

Evolution of Antibiotic Resistance Genes: The DNA Sequence of a Kanamycin Resistance Gene from *Staphylococcus Aureus*¹

Gary S. Gray

Biogen S.A., Switzerland

Walter M. Fitch

University of Wisconsin—Madison

The kanamycin resistance gene from *Staphylococcus aureus* has been sequenced and its structure compared with similar genes isolated from *Streptomyces fradiae* and from two transposons, Tn5 and Tn903, originally isolated from *Klebsiella pneumoniae* and *Salmonella typhimurium*, respectively. The genes are all homologous but, since their common ancestor, have undergone extensive divergence, with more than 43% divergence between the closest pair. The phylogeny of the genes cannot be made congruent to the phylogeny of the taxa from which they were isolated without requiring rather improbable differences in rates. One is therefore led to conclude that there have been multiple occurrences of gene transfer between these species. Thus, although they are homologous, they are neither orthologous nor paralogous. It is suggested that homologous genes of this type be called xenologous.

Introduction

Staphylococcus aureus, resistant to aminocyclitol antibiotics, produces a variety of mechanistically different aminocyclitol-modifying enzymes including the aminocyclitol-3'-phosphotransferase (APH[3']-III). Similar enzymes have also been described in the gram-positive genus *Streptococcus* (Courvalin et al. 1980a), the Enterobacteriaceae (Smith 1978) and Pseudomonadiaceae (Matsushashi et al. 1975), aminocyclitol-producing *Bacillus circulans* (Courvalin et al. 1977), and actinomycetes (Davies et al. 1979; Thompson et al. 1980). The APH(3') activities from various sources all catalyze the phosphorylation of neomycin and kanamycin and can be subgrouped according to their reactivity with other aminocyclitol antibiotics (Davies and Smith 1978). Although the APH(3') genes and their gene products are functionally related, studies employing both immunochemical methods and DNA hybridization have shown only the staphylococcal and streptococcal genes to be related (Davies and Smith 1978; Smith 1978; Courvalin et al. 1980a). However, comparison of the DNA sequences of the APH(3') genes from trans-

1. Key words: kanamycin resistance, evolution of antibiotic resistance, aminocyclitol-3'-phosphotransferase gene sequence, transfection.

Address for correspondence and reprints: Gary S. Gray, Biogen S.A., 46, Route des Acacias, 1227 Carouge/Geneva, Switzerland.

Mol. Biol. Evol. 1(1):57-66.

© 1983 by The University of Chicago. All rights reserved.

0737-4038/83/0101-0009\$02.00

posons Tn5 and Tn903 of gram-negative bacteria and from *Streptomyces fradiae* revealed marked similarity (Thompson and Gray 1983). It has been hypothesized that an aminocyclitol-producing bacterium was the source of the genes for the widely disseminated aminocyclitol phosphotransferases (Benveniste and Davies 1973). To extend the analyses of genes encoding aminocyclitol-modifying enzymes and to test that hypothesis, we report here the DNA sequence of the APH(3')-III gene encoded by an *S. aureus* plasmid and present a detailed comparison of the four known APH(3') gene sequences. These results confirm the close structural relationship among five phosphotransferase genes. They also show that the hypothesis about the source of the gene in resistant strains is correct.

Material and Methods

The aminocyclitol-3'-phosphotransferase gene from *Staphylococcus aureus* plasmid pSH2 was cloned in *Escherichia coli* (Courvalin and Fianndt 1980) and subsequently subcloned after partial digestion with *Hind*III. The resulting recombinant plasmid, pAT48, confers kanamycin resistance to *E. coli* and possessed 3.5 kbp of inserted DNA which contained two internal *Hind*III sites. An attempt to subclone the large insert-specific *Hind*III fragment from pAT48 by selection of transformants for kanamycin resistance was unsuccessful, suggesting that a DNA sequence surrounding at least one of the two *Hind*III sites was essential for transcriptional or translational activity. Therefore, DNA sequencing was initiated at the ends of this large *Hind*III fragment. The DNA sequence around one of the *Hind*III sites proved to be similar to that found for other APH(3') genes (Thompson and Gray 1983); subsequently, adjacent DNA sequences were determined by the method of Maxam and Gilbert (1980) according to the strategy in figure 1.

