Leder, P. (1973) *Adv. Protein Chem. 27*, 213–242.
Lill, R., Robertson, J. M., & Wintermeyer, W. (1984) *Biochemistry 23*, 6710–6717.
Lührmann, R., Eckhardt, H., & Stöffler, G. (1979) *Nature (London) 280*, 423–425.
Modolell, J., Girbés, T., & Vázquez, D. (1975) *FEBS Lett. 60*, 109–113.
Nierhaus, K. H., & Rheinberger, H.-J. (1984) *Trends Biochem. Sci. (Pers. Ed.) 9*, 428–432.
Noll, H. (1966) *Science (Washington, D.C.) 151*, 1241–1245.
Paulsen, H., & Wintermeyer, W. (1986) *Biochemistry* (in press).
Paulsen, H., Robertson, J. M., & Wintermeyer, W. (1982) *Nucleic Acids Res. 10*, 2651–2663.
Paulsen, H., Robertson, J. M., & Wintermeyer, W. (1983) *J. Mol. Biol. 167*, 411–426.
Peters, M., & Yarus, M. (1979) *J. Mol. Biol. 134*, 471–491.
Record, M. T., Jr., Lohman, T. M., & deHaseth, P. L. (1976) *J. Mol. Biol. 107*, 145–158.
Record, M. T., Jr., Anderson, C. F., & Lohman, T. M. (1978) *Q. Rev. Biophys. 11*, 103–178.
Rheinberger, H.-J., & Nierhaus, K. H. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4213–4217.
Rheinberger, H.-J., Sternbach, H., & Nierhaus, K. H. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 5310–5314.
Robertson, J. M., & Wintermeyer, W. (1981) *J. Mol. Biol. 151*, 57–79.
Robertson, J. M., Kahan, M., Wintermeyer, W., & Zachau, H. G. (1977) *Eur. J. Biochem. 72*, 117–125.
Schmitt, M., Neugebauer, U., Bergmann, C., Gassen, H. G., & Riesner, D. (1982) *Eur. J. Biochem. 127*, 525–529.
Thompson, R. C., & Karim, A. M. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4922–4926.
Watanabe, S. (1972) *J. Mol. Biol. 67*, 443–457.
Wettstein, F. O., & Noll, H. (1965) *J. Mol. Biol. 11*, 35–53.
Winkler-Oswatitsch, R., & Eigen, M. (1979) *Angew. Chem. 91*, 20–51.
Wintermeyer, W., & Zachau, H. G. (1979) *Eur. J. Biochem. 98*, 465–475.
Wintermeyer, W., & Robertson, J. M. (1982) *Biochemistry 21*, 2246–2252.
Wintermeyer, W., Lill, R., Paulsen, H., & Robertson, J. M. (1984) in *Mechanisms of Protein Synthesis* (Bermek, E., Ed.) pp 76–89, Springer, Berlin and Heidelberg.
Wintermeyer, W., Lill, R., Paulsen, H., & Robertson, J. M. (1986) in *Structure, Function, and Genetics of Ribosomes* (Hardesty, B., & Kramer, G., Eds.) Springer, New York (in press).
Wurmbach, P., & Nierhaus, K. H. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2143–2147.

Biosynthetic *alr* Alanine Racemase from *Salmonella typhimurium*: DNA and Protein Sequence Determination†

Nicholas G. Galakatos, Elisabeth Daub, David Botstein, and Christopher T. Walsh*

Departments of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received October 15, 1985; Revised Manuscript Received January 13, 1986

ABSTRACT: The nucleotide sequence of the *alr* gene encoding the biosynthetic alanine racemase in *Salmonella typhimurium* is reported. The sequence was determined by the dideoxy chain termination method of Sanger mostly from recombinants derived from shotgun and specific subcloning of a 2.6-kilobase region containing the *alr* gene. The final bridging of nonoverlapping contiguous sequences was accomplished with the use of synthetic site-specific primers. The *alr* gene was found to be 1077 base pairs in length encoding a protein of 359 amino acid residues. Comparison of *alr* with the *dadB* gene encoding the catabolic alanine racemase in *S. typhimurium* revealed almost identical size (1077 vs. 1068 base pairs) and 52% sequence identity. The respective gene products displayed 43% homology, which includes a decapeptide bearing the pyridoxal 5′-phosphate binding site.

