
Base specificity of mismatch repair in *Streptococcus pneumoniae*

Jean-Pierre Claverys, Vincent Méjean, Anne-Marie Gasc, Francis Galibert* and Armand-Michel Sicard

Centre de Recherche de Biochimie et de Génétique Cellulaires du C.N.R.S., 118 route de Narbonne, 31062 Toulouse Cedex, and *Laboratoire d'Hématologie Expérimentale, Centre Hayem, Hôpital St. Louis, Paris, 75010, France

Received 26 March 1981

ABSTRACT

DNA sequence analysis was undertaken to investigate the structural basis of mutations showing different integration efficiencies in *Streptococcus pneumoniae*. Wild type, mutant and revertant sequences at two sites in the *amiA* locus were determined. It appears that markers which transform efficiently or inefficiently can result from single base pair changes. A low efficiency (LE) marker corresponds to a C:G to T:A change and a high efficiency (HE) marker to a G:C to T:A change. In the latter case, two mismatches, G/A and T/C, can exist at the heteroduplex stage in transformation; only T/C appears to be recognized by the *hex* system which controls transforming efficiencies in pneumococcus. Each of the recognized mismatches, T/G and C/A, which result from transitional change, and T/C appears to involve at least one pyrimidine. It is proposed that the mismatch repair system of *S. pneumoniae* is directed against mismatched pyrimidines. DNA sequence analysis also reveals that short deletions (33 or 34 bases long) behave as very high efficiency markers, confirming that deletions are not recognized by the *hex* system.

INTRODUCTION

Genetic transformation in *Streptococcus pneumoniae* involves the insertion of a single strand fragment of donor DNA into the DNA of the recipient bacterium (1). When a mutational difference between the donor and recipient strains occurs within the heteroduplex region, successful integration leads to transformation. The probability for successful integration has a characteristic value for each genetic marker. Markers within the same gene are generally divided into several classes according to their integration efficiency (2, 3). Markers referred to as high efficiency (HE) give yields of transformants approaching 0.5 per bacterial equivalent of transforming DNA taken up by the cells. Other markers described as low efficiency (LE) show a 10 to 20 fold reduction in yields of transformants for the same amount of DNA taken up. Markers which are 1.5 to 2 times as efficient as HE markers have also been described (3,4) and are referred to as very high efficiency markers (VHE) (5).

Excision of donor DNA induced by mismatched base pairs of donor-recipient heteroduplexes was proposed to account for the low efficiency of some markers (6). This hypothesis was substantiated by the identification of mutant strains, denoted *hex*⁻, which are transformed with very high efficiency for all single site markers (7,8). Moreover, the separation of complementary strands of denatured DNA and subsequent preparation of annealed molecules led to the measurement of the efficiency of individual strands (referred to as light = L and heavy = H). It was found that both strands are equally low in efficiency for LE markers (9), and they are both very efficient for VHE markers (10, 9). For HE markers, complementary strands are not equally effective in producing transformants (10, 9). Within the same gene, HE markers with opposite strand preference can be found (9). This phenomenon is also controlled by the *hex* system and interpreted as recognition of one of the two reciprocal donor-recipient heteroduplexes. Additional evidence for excision of mismatched base pairs at the donor-recipient heteroduplex level comes from studies of the kinetics of destruction of LE markers : these kinetics are consistent with elimination of the marker after heteroduplex DNA has been formed (11).

In order to understand the nature of the mismatches recognized by the *hex* system, we have investigated the mutational changes at two sites within the *amiA* locus of pneumococcus by nucleotide sequence analysis. Mutations of this gene confer resistance to aminopterin and have been widely used as markers in transformation studies (2, 5, 9). A set of markers representative of all efficiency classes has been genetically mapped (12) by using the possibility to select under appropriate conditions *amiA*⁺ (aminopterin sensitive) as well as *amiA*⁻ (aminopterin resistant) phenotypes. The recent cloning of *amiA* fragments (13) has allowed us to carry out DNA sequence analysis of mutational changes within this locus. We report our results on the relationship between base mispairing and integration efficiencies.