Results

Figure 2 shows the primary structure of the APH(3') gene of *Staphylococcus aureus* and the predicted protein sequence for the open reading frame of 789 base pairs (fig. 2, residues 293–1082). Within the coding region, there were 62 translation termination codons distributed among the other five possible reading frames. The translation initiation site of the staphylococcal APH(3') gene was assigned to the methionine codon at residue 293; initiation at this codon would yield a protein of MW = 30,724 (MW = molecular weight, in daltons). This estimate agrees with the value found for purified APH(3') protein (MW = 29,000 ± 300) (Smith 1978) and that determined for the insert-specific protein (MW = 31,000) obtained after coupled in vitro transcription and translation of pAT48 DNA. Inspection of the DNA sequences immediately preceding the proposed initiation codon reveals a ribosome binding site (GGAAGG, residues 277–282) and transcription initiation sequences (residues 230–260 or 210–240) similar to those found for other staphylococcal genes (Horinouchi and Weisblum 1982a, 1982b). The region following the translational stop codon (residue 1033) contains an inverted repeated (IR) sequence of 19 bp which presumably functions in RNA transcription termination (Rosenberg and Court 1979); similar sequences occur after the staphylococcal chloramphenicol (Horinouchi and Weisblum 1982a) and erythromycin resistance genes (Horinouchi and Weisblum 1982b) and many gram-negative sequences (Rosenberg and Court 1979). The two *Hind*III sites located in the IR sequence (residues 1109–1122) mark the end of the large *Hind*III fragment; their location suggests that the failure to obtain kanamycin resistant clones when sub-

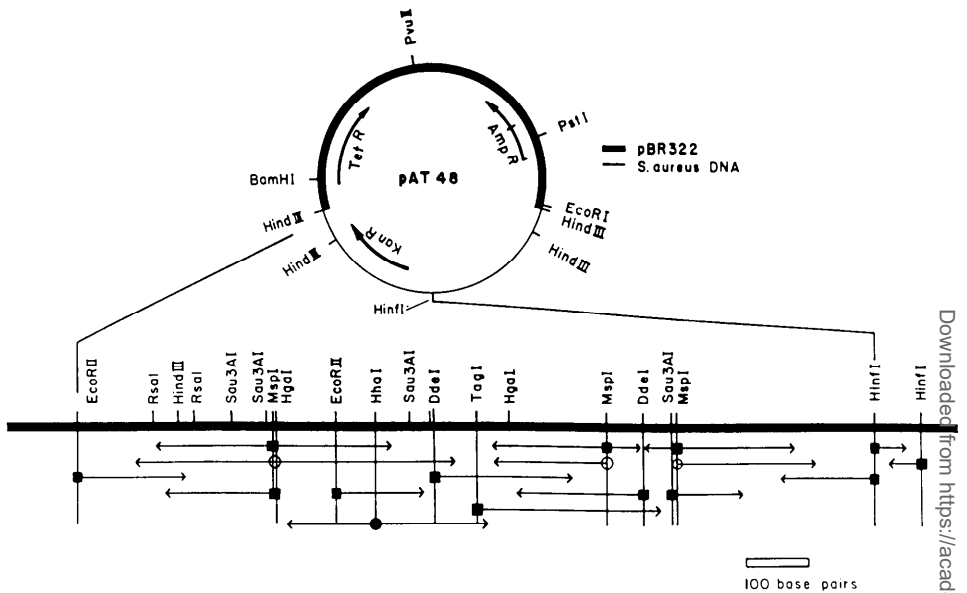


FIG. 1.—Restriction enzyme map and DNA sequencing strategy for the *Staphylococcus aureus* aminocyclitol-3'-phosphotransferase gene. Plasmid pAT48 was constructed by ligation of partially *Hind*III digested recombinant plasmid pAT4 (Courvalin and Fianndt 1980) and *Hind*III cleaved pBR322. The restriction map shows only the relevant restriction endonuclease sites. DNA sequencing reactions were performed according to the method of Maxam and Gilbert (1980). The direction and extent of each sequence determination are shown by the horizontal arrows below the restriction map; isolated DNA restriction fragments were end-labeled at the 5' end with [γ - 32 P]-ATP and polynucleotide kinase (Maxam and Gilbert 1980), ■; or on the 3' strand by [α - 32 P]-cordycepin triphosphate and terminal transferase (Tu and Cohen 1980), ●; and by "filling in" with [α - 32 P]-dNTP's and DNA polymerase, large fragment (Backman et al. 1976), ○.