A key building block in the peptidoglycan layer of bacterial cell walls is D-alanine. The biosynthetic route to D-alanine is by racemization of L-alanine. Recent studies from these laboratories (Wasserman et al., 1983) demonstrated the presence of two genes, *dadB* and *alr*, from *Salmonella typhimurium*, both encoding alanine racemases.

The *dadB* gene encodes an alanine racemase that maps adjacent to the *dadA* gene, which in turn encodes a D-alanine dehydrogenase. The *dadB* and *dadA* gene products permit *Salmonella* to grow on L-alanine as a source of carbon and nitrogen, and thus the *dadB* alanine racemase may have primarily a catabolic function. The *dadB* gene has been sequenced, the encoded alanine racemase purified to homogeneity (Wasserman et al., 1984), and its molecular basis of susceptibility to β-haloalanine antibacterials determined (Badet et al., 1984).

The second alanine racemase gene in *S. typhimurium* is *alr*, which has now been mapped to minute 91 and isolated from a λ library (E. Daub et al., unpublished results). In the accompanying paper (Esaki & Walsh, 1986), we describe the purification of the cloned *alr*-encoded alanine racemase to homogeneity, its enzymatic characteristics and susceptibility to haloalanines, and the *N*-terminal and active site protein

† Financial support from the National Science Foundation (Grant PCM 8308969) is gratefully acknowledged.

* Address correspondence to this author at the Department of Chemistry.

0006-2960/86/0425-3255$01.50/0

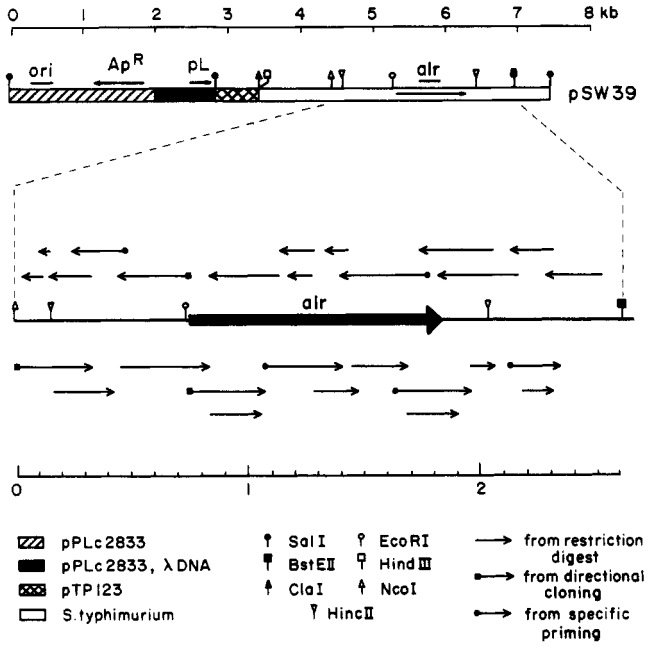

FIGURE 1: DNA sequencing scheme for the *alr* region of plasmid pSW39. The composition of pSW39 (Materials and Methods) is shown at the top. The isolated 2.6-kb *Nco*I–*Bst*EII fragment was subcloned into M13 vectors by shotgun or directional cloning, and the resulting single-stranded templates were sequenced by the dideoxy chain termination method with either the universal or site-specific sequencing primers.

sequences. In this paper, we report the nucleotide sequence of the *S. typhimurium alr* alanine racemase gene and the encoded polypeptide sequence. We also illustrate the correlations in primary structure between the *dadB* and *alr* genes and gene products.

Materials and Methods

Materials. Plasmids pTP123 and pPLc2833 were graciously provided by Drs. A. Poteete and W. Fiers, respectively. *Escherichia coli* strain JM101, M13mp8 RF DNA, M13mp9 RF DNA, the universal sequencing primer 5′-GTAAAACGACGGCCAGT-3′, deoxy- and dideoxynucleoside triphosphates, and deoxyadenosine 5′-(α-[^{35}S]thiotriphosphate) (>600 mCi/mmol) were purchased from Amersham. All restriction enzymes and T4 DNA ligase were from New England Biolabs. Calf intestinal alkaline phosphatase (25 units/μL) and the Klenow fragment of DNA polymerase I (5 units/μL) were from Boehringer-Mannheim.