MATERIALS AND METHODS

Bacterial strains and plasmids

All pneumococcal and *Escherichia coli* strains were previously described (13). Three recombinant plasmids were used for this study : pR10, pR15 and pR16. Each of these plasmids carries a 1050 bp long *Bam*HI - *Eco*RI fragment of the *amiA* locus inserted between the *Bam*HI and *Eco*RI sites of pBR325 (14). pR10 which carries two mutations, *amiA29* and *amiA9*, was constructed and isolated by *in vitro* recombination as previously described for other

amiA fragments (13). Construction of the pneumococcal strain carrying these two closely linked mutations has been described (15). pR15 and pR16 which carry respectively 9 *rev* and 29 *rev* were constructed by "plasmid rescue" in a way similar to that described by Perucho et al. (16; Mejean, Claverys and Sicard, unpublished results).

Enzyme and chemicals

Restriction endonucleases were purchased from New England Biolabs and used as recommended by the manufacturer. Bacterial alkaline phosphatase was from P.L. Biochemicals and DNA polymerase I and polynucleotide kinase were from Boehringer Mannheim. Chemicals utilized were : dimethyl sulfate (Aldrich), hydrazine (Eastman Kodak), acrylamide and bisacrylamide (2 fold crystallized grade, Serva). Piperidine (Merck) was redistilled under vacuum.

Sequencing procedure

10 to 20 picomoles of each plasmid DNA were prepared as described (13) and fully digested with *EcmH1*, DNA was then either 5' end labeled with ^{32}P γ ATP (NEN, above 3000 Ci/mM) using polynucleotide kinase (17) after dephosphorylation at 65°C with alkaline phosphatase (removed by alkaline treatment; 18) or 3' end labeled with ^{32}P α dGTP (NEN, 2500 Ci/mM) using DNA polymerase I (19). End labeled DNA was then cleaved with *EcoRI* and the cloned fragment (1050 bp long) was separated from the vector fragment (4550 bp long) by electrophoresis in a 2 % agarose gel. DNA was electroeluted from agarose gel and chromatographed on a DE52 column (Whatman). Chemical reactions, specific for G, A + G, C + T, C and A/C where A gives a stronger band than C, were performed (17, 19, 20). Sequence reaction products dissolved in the loading solution were heat denatured and fractionated on acrylamide sequencing gels (0.8 mm thick).

25 % acrylamide gels (400 mm long) were run with the bromophenol blue at 90 mm (5' end label) or at 150 mm (3' end labeled) from the top to determine the first 30 nucleotides. Further nucleotide sequences were obtained with i) 16 % acrylamide gels (400 mm long) in which the phenol cyanol FF dye had migrated half way down (5' end label) or two thirds down (3' end label), ii) 8 % acrylamide gels (800 mm long), in which the cyanol FF had migrated to the bottom, iii) 6 % acrylamide gels (800 mm long) in which the phenol cyanol FF had migrated 1.3 times the length of the gel.

RESULTS

Integration efficiencies and choice of markers

A partial genetic map of the *amiA* locus which shows the position of some restriction sites pertinent to this study is shown in fig. 1. From the possible markers, the *amiA29* and *amiA9* sites were retained for DNA sequence analysis. These sites were chosen because a large spectrum of integration efficiencies was covered (LE, HE and VHE, see below) and because they are located close to a *Bam*HI restriction site, a favoured situation for sequence analysis since this site can be used for end labeling.

In a wild type recipient, *amiA29* is a HE marker with a strand preference toward the H strand and *amiA9* is an LE marker (2, 9). Spontaneous revertants of these mutations have been isolated which exhibit a wild type phenotype but are not true reversions to wild type, since the efficiency of each marker is changed when their respective reversions are used as recipient (4). Thus, *amiA29* behaves as a VHE marker when the 29 *rev* strain is used as recipient and *amiA9* behaves as a HE marker in the 9 *rev* recipient (4). In the latter case, the strand preferentially integrated is the L strand (9). In addition, we have been able to determine the efficiency of the mismatches between wild type and 29 *rev* sequences and between wild type and 9 *rev* sequences. This was done by testing the effect of these mismatches on the inte-