cloning the *S. aureus* large *Hind*III fragment was due to loss of proper transcription termination.

The proposed amino acid sequence for the *S. aureus* APH(3')-III gene and those predicted for the resistance genes of transposons Tn5 and Tn903 and the neomycin-producing *Streptomyces fradiae* are compared in figure 3. The amino acid sequences exhibit regions of extensive similarity, with the carboxyl terminal segments being the most similar. A conserved cysteine residue at position 145 is flanked by two regions (residues 115–175) containing a high frequency of acidic amino acids which may function in the binding of the basic aminocyclitol antibiotics to the modifying enzyme.

That the sequences are very divergent is shown by counting the number of positions with different numbers of nucleotides in them. They all may be different (1:1:1:1), or may have only one pair (2:1:1), two pairs (2:2), a triplet (3:1), or a quartet (4) of nucleotides alike. These cases number 26, 270, 103, 246, and 213, respectively, and include the termination codon. A gapped nucleotide was counted as identical to the most frequent nucleotide in this count so that the differences reflect nucleotide substitutions exclusive of insertions or deletions and are therefore conservative with respect to total divergence. There are 21 distinguishable gaps comprising 39 positions. Less than 25% of the positions have no nucleotide differences, while more than 42% have three or more different nucleotides.

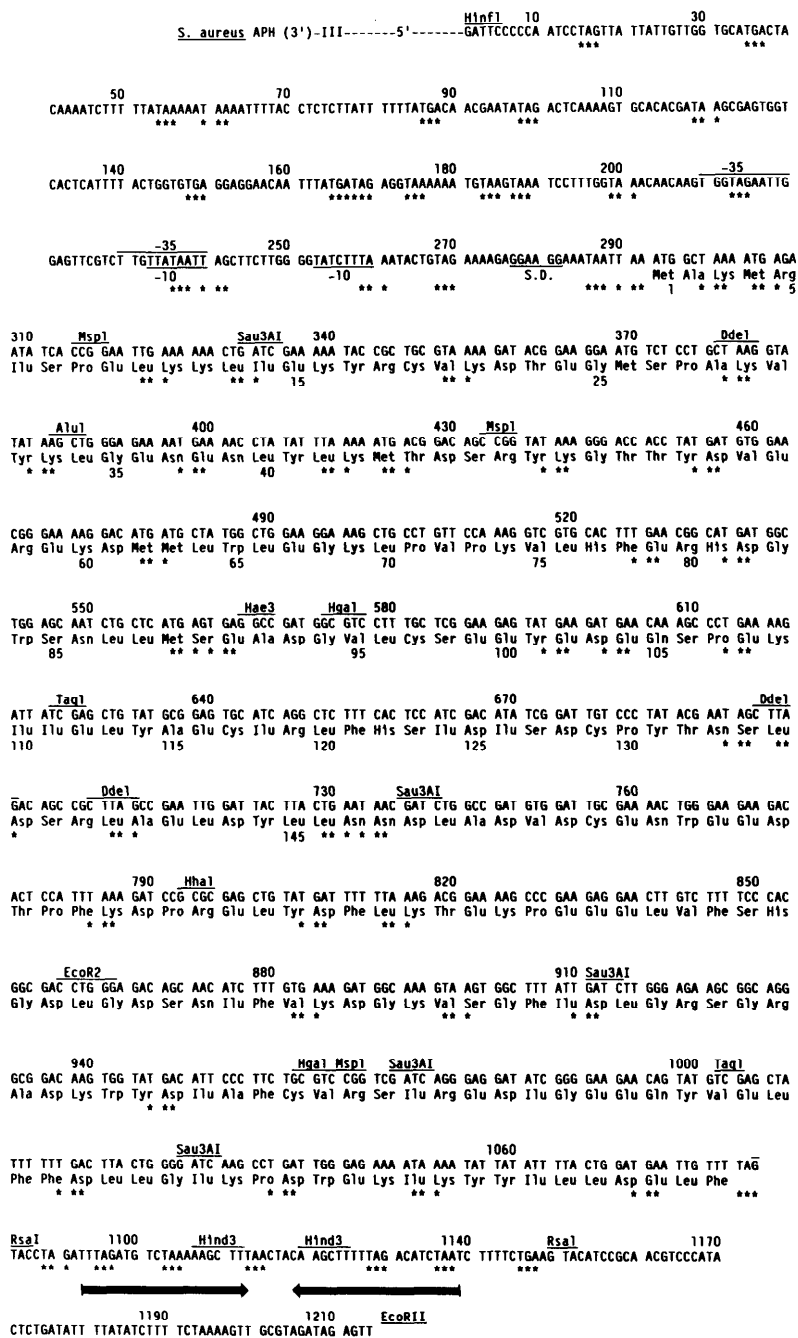


FIG. 2—DNA sequence of the aminocyclitol-3'-phosphotransferase gene of *Staphylococcus aureus*. The sequence represented represents the sense strand of DNA commencing at the 5' end with the HinfI site. The nucleotide residue corresponding to the proposed initiation codon is located at residue 293. Pregene sequences include a Shine and Dalgarno (1974) sequence (S.D.; residue 280) and two possible promoter sequences (−10, −35; residue 210–240 and 230–260). The predicted amino acid sequence is shown below the DNA sequence and the location of translational stop codons indicated (***). Arrows indicate the block of inverted repeated symmetric sequences immediately following the predicted coding region.

The phylogeny was investigated using a parsimony analysis (Fitch 1971), for which only the 103 2:2 cases are informative and, for any one nucleotide position, may be assigned to one of three ordered sets, MNMN = 48, MMNN = 31, and MNNM = 24, where M and N represent the two kinds of nucleotides in the same sequence order as in figure 3 and where $48 + 31 + 24 = 103$ accounts for all the