Elutip-d prefilters and columns were obtained from Schleicher & Schuell. 5′-(Dimethoxytrityl)deoxynucleotide triphosphates, solid-phase resins, and reagents for oligonucleotide synthesis were from Biosearch. IPTG[1] and X-Gal were from Sigma.

Construction of Plasmid pSW39. The overproducing plasmid pSW39 is a derivative of pSW22, which has been previously described (Wasserman et al., 1983). Initially, a 4.0-kb *Cla*I–*Cla*I fragment of pSW22 was subcloned into pTP123, a pBR322 derivative, and this plasmid was designated pSW24. This construction contained a 4.6-kb *Sal*I–*Sal*I fragment coding for the *alr* alanine racemase, of which 0.62 kb was from pTP123 (Figure 1). The *Sal*I–*Sal*I fragment from pSW24 was then ligated into the *Sal*I site of pPLc2833, which is an expression vector containing the pL promoter of λ (Remaut et al., 1983). The ligation mixture was transformed into DB4949 (λ) (Wasserman et al., 1983), selecting ampicillin resistance at 41 °C. Restriction analysis allowed the selection of the desired structure for pSW39.

Table I: Sequences of Synthetic Primers

sequence	orientation/locus[a]
5′-TGTCTGCATACCGCCGT-3′	(+) 1027–1043
5′-TACTGGTCAATGGTCGT-3′	(+) 1610–1626
5′-CAAAGTGAATCTGAGCA-3′	(+) 2093–2109
5′-GTTCACCGGACGTTTAT-3′	(−) 481–497
5′-TAAGCACTTACTTTTGT-3′	(−) 1759–1775

[a] Orientation: 5′-*Nco*I–*Bst*EII-3′ = (+); reverse complement = (−). Locus: in bp measured in the (+) orientation from the *Nco*I site.

Subcloning of alr Alanine Racemase Gene. The 2.6-kb *Nco*I–*Bst*EII fragment of pSW39 containing the *alr* gene was isolated by agarose gel electrophoresis (1%) followed by electroelution and was purified by the Elutip-d procedure of Schleicher & Schuell. It was then subcloned into M13 RF vectors in two ways:

(a) The fragment was digested in separate reactions with *Hae*III, *Alu*I, *Hin*cII, and *Sau*3AI, and the mixture was extracted with phenol and precipitated with ethanol. The resulting restriction products were then treated with T4 DNA ligase in the presence of M13mp8, which had been linearized with *Sma*I or *Bam*HI (for the *Sau*3AI digest only) and dephosphorylated with CIP (Messing, 1983).

(b) The protruding ends of the 2.6-kb segment were filled in with the Klenow fragment of DNA polymerase I and deoxynucleotide triphosphates (Maniatis et al., 1982). The product was then digested with *Eco*RI, extracted with phenol, and precipitated with ethanol. The resulting mixture of two fragments (0.75 and 1.85 kb) was then directionally inserted in separate reactions and under standard conditions into M13mp8 and M13mp9 RF vectors that had been treated with *Eco*RI and *Sma*I and then with CIP (Messing & Vieira, 1982).

In all cases, the ligation mixture was used directly to transform competent *E. coli* strain JM101. For each ligation reaction, 36 colorless plaques were selected from YT plates containing 2% X-Gal and 0.6% IPTG (Messing, 1983).

DNA Sequence Analysis of alr Region. Following plaque purification, single-stranded DNA templates of the M13 clones were generated and sequenced by the chain termination method (Sanger et al., 1977) with the universal primer as the origin of chain elongation and deoxyadenosine 5′-(α-[^{35}S]-thiotriphosphate) as the label (Kunkel et al., 1981). To expedite the process, each set of single-stranded recombinants was prescreened, first by size on agarose gel electrophoresis (0.7%) and then by single-track Sanger sequence analysis (T screening).