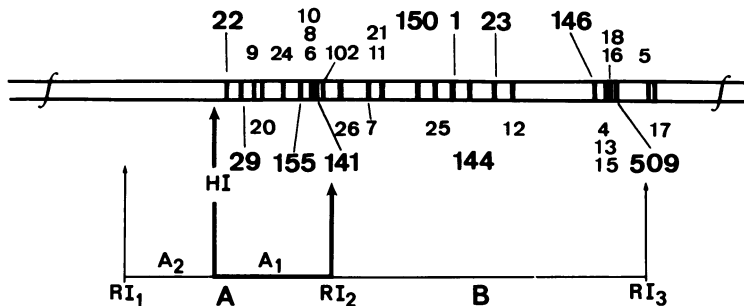


FIGURE 1 : Physical and genetic map of the *amiA* locus. Markers are ordered according to their map position established from genetic crosses (12). RI and HI designate respectively *Eco*RI and *Bam*HI restriction sites. Correlation between genetic and physical distances and location of *Eco*RI sites 1, 2 and 3 have been described (21). *Eco*RI generated fragments A and B are respectively 1800 bp and 2500 bp long. Fragments A1 and A2 are byproducts of the *Bam*HI hydrolysis of the A fragment. Large numbers correspond to HE and VHE markers, small numbers to LE markers.

gration of neighboring markers normally integrated as HE markers. Depending on whether the integration of the markers is affected (exclusion effect) or unaffected by the presence of a given mismatch, the efficiency of this mismatch can be deduced (Table 1).

The mismatch between 29 *rev* and wild type does not affect the integration of *amiA22* or *amiA144*, two HE markers with opposite strand preference. This mismatch is therefore concluded to be of the VHE type. The mismatch between 9 *rev* and wild type, affecting the integration of both *amiA22* and *amiA144* markers is deduced to be of the LE type. The smaller effect on *amiA144* than on *amiA22* integration appears to result from the respective distances between these markers and the *amiA9* site (see fig.1). Indeed, frequent independent integration of *amiA9*⁺ and *amiA144* sites can account for the relatively low exclusion effect observed.

Starting material for DNA sequence analysis was constituted by three recombinant plasmids (pR10, pR15 and pR16). Each carried a *Bam*HI - *Eco*RI fragment (A1) of the *amiA* locus (see fig. 1) inserted between the *Eco*RI and *Bam*HI sites of pBR325. These recombinant plasmids harbor unique *Bam*HI and *Eco*RI sites bracketing the cloned DNA segment. Plasmid pR10 carried *amiA29* and *amiA9* mutations; plasmid pR15 carried 9 *rev* and *amiA29*⁺; plasmid pR16 carried 29 *rev* and *amiA9*⁺.

Partial wild type sequence of the A1 fragment

The strategy used for sequence determination was to linearize the purified plasmid DNA with *Bam*HI, to label either the 5' or the 3' end and to digest with *Eco*RI prior to fractionation of vector (4550 bp) and cloned fragments (1050 bp) by electrophoresis in agarose gels. As shown in the sequen-

TABLE 1 : Exclusion effect of mismatches between wild type and 29 *rev* or 9 *rev* sequences.

Recipient strain	Donor DNA	
	<i>amiA22</i> (<i>ami</i> ⁺ for all other sites)	<i>amiA144</i> (<i>ami</i> ⁺ for all other sites)
29 <i>rev</i>	0.92 ([±] 0.10)	1.12 ([±] 0.25)
9 <i>rev</i>	0.20 ([±] 0.05)	0.63 ([±] 0.08)

Results are expressed as transforming efficiencies relative to our standard reference marker *str41*. *Str41* is a HE marker conferring resistance to streptomycin and is carried on both donor DNAs. *amiA22* and *amiA144* are HE markers (relative transforming efficiency close to 1) with opposite strand preference, H strand for *amiA22* and L strand for *amiA144* (9).

cing gel autoradiogram given in fig. 2, five different reactions were used (G, A + G, C + T, C, A/C). The A1 fragment was sequenced over 320 nucleotides starting from the *Bam*HI site (fig. 3). In one instance a base did not react as expected : 5' end labeled fragments gave a band for all chemical reactions although stronger for A + G and A/C than for other reactions at position 221 (fig. 4). This result led to the possibility that more than one base could be present. Nevertheless the existence of only one A:T base pair at this position was shown by analysis of the other DNA strand after labeling of the 3' end (fig. 4).

