antigens. Indeed, our observation that an H-2 cDNA sequence could be interpreted as a patchwork of two others (29) initially called our attention to heteroduplex correction.

Our results are only valid for *E. coli*, where heteroduplex correction may have played a role—for instance, in the evolution of temperate phage genomes. Whether, as we predict, an extrapolation will hold for genes of the higher eukaryotes requires experimentation in animal cells.

We are grateful to Dr. M. Radman for helpful discussions and to Drs. M. Cochet and J. P. Abastado for advice in Maxam–Gilbert sequence analysis. This work was supported by Institut National de la Santé et de la Recherche Médicale (S.C. 20 and P.R.C. 124031), Centre National de la Recherche Scientifique (E.R. 201 and A.T.P. 90.5039), and the Fondation pour la Recherche Médicale Française.

1. Radding, C. M. (1978) *Annu. Rev. Biochem.* **47,** 847–880.
2. Radman, M., Wagner, R. E., Jr., Glickman, B. W. & Meselson, M. (1980) in *Progress in Environmental Mutagenesis*, ed. Alacevic, M. (Elsevier/North-Holland, New York), pp. 121–130.
3. Bourguigon Van Horen, F., Brotcorn, A., Caillet-Fauquet, P., Diver, W. P., Dohet, C., Doubleday, O. P., Lecomte, P., Maenhaut-Michel, G. & Radman, M. (1982) *Biochimie* **64,** 559–564.
4. Bale, A., d'Alarcao, M. & Marinus, M. G. (1979) *Mutat. Res.* **59,** 157–165.
5. Wildenberg, J. & Meselson, M. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 2202–2206.
6. White, R. L. & Fox, M. S. (1975) *Mol. Gen. Genet.* **141,** 163–171.
7. Radman, M., Dohet, C., Bourguignon, M. F., Doubleday, O. P. & Lecomte, P. (1981) in *Chromosome Damage and Repair*, eds. Seeberg, E. & Kleppe, K. (Plenum, New York), pp. 431–445.

8. Pukkila, P. J., Peterson, J., Herman, G., Modrich, P. & Meselson, M. (1983) *Genetics* **104**, 571–582.
9. Wagner, R., Jr., & Meselson, M. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 4135–4139.
10. Lai, C. J. & Nathans, D. (1975) *Virology* **66**, 70–81.
11. Miller, L. K., Cooke, B. A. & Fried, M. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3073–3077.
12. Wilson, J. H. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3503–3507.
13. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. (1978) *Science* **202**, 11–17.
14. Kourilsky, P. (1983) *Biochimie* **65**, 85–93.
15. Klein, J. (1978) *Adv. Immunol.* **26**, 55–146.
16. Gachelin, G., Dumas, B., Abastado, J. P., Cami, B., Papamatheakis, J. & Kourilsky, P. (1982) *Ann. Immunol. (Inst. Pasteur)* **133**, 3–20.
17. Kimball, E. S. & Coligan, J. E. (1983) *Contemp. Top. Immunol.* **9**, 1–58.
18. Hood, L., Steinmetz, M. & Malissen, B. (1983) *Annu. Rev. Immunol.* **1**, 529–568.
19. Murray, N. Z., Brammar, W. J. & Murray, K. (1977) *Mol. Gen. Genet.* **150**, 53–61.
20. Kvist, S., Brégégère, F., Rask, L., Cami, B., Garoff, H., Daniel, F., Wiman, K., Larhammar, D., Abastado, J. P., Gachelin, G., Peterson, P. A., Dobberstein, B. & Kourilsky, P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2772–2776.
21. Brégégère, F., Abastado, J. P., Kvist, S., Rask, L., Lalanne, J. L., Garoff, H., Cami, B., Wiman, K., Larhammar, D., Peterson, P. A., Gachelin, G., Kourilsky, P. & Dobberstein, B. (1981) *Nature (London)* **292**, 78–81.
22. Kalderon, D., Ooska, B. A., Ely, B. K. & Smith, A. E. (1982) *Nucleic Acids Res.* **10**, 5161–5181.
23. Dagert, M. & Erlich, S. D. (1979) *Gene* **5**, 23–28.
24. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
25. Roychoudhyry, R. & Wu, R. (1980) *Methods Enzymol.* **65**, 43–47.
26. Gelinas, R. E., Myers, P. A. & Roberts, R. J. (1977) *J. Mol. Biol.* **114**, 169–180.
27. Geier, G. & Madrich, P. (1979) *J. Biol. Chem.* **254**, 1408–1413.
28. McClelland, M. (1981) *Nucleic Acids Res.* **9**, 5859–5866.
29. Lalanne, J. L., Brégégère, F., Delarbre, C., Abastado, J. P., Gachelin, G. & Kourilsky, P. (1982) *Nucleic Acids Res.* **10**, 1039–1049.
30. Slightom, J. L., Blechl, A. E. & Smithies, O. (1980) *Cell* **21**, 627–638.
31. Miyata, T., Yasumaya, T., Yamawaki-Katooka, Y., Obata, M. & Honjo, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2143–2147.
32. Baltimore, D. (1981) *Cell* **24**, 592–594.
33. Egel, R. (1981) *Nature (London)* **290**, 191–192.
34. Yeh, W. K. & Ornston, L. N. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5365–5369.
35. Yeh, W. K., Shih, C. & Ornston, L. N. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3794–3797.
36. Lopez de Castro, J. A., Strominger, J. L., Strong, D. M. & Orr, H. T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3813–3817.
37. Schulze, D. J., Pease, L. R., Geier, S. S., Reyes, A. A., Sarmiento, L. A., Wallace, R. B. & Nathenson, S. G. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 2008–2011.
38. Weiss, E. H., Mellor, A., Golden, L., Fahener, K., Simpson, E., Hurst, J. & Flavell, R. A. (1983) *Nature (London)* **301**, 671–674.
39. Bastia, D. (1978) *J. Mol. Biol.* **124**, 601–639.
40. Rougeon, F., Kourilsky, P. & Mach, B. (1975) *Nucleic Acids Res.* **2**, 2365–2378.
41. Birnbaum, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.