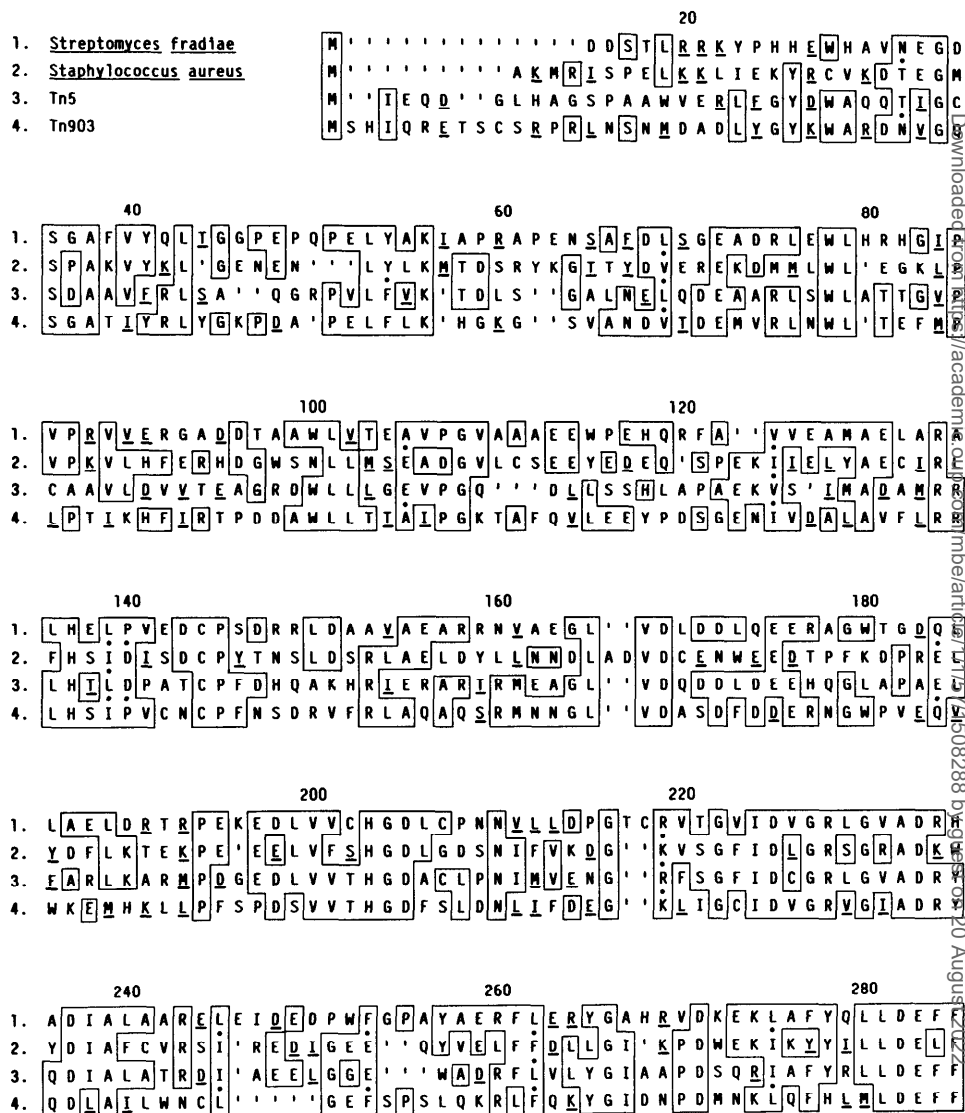


FIG. 3.—Comparison of the amino acid sequences of aminocyclitol-3'-phosphotransferases of *Staphylococcus aureus*, *Streptomyces fradiae* (Thompson and Gray 1983), and gram-negative transposons Tn5 (Beck et al. 1982) and Tn903 (Oka et al. 1981). Sequence locations containing identical amino acid residues are enclosed within solid lines, and sites of conservative amino acid replacements are enclosed within dotted lines. Conservative means within the groups [D,E], [K,R], [S,T], [F,Y], and [I,L,V,M]. Gaps have been inserted to improve the sequence alignment. The one-letter amino acid codes are A = ala, C = cys, D = asp, E = glu, F = phe, G = gly, H = his, I = ileu, K = lys, L = leu, M = met, N = asn, P = pro, Q = gln, R = arg, S = ser, T = thr, V = val, W = trp, and Y = tyr.

cases. There are only three possible unrooted tree structures, and a tree requiring only one nucleotide substitution for one of these three sets will require two substitutions in each of the other two trees. Thus the three possible unrooted topologies for these crucial positions will require $48 + 2(24 + 31) = 158$, $31 + 2(48 + 24) = 175$, and $24 + 2(48 + 31) = 182$ nucleotide substitutions. To each of these must be added 864 additional substitutions for the uniquely derived nucleotides. These trees are shown in their respective order at the top of figure 4.

Of the 1,022 nucleotide substitutions in the left tree, 242 do not change the encoded amino acid. The number of nucleotide substitutions in the first, second, and third codon positions is 313, 260, and 449, respectively. The ratio of transversions to transitions was 666/356 overall and not greatly different by codon position.

Discussion

The alignment in figure 3 shows sufficient amino acid identities and similarities that one can hardly doubt that the sequences are homologous (i.e., have a common ancestor). In this analysis, codon gaps were constrained to be in common where possible and to include only whole codons. The *Streptomyces fradiae* DNA se-