Since the A1 fragment is internal to the *amiA* gene, it is of interest to look at the open reading frames which could indicate which strand is transcribed. The reading frame of the nucleotide sequence of the A1 fragment will be defined from the *Bam*HI cleavage site : reading frame 1 will be GAT,

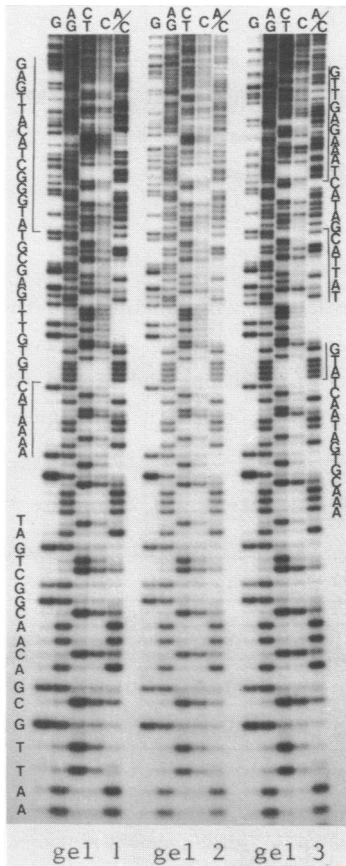


FIGURE 2 : Autoradiogram of sequencing gels. 5' end labeled A1 fragment obtained from plasmid pR10 (gel 1), pR15 (gel 2) and pR16 (gel 3) were subjected to the Maxam-Gilbert chemical degradation. Products were fractionated by electrophoresis on 16 % acrylamide gel. Gel 1, 2 and 3 cover the sequence from nucleotide 18 to nucleotide 108.

Mutational changes at the *amiA29* site

Fig.5 shows sequencing gel autodiagrams used to determine wild type, *amiA29* and 29 *rev* nucleotide sequences. The *amiA29* mutation is a frameshift mutation which occurred by replacement of three bases GCA₁₃₁ in the sequence AGCATGGATTCT by four bases TGGC to give the sequence AGCATTGGCTTCT. This frameshift mutation leads rapidly to a stop of translation at TAA₁₆₇. Reversion of the *amiA29* mutation occurred spontaneously by deletion of 33 nucleotides as compared to wild type sequence, from residue 120 to residue 153. This dele-