The bridging of nonoverlapping contiguous regions of the sequence was accomplished by selective priming of the single-stranded recombinants generated by directional cloning. As an intrinsic control for specificity and purity, the primer sequences were designed to complement regions 25–50 bp upstream of the end of established sequences. Five such sequence-specific 17-mers (Table I) were synthesized as described below and were annealed to the templates under standard conditions (Messing, 1983) to allow sequencing.

Finally, for regions of high G,C content or apparent G stacking (see Computational Methods), the Sanger reaction was also run with deoxyinosine triphosphate in lieu of deoxyguanosine triphosphate (Mills & Kramer, 1979).

[1] Abbreviations: CIP, calf intestinal alkaline phosphatase; dNTP, deoxynucleotide triphosphate; IPTG, isopropyl β-D-thiogalactopyranoside; X-Gal, 5-bromo-4-chloro-3-indoyl β-D-galactopyranoside; DMT, 4′,4′-dimethoxytrityl; PLP, pyridoxal 5′-phosphate; SDS, sodium dodecyl sulfate; kb, kilobase; bp, base pair(s); Tris, tris(hydroxymethyl)aminomethane; EDTA, ethylenediaminetetraacetic acid.

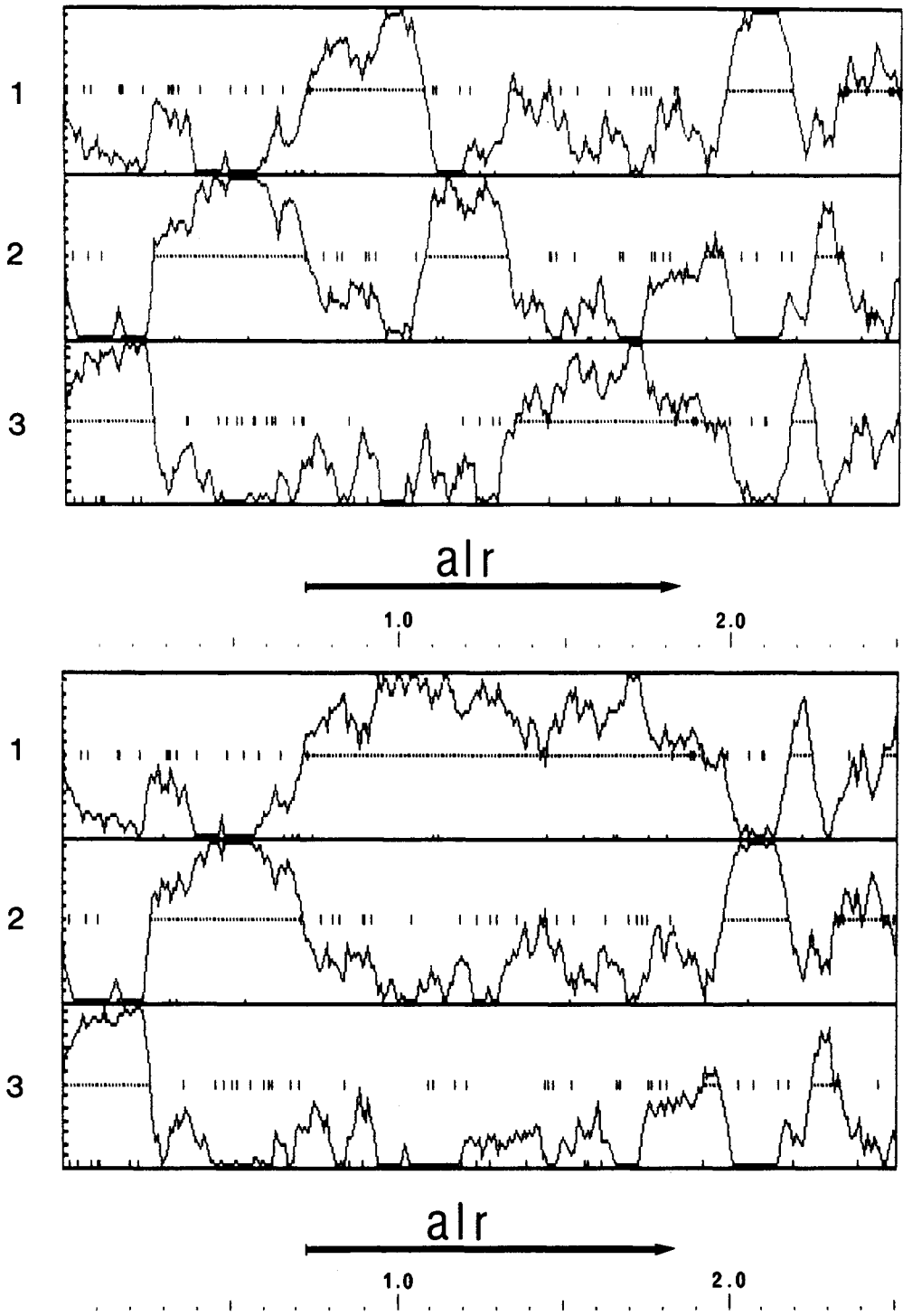

FIGURE 2: Identification of sequence artifacts by codon preference. *S. typhimurium* coding probability plots (window = 25 codons) of the three 5′–3′ reading frames of the 2.6-kb *Nco*I–*Bst*EII sequence. The dotted line across the half-point of each frame corresponds to 50% coding probability. Vertical bars on the dotted line and at the bottom of each frame