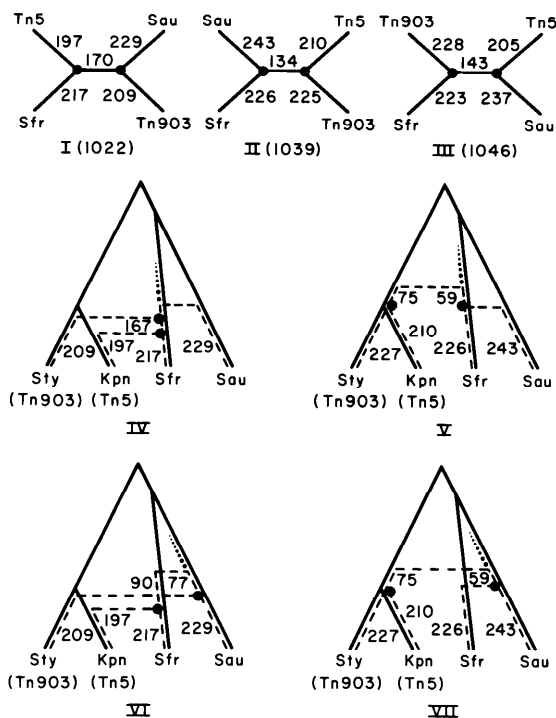


FIG. 4.—The top shows the three possible unrooted tree topologies for the four APH(3') sequences and the mean distribution of the minimum number of nucleotide substitutions required to account for their sequence differences. The lower portion shows, with dashed lines, trees I and II superimposed on the presumed history of the taxa from which the sequences were obtained (IV–VII, respectively). Horizontal dashed lines imply transmission between species. The values 75 and 59 in trees V and VII sum to the 134 substitutions on the internodal interval of tree II. They are to be associated with the vertical portions between the nodes (●) rather than the horizontal portion. Sfr = *Streptomyces fradiae*, Sau = *Staphylococcus aureus*, Kpn = *Klebsiella pneumoniae*, Sty = *Salmonella typhimurium*.

quence contained an additional 12 nucleotides at the 5' end because of the possibility of translational initiation at either of the methionine codons located here.

While parsimony clearly favors tree I (fig. 4), the total amount of change is so great as to make any definite conclusion uncertain. One is no better off using a method based on overall amounts of change (e.g., Fitch and Margoliash 1967) since any pair of sequences is about as divergent as any other pair (table 1, range 397–454 base differences), although again tree I is marginally preferred over tree II. If the nucleotide sequences are ignored and one resorts to the amino acid sequences, tree III is preferred on the basis of minimum base differences (see table 1) when an average linkage method such as that of Fitch and Margoliash (1967) is used.

Since two of the sequences are encoded by transposable elements, it is probably not meaningful to ask if tree I agrees with the presumed phylogeny of the four taxa that are host to these sequences. It is instructive, however, to inquire about the relation between the two since any discrepancy between them would be accountable by, and expected on the basis of, a plasmid-mediated acquisition by its host of another taxon's defense against aminocyclitol antibiotics. For this purpose, we assume that the first maker of the antibiotic had to have the enzymatic defense mechanism to protect itself from its own weaponry. We shall assume for the moment that the *Streptomyces fradiae* lineage serves that role since it is the only taxon present that makes aminocyclitols. This origination is represented by the small dotted lines uppermost on trees IV and V in figure 4 where topologies I and II have been superimposed (dashed lines) on the presumed phylogeny of the four taxa (solid lines).

The implication of tree IV is that all three of the other lineages independently acquired their gene copies from the *fradiae* line. This can be reduced to two acquisitions in tree V by having the *Klebsiella*(Tn5)-*Salmonella*(Tn903) lineage acquire the gene (transposon) prior to the divergence of those two genera. The principal argument in favor of this version is that by splitting the 134 substitutions into 75 and 59 on the upper left and right dashed lines of descent, one obtains a solution with a magnificently uniform rate of descent in all four lines.

However, since the two topologies (species and gene) are identical, no independent gene acquisitions need to be postulated, as can be seen by stretching the upper bifurcation point on the *fradiae* line upward to the apex representing the common ancestor of all four taxa until the branch points of both trees are superimposed. The principal argument against this is the correspondingly large distortion of uniformity of rate of change with time as more than 40% of all the

Table 1
Base Differences for APH(3') Sequences

	TN5	Sfr	903	Sau
TN5.....	...	231	236	242
Sfr.....	397	...	240	271
903.....	421	445	...	261
Sau.....	422	454	420	...