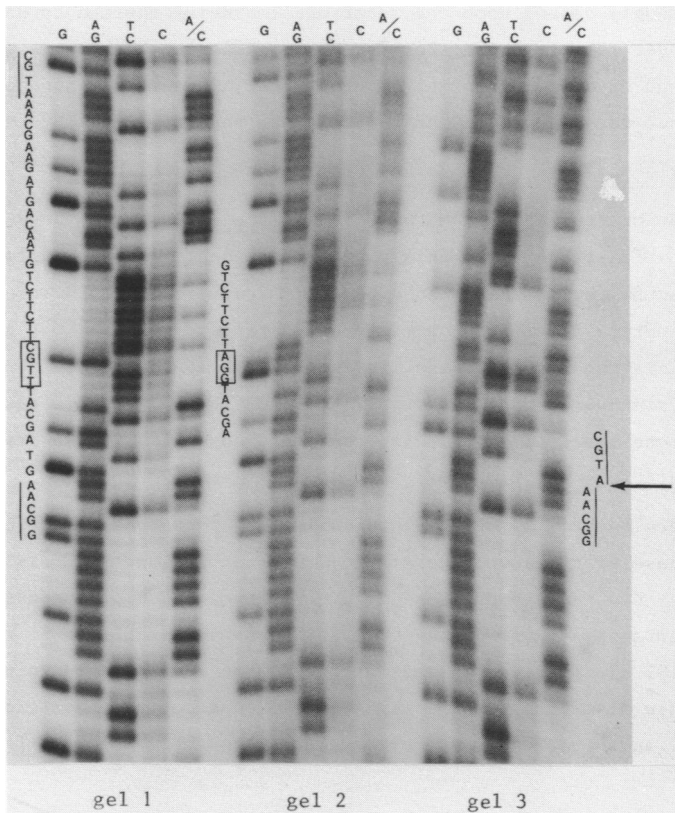


FIGURE 5 : *amiA29* site sequencing gels. 5' end labeled Al fragments were subjected to the Maxam and Gilbert chemical degradation. Products were fractionated by electrophoresis on 8% acrylamide gels. Sequence changes between *amiA29* (gel 1) and wild type (gel 2) are shown in boxes. The position of the deletion corresponding to 29 *rev*(gel 3) is indicated by an arrow. Sequences bracketing the deletion are indicated by vertical bars (gel 1 and 3). Gels 1 and 2 cover the sequence from nucleotide 105 to nucleotide 158 ; gel 3 from nucleotide 105 to nucleotide 193.

tion covers the *amiA29* site (around position 130) and thus restores a normal reading frame, leading to a protein 11 amino-acids shorter than the wild type protein. It is interesting to note that this deletion has occurred between two four base repeats (GCAA₁₂₁ and GCAA₁₅₄) leaving only one intact GCAA sequence, a result reminiscent of that of Farabough et al. in *Escherichia coli* (24).

DISCUSSION

Transformation involves the insertion of a strand of donor DNA into a recipient molecule to form a heteroduplex region (1). Efficiencies of transformation strongly depend upon the mutations (markers) carried by donor DNA (2, 3). In order to relate transforming efficiencies to the nature of mismatches present on donor-recipient heteroduplex, nucleotide sequence analysis was carried out for six alleles at two sites of the *amiA* locus.

Single base pair changes and transforming efficiencies

Mutation from wild type to *amiA9* occurred by a single base pair change, the C:G to T:A transition. This single base change leads to a low efficiency of transformation in a wild type (*hex*⁺) recipient. Depending upon the strand which enters a cell, donor-recipient heteroduplexes exhibit either a G/T or A/C mismatch. Each of these mismatches thus appears to be recognized by the *hex* system which controls transforming efficiencies. This result agrees with suggestions based on the finding that mutagens which induce transitions lead to LE mutations (2, 3, 25).

Reversion from *amiA9* to 9 *rev* (wild type phenotype) occurred also by a single base pair change, the A:T to C:G transversion. *amiA9* is integrated into the 9 *rev* recipient with a high efficiency and opposite strands are not equally effective in producing transformants, a characteristic feature of HE markers (10, 9). This result was interpreted as specific recognition of one of two possible mismatches by the *hex* system. Donor-recipient heteroduplexes between *amiA9* and 9 *rev* can carry either A/G (l chain *amiA9*) or C/T (r chain *amiA9*) mismatches.

Which one of the two reciprocal mismatches A/G or C/T is recognized ?

We have shown that the L strand carrying *amiA9* is five times more efficient than the H strand in transformation of the 9 *rev* recipient : this is interpreted as recognition of H *amiA9*/L 9 *rev* mismatch by the *hex* system (9). If the H strand of the *amiA* locus can be identified as the l or r chain, this will give us the answer as to whether A/G or C/T is the recognized mis-

match. We have previously shown that the rate of phenotypic expression of *amiA* markers in a *hex*⁻ recipient is faster when they are introduced *via* the H strand (9). This result indicates that the H strand is the strand transcribed. Thus we can identify the H strand as the r chain. This identification is supported by the fact that only one reading frame out of a possible six is open and that the *amiA9* mutation leads to the appearance of a stop codon in this open reading frame. Therefore the mismatch H *amiA9*/L 9 *rev* corresponds to C/T (r chain *amiA9*) which is the third mismatch recognized by the *hex* system.

Multisite base pair changes and transforming efficiencies

DNA sequence analysis also reveals that multiple base pair changes can be associated with LE (9 *rev* facing wild type) or HE (*amiA29* facing wild type) mismatches corrected by the *hex* system. Nucleotide sequence shows that two base pair changes occurred between 9 *rev* and wild type, around residue 239 -TAC- for 9 *rev* instead of -CAA-. Thus two mismatched base pairs normally recognized as LE and HE are present, separated by a single base pair. This double base pair change is still recognized as LE by the *hex* system.