represent stop and start codons, respectively. The length of the *alr* gene is shown above the size marker. (Top) Sequence containing two compressed-G regions; the artifact loci at 1.08 and 1.33 kb are evidenced by the sequential change in reading frames from 1 to 2 and 2 to 3, respectively. (Bottom) Corrected sequence. Following the resolution of the compressed-G regions with deoxyinosine triphosphate, a single open-reading frame covering the full length of *alr* appears at frame 1.

Synthesis of Sequencing Primers. Oligonucleotide synthesis was performed on a Biosearch Model Sam One DNA synthesizer by the phosphotriester method of Itakura (Tan et al., 1983). At the end of the solid-phase synthesis, the fully blocked, resin-bound oligomer was subjected to an additional DMT-deblocking cycle and was then treated successively with pyridine aldoxime (25 °C, 0.5 h) and 28% ammonium hydroxide (55 °C, 5 h) to afford phosphate deprotection and release from the support, respectively. Preparative polyacrylamide denaturing gel electrophoresis followed by elution of the isolated band with 0.1 M ammonium bicarbonate (37 °C, 24 h) yielded pure material that was desalted by passage through a G-50 Sephadex column preequilibrated with TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0).

In an alternative purification scheme, the final DMT-deblocking step was omitted, and the oligomer was purified by the method of Khorana (Lo et al., 1984). In our hands, the former protocol proved to be more convenient and efficient.