NOTE.—The portion of the table above the diagonal shows the minimum base differences as if only the amino acid sequences in fig. 3 were known; the portion below the diagonal shows the actual base differences observed when the nucleotide sequences are examined in the same coding alignment.

required changes must be incorporated in the relatively small amount of evolutionary time (crudely, 20%) since the common ancestor of *Klebsiella* and *Salmonella* compared with the sum of the other evolutionary times on the tree. Of course, being resident on a transposon may somehow contribute the necessary additional increase in rate. But there is no reason to stretch the point to avoid independent gene acquisitions since they cannot be avoided. Courvalin (personal communication) has sequenced this gene isolated from a transposon from *Streptococcus faecalis*, and it is identical to that of *Staphylococcus aureus* reported here. These two species are much too distantly related for at least one of these two genes not to have been recently acquired by transfection. A similar situation would appear to have occurred in the histones of sea urchins (Busslinger et al. 1982), except that the responsible vector is unknown in their case.

The trees in figure 4 make it clear once again that "homology is not enough. That is to say, any attempt to use homologous sequences to infer the phylogeny of the taxa is in danger of error if the sequences are not the right type of homology (Fitch 1970). If the sequences diverged because the taxa containing the gene diverged, this is the subcategory of homology called orthology, and a good species phylogeny requires orthologous sequences. If the sequences diverged following a gene duplication, this is the subcategory of homology called paralogy, and mixing paralogous sequences such as the alpha and beta hemoglobins (one or the other for each taxon) would be a good method for getting a bad species phylogeny. Clearly transfection (and symbiosis and parasitism as well) represents a way that cells and organisms have acquired foreign genes in the past, and, since they are neither orthologous nor paralogous but are clearly homologous (*sensu strictu*), perhaps they should be called xenologous (xeno = foreign).

Finally, it should be noted that it would be premature to assume that these genes originated in the *Streptomyces* line. Since *Bacillus circulans* also produces an aminocyclitol antibiotic, one might as easily believe, on the limited evidence available, that these genes first appeared in the *Bacillus* lineage which, phylogenetically, would form a branch off the *Staphylococcus* line in figure 4 (trees VI and VII). Remarkably, shifting the origin to the latter lineage would have little effect on the preceding discussion. The double trees of figure 4 VI and VII differ from IV and V only in that shift. Hypotheses VI and VII do, however, raise a new question, namely, Why don't staphylococcal species make aminocyclitols if their ancestors did? One is still left with a narrow choice between tree I and tree II on the grounds of parsimony versus uniform rates, neither of which is a given principle of nature (tree III is the worst on both grounds). Nor can one dismiss the possibility that different lineages acquired their defenses from different offensive lineages such as other aminocyclitol producing actinomycetes. The homology of these sequences is clear, as is their transfection, and one cannot help being tantalized by the prospect and potential of further sequence information to resolve the question of their origins.

Note added in proof.—C. J. Herbert, I. G. Giles, and M. Akhtar (FEBS Letters 160:67–71, 1983) have now sequenced the APH gene from *Bacillus circulans*. It is homologous to the genes in this paper and has a range of distances to these genes similar to that which these genes have among themselves. The *B. circulans* gene is marginally closer to the *Staphylococcus aureus* gene than to the others. This seems to imply that either the *Streptomyces* or the *Bacillus* lineage (probably the latter) acquired this defense by transfection relatively recently in its history

and raises the question whether its offensive weaponry might not also have been obtained by transfection.

Acknowledgments

We thank Julian Davies for helpful discussions and a critical review of this manuscript and Sylvia Robertson and Ron Niece for technical assistance in the computing. This investigation was supported by Biogen S.A. and National Science Foundation grant DEB-78-14197.