Change from wild type to *amiA29* involves the replacement of three bases GGA by four bases TTGC. *amiA29* is integrated in the wild type recipient with a high efficiency and exhibits a strand preference characteristic of HE markers. This behaviour is similar to that of *amiA9* in the 9 *rev* recipient (see above) except that the more efficient strand is the H strand (9). This was interpreted as recognition of the L *amiA29*/H 9 *rev* mismatch by the *hex* system. Since we have identified the H strand as the r chain, this mismatch is :

l <i>amiA29</i> chain	A	T	T	T	G	C	T	T
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
r wild type chain	T	A	C	C	T	A	A	A

The only single base pair mismatches present here are the C/T type, i.e. mismatches that were shown in a previous section to be recognized by the *hex* system. The complementary donor-recipient heteroduplex between L wild type/H *amiA29* is as follow :

l wild type chain	A	T	G	G	A	T	T
	⋮	⋮	⋮	⋮	⋮	⋮	⋮
r <i>amiA29</i> chain	T	A	A	A	C	G	A

In this case, the only mismatches present are the A/G type which, as shown in a previous section, is not recognized by the *hex* system.

Thus high efficiency of transformation and strand preference can be related to specific recognition by the *hex* system of the C/T mismatch and not of the reciprocal A/G mismatch. This seems to be true when these mismatches are present either singly (as seen with *amiA9*) or in multiples (as seen with

amiA29).

Short deletions result in very high efficiency of transformation

The last results obtained by DNA sequence analysis concern the transformation efficiency of short deletions. We found that reversion from *amiA29* to 29 *rev* (wild type phenotype) has occurred by deletion of 34 bases. *amiA29* is integrated into the 29 *rev* recipient as a VHE marker and we have provided evidence (see results) that mismatching between wild type and 29 *rev* sequences is also of the VHE type. Thus in both cases, a deletion of 33 or 34 bases is not recognized by the *hex* system and behaves as a VHE marker. This result is not very surprising since it is known that larger deletions are not recognized by the *hex* system (7,9,26) and that deletions as large as 200-240 base pairs exhibit a high efficiency of transformation (9,26). This raises the possibility that VHE markers are not single site mutations.

Specificity of the *hex* system

The insertion of a strand of donor DNA carrying a point mutation into a recipient molecule can lead to eight possible mismatches at the heteroduplex stage. The four of them which have been found in the present study reveal part of the specificity of the *hex* system. It should be noted that among those mismatches which are recognized by the *hex* system, all contain at least one pyrimidine : A/C, G/T and C/T, whereas the mismatch A/G which escapes the *hex* action involves two purines. This result opens the possibility that the *hex* mismatch repair system is directed against mispaired pyrimidines. If this is true, we can predict that the four remaining mismatches not encountered in this study will correspond to HE mutations. Indeed transversional changes A:T to T:A and G:C to C:G lead respectively to the following mismatched sets of pairs : (A/A, T/T) and (G/G, C/C) which contain only one mismatch involving a pyrimidine. Support for this hypothesis comes from results at the *amiA6* site where various mismatches between wild type, mutant and revertant sequences are of the HE type (Claverys and Sicard, unpublished results). In any case, further DNA sequence analysis for other markers should give a complete picture of the specificity of the pneumococcal mismatch repair system.

Another open question concerns frameshift mutations. It has been shown that acridine induced mutations were mainly of the LE type and evidence has been presented that these mutations were frameshift mutations (27). Our results clearly suggest that at least one frameshift mutation is not recognized by the *hex* system. It is possible that acridine induced mutations are

complex mutations which frequently harbor mismatches recognized by the *hex* system when paired with wild type sequence. This could explain the low efficiency of transformation of these mutations without need for recognition of frameshift mutations *per se* by the *hex* system.

The last point concerns the mechanism of co-correction of VHE or HE markers with adjacent LE markers (3,4,6). This co-correction can result from excision of a large tract of bases once a mismatch is recognized by the *hex* system. Alternatively, it has been suggested that once triggered by an LE mismatch, the *hex* system becomes capable of short patch repair for all mismatches present on the heteroduplex (28). Since we have found that VHE type transformations at the *amiA29* site (which are normally co-corrected with the LE mismatch involving *amiA9*) do not involve a mismatch but a short deletion, one must assume either that *hex* recognition is followed by long patch correction or that once triggered the *hex* system not only excises all mismatches it finds, but also sequences corresponding to deletions.