Computational Methods. The assembly of the sequence data generated from the M13 recombinants was performed

```
                -50                             -30                             -10
ATG CGG GAC CGC AAT ACG ATG ATG AGT AAC TCT CCG TCA TTC TTT TAA CAA GGA ATT CAA

                10                              30                              50
ATG CAA GCG GCA ACA GTC GTC ATT AAC CGC CGC GCT CTG CGA CAC AAC CTG CAA CGT CTG
Met Gln Ala Ala Thr Val Val Ile Asn Arg Arg Ala Leu Arg His Asn Leu Gln Arg Leu

                                                *
                70                              90                              110
CGT GAA CTG GCG CCT GCC AGT AAG CTG GTT GCG GTG GTG AAA GCG AAC GCT TAT GGA CAC
Arg Glu Leu Ala Pro Ala Ser Lys Leu Val Ala Val Val Lys Ala Asn Ala Tyr Gly His

                130                             150                             170
GGT CTT CTG GAG ACC GCG CGA ACG CTC CCT GAT GCT GAC GCT TTT GGC GTG GCG CGT CTT
Gly Leu Leu Glu Thr Ala Arg Thr Leu Pro Asp Ala Asp Ala Phe Gly Val Ala Arg Leu

                190                             210                             230
GAA GAG GCT CTA CGT CTG CGA GCG GGC GGG ATC ACG CAG CCA ATC CTG CTG CTG GAG GGT
Glu Glu Ala Leu Arg Leu Arg Ala Gly Gly Ile Thr Gln Pro Ile Leu Leu Leu Glu Gly

                250                             270                             290
TTT TTC GAC GCC GCC GAT TTG CCG ACC ATT TCC GCG CAA TGT CTG CAT ACC GCC GTA CAT
Phe Phe Asp Ala Ala Asp Leu Pro Thr Ile Ser Ala Gln Cys Leu His Thr Ala Val His

                310                             330                             350
AAT CAA GAG CAG CTT GCC GCC CTG GAG GCG GTG GAG CTG GCG GAG CCG GTA ACG GTC TGG
Asn Gln Glu Gln Leu Ala Ala Leu Glu Ala Val Glu Leu Ala Glu Pro Val Thr Val Trp

                370                             390                             410
ATG AAG CTG GAT ACC GGT ATG CAT CGT CTC GGC GTG CGT CCC GAA GAG GCG GAG GCG TTC
Met Lys Leu Asp Thr Gly Met His Arg Leu Gly Val Arg Pro Glu Glu Ala Glu Ala Phe

                430                             450                             470
TAC CAG CGT CTG ACG CAC TGT AAA AAT GTA CGC CAG CCG GTG AAT ATC GTC AGC CAT TTT
Tyr Gln Arg Leu Thr His Cys Lys Asn Val Arg Gln Pro Val Asn Ile Val Ser His Phe

                490                             510                             530
GCC CGT GCG GAT GAG CCG GAA TGC GGC GCT ACC GAA CAT CAG CTC GAC ATT TTT AAT GCC
Ala Arg Ala Asp Glu Pro Glu Cys Gly Ala Thr Glu His Gln Leu Asp Ile Phe Asn Ala

                550                             570                             590
TTC TGT CAG GGT AAA CCC GGT CAG CGC TCT ATT GCC GCG TCT GGC GGT ATC CTG CTG TGG
Phe Cys Gln Gly Lys Pro Gly Gln Arg Ser Ile Ala Ala Ser Gly Gly Ile Leu Leu Trp

                610                             630                             650
CCG CAG TCT CAC TTT GAC TGG GCG CGT CCG GGC ATC ATT TTG TAT GGC GTA TCG CCG CTG
Pro Gln Ser His Phe Asp Trp Ala Arg Pro Gly Ile Ile Leu Tyr Gly Val Ser Pro Leu

                670                             690                             710
GAG CAC AAA CCC TGG GGG CCG GAT TTT GGT TTT CAG CCG GTG ATG TCC TTA ACC TCC AGT
Glu His Lys Pro Trp Gly Pro Asp Phe Gly Phe Gln Pro Val Met Ser Leu Thr Ser Ser

                730                             750                             770
TTG ATC GCG GTG CGT GAC CAC AAA GCG GGC GAA CCG GTG GGC TAC GGC GGG ACA TGG GTG
Leu Ile Ala Val Arg Asp His Lys Ala Gly Glu Pro Val Gly Tyr Gly Gly Thr Trp Val

                790                             810                             830
AGT GAG CGC GAC ACG CGC CTG GGC GTG GTG GCG ATG GGT TAT GGC GAT GGC TAC CCA CGA
Ser Glu Arg Asp Thr Arg Leu Gly Val Val Ala Met Gly Tyr Gly Asp Gly Tyr Pro Arg

                850                             870                             890
GCG GCG CCT TCC GGT ACG CCA GTA CTG GTC AAT GGT CGT GAA GTT CCG ATT GTC GGG CGG
Ala Ala Pro Ser Gly Thr Pro Val Leu Val Asn Gly Arg Glu Val Pro Ile Val Gly Arg

                910                             930                             950
GTG GCG ATG GAT ATG ATT TGC GTA GAT TTG GGG CCA AAC GCG CAG GAT AAC GCG GGC GAT
Val Ala Met Asp Met Ile Cys Val Asp Leu Gly Pro Asn Ala Gln Asp Asn Ala Gly Asp

                970                             990                             1010
CCG GTG GTC TTA TGG GGT GAA GGT CTG CCG GTT GAA CGT ATC GCT GAA ATG ACA AAA GTA
Pro Val Val Leu Trp Gly Glu Gly Leu Pro Val Glu Arg Ile Ala Glu Met Thr Lys Val

                1030                            1050                            1070
AGT GCT TAC GAA CTT ATC ACG CGC CTG ACC TCA AGG GTG GCG ATG AAG TAT ATT GAT TAA
Ser Ala Tyr Glu Leu Ile Thr Arg Leu Thr Ser Arg Val Ala Met Lys Tyr Ile Asp ***

ATA CGC GGC GGG CCG GGT GGC GCT TGC GCT TAT CCG GCT TGT ATC GCG CAT TGG CTG TAG
```