LITERATURE CITED

- BACKMAN, K., M. PTASHNE, and W. GILBERT. 1976. Construction of plasmids carrying the *cI* gene of bacteriophage λ . *Proc. Natl. Acad. Sci. USA* **73**:4174–4178.
- BECK, E., G. LUDWIG, E. AUERSWALD, B. REISS, and H. SCHALLER. 1982. Nucleotide sequence and exact localization of the neomycin phosphotransferase gene from transposon Tn5. *Gene* **19**:327–336.
- BENVENISTE, R., and J. DAVIES. 1973. Aminoglycoside antibiotic-inactivating enzymes in actinomycetes similar to those present in clinical isolates of antibiotic-resistant bacteria. *Proc. Natl. Acad. Sci. USA* **70**:2276–2280.
- BUSSLINGER, M., S. RUSCONI, and M. L. BIRNSTIEL. 1982. An unusual evolutionary behavior of a sea urchin histone gene cluster. *EMBO J.* **1**:27–33.
- COURVALIN, P., C. CARLIER, and E. COLLATZ. 1980a. Evolutionary relationships between plasmid-mediated aminoglycoside-modifying enzymes from gram-positive and gram-negative bacteria. Pp. 95–112 in G. GRASSI and L. SABATH, eds. *New trends in antibiotics: research and therapy*. Elsevier/North-Holland, New York.
- . 1980b. Plasmid-mediated resistance to aminocyclitol antibiotics in group D streptococci. *J. Bacteriol.* **143**:541–551.
- COURVALIN, P., and M. FIANDT. 1980. Aminoglycoside-modifying enzymes of *Staphylococcus aureus*: expression in *Escherichia coli*. *Gene* **9**:247–269.
- COURVALIN, P., B. WEISBLUM, and J. DAVIES. 1977. Aminoglycoside-modifying enzyme of an antibiotic-producing bacterium acts as a determinant of antibiotic resistance in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **74**:999–1003.
- DAVIES, J., C. HOUK, M. YAGISAWA, and T. J. WHITE. 1979. Occurrence and function of aminoglycoside-modifying enzymes. Pp. 166–169 in *Genetics of industrial microorganisms—1978*. American Society of Microbiology, Washington, D.C.
- DAVIES, J., and D. I. SMITH. 1978. Plasmid-determined resistance to antimicrobial agents. *Annu. Rev. Microbiol.* **32**:469–518.
- FITCH, W. M. 1970. Distinguishing homologous from analogous proteins. *Syst. Zool.* **19**:99–113.
- . 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20**:406–416.
- FITCH, W. M., and E. MARGOLIASH. 1967. The construction of phylogenetic trees—a generally applicable method utilizing estimates of the mutation distance obtained from cytochrome *c* sequences. *Science* **155**:279–284.
- HORINOUCHI, S., and B. WEISBLUM. 1982a. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* **150**:804–814.
- . 1982b. Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibiotics. *J. Bacteriol.* **150**:815–825.

- MATSUHASHI, Y., M. YAGISAWA, S. KONDO, T. TAKEUCHI, and H. UMEZAWA. 1975. Aminoglycoside 3'-phosphotransferases I and II in *Pseudomonas aeruginosa*. J. Antibiot. (Tokyo) **28**:442-447.
- MAXAM, A., and W. GILBERT. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499-560.
- OKA, A., H. SUGISAKI, and M. TANAKAMI. 1981. Nucleotide sequence of the kanamycin resistance transposon Tn903. J. Mol. Biol. **147**:217-236.
- ROSENBERG, M., and D. COURT. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319-353.
- SHINE, J., and L. DALGARNO. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. **71**:1342-1346.
- SMITH, D. I. 1978. Purification and characterization of aminoglycoside-3'-phosphotransferases. Ph.D. thesis. University of Wisconsin—Madison.
- THOMPSON, C., and G. GRAY. 1983. Nucleotide sequence of a streptomycete aminoglycoside phosphotransferase gene and its relationship to phosphotransferases encoded by resistance plasmids. Proc. Natl. Acad. Sci. USA **80**:5190-5194.
- THOMPSON, C., J. M. WARD, and D. HOPWOOD. 1980. DNA cloning in *Streptomyces* resistance genes from antibiotic-producing species. Nature **286**:525-527.
- TU, C.-P., and S. COHEN. 1980. 3'-end labeling of DNA with [α - 32 P] cordycepin-5'-triphosphate. Gene **10**:177-183.

ROY J. BRITTEN, reviewing editor

Revision received September 2, 1983.