ACKNOWLEDGEMENT

We thank B. Stevens and J.P. Bouché for their help during the preparation of this manuscript. This work was partly supported by an I.N.S.E.R.M. grant (7279104030).

REFERENCES

* Normal base pairs will be written as G:C whereas mismatched base pairs will be symbolized G/T. The first written base of a pair will normally correspond to the base located on the L strand.

1. Fox, M.S. and Allen, M.K. (1964) Proc. Nat. Acad. Sci. USA 52, 412-419.
2. Ephrussi-Taylor, H., Sicard, A.M. and Kamen, R. (1965) Genetics 51, 455-475.
3. Lacks, S. (1966) Genetics 53, 207-235.
4. Sicard, A.M. and Ephrussi-Taylor, H. (1966) C.R. Acad. Sci. Paris 262, 2305-2308.
5. Tiraby, G. and Sicard, A.M. (1973) J. Bacteriol. 116, 1130-1135.
6. Ephrussi-Taylor, H. and Gray, T.C. (1966) J. Gen. Physiol. 49/2, 211-231.
7. Lacks, S. (1970) J. Bacteriol. 101, 373-383.
8. Tiraby, G. and Fox, M.S. (1973) Proc. Nat. Acad. Sci. USA 70, 3541-3545.
9. Claverys, J.P., Roger, M. and Sicard, A.M. (1980) Mol. Gen. Genet. 178, 191-201.
10. Roger, M. (1977) J. Bacteriol. 129, 298-304.
11. Shoemaker, N.B. and Guild, W.R. (1974) Molec. Gen. Genet. 128, 283-290.
12. Sicard, A.M. and Ephrussi-Taylor, H. (1965) Genetics 52, 1207-1227.

13. Claverys, J.P., Louarn, J.M. and Sicard, A.M. (1981) *Gene* 13, 65-73.
14. Bolivar, F. (1978) *Gene* 4, 121-136.
15. Claverys, J.P., Lefèvre, J.C. and Sicard, A.M. (1979) in *Transformation 1978*, Glover, S.W. and Butler, L.O. Eds, pp 135-150 Costwold, Oxford.
16. Perucho, M., Hanahan, D., Lipsich, L. and Wigler, M. (1980) *Nature* 285, 207-210.
17. Maxam, A.M. and Gilbert, W. (1977) *Proc. Nat. Acad. Sci. USA* 74, 560-564.
18. Knoecker, W.D. and Laskowski, M.S.R. (1977) *Analyt. Biochem.* 79, 63-73.
19. Maxam, A.M. and Gilbert, W. (1980) in *Methods in Enzymology*, Grossman, L. and Moldane, K. Eds, vol. 65, pp 499-560, Academic Press, London.
20. Galibert, F., Hérisse, J. and Courtois, G. (1979) *Gene* 6, 1-22.
21. Claverys, J.P., Lataste, H. and Sicard, A.M. (1979) in *Transformation 1978*, Glover, S.W. and Butler, L.O. Eds, pp 161-169 Costwold, Oxford.
22. Drake, J.W. (1970) In *Molecular Basis of Mutation*, pp 123-145, Holden Day, San Francisco.
23. Gasc, A.M., Vacher, J., Buckingham, R. and Sicard, A.M. (1979) *Molec. Gen. Genet.* 172, 295-301.
24. Farabough, P.J., Schmeissner, U., Hofer, M. and Miller, J. (1978) *J. Mol. Biol.* 126, 847-863.
25. Gasc, A.M. and Sicard, A.M. (1972) *C.R. Acad. Sci. Paris* 275, 285-287.
26. Lataste, H., Claverys, J.P. and Sicard, A.M. (1980) *J. Bacteriol.* 144, 422-424.
27. Gasc, A.M. and Sicard, A.M. (1978) *Genetics* 90, 1-18.
28. Guild, W.R. and Shoemaker, N.B. (1976) *J. Bacteriol.* 125, 125-135.