FIGURE 3: DNA sequence and translation of *alr* gene. Coding region starts at position 1 and is terminated at position 1080. The two peptides for which sequences were determined from the purified gene product (Esaki & Walsh, 1986) are underlined; discrepancies between the predicted and observed residues are doubly underlined. The lysyl residue involved in the binding of PLP is indicated by an asterisk.

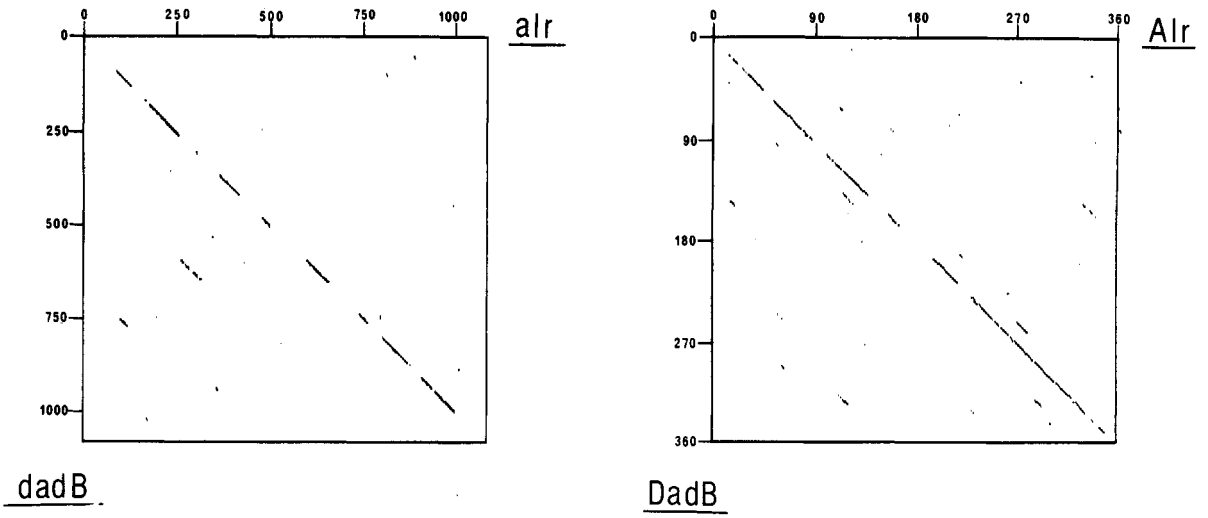

FIGURE 4: Diagonal dot matrix comparison of the gene and gene-product sequences of *alr* and *dadB*. The abscissa represents the primary sequence of *alr* while the ordinate corresponds to the sequence of *dadB*. (Left) Absolute homology of the gene sequences plotted on the basis of a unitary scoring matrix with a window of 45 nucleotides at 0.04% probability. (Right) Conservative homology between the two polypeptide chains calculated according to a mutation data scoring matrix and with a window of 15 amino acid residues at 0.25% probability.

```
                1         5         10        15        20        25        30        35        40        45        50        55
ALR             M - Q A A T V V I N R R A L R H N L Q R L R E L A P A S K L V A V V [K] A N A Y G H G L L E T A R T - L P D A D
DADB            M T R P I Q A S L D L Q V M K Q N L A I V R R A A P E A R V W S V V [K] A N A Y G H G I - E R V W S A L G A T D
Common          M                             N L       R     A P                 V V [K] A N A Y G H G       E           L       D

                56        60        65        70        75        80        85        90        95        100       105       110
ALR             A F G V A R L E E A L R L R A G G I T Q P I L L L E G F F D A A D L P T I S A Q C L H T A V H N Q E Q L A A L
DADB            G F A M L N L E E A I T L R E R G W K G P I L M L E G F F H A Q D L E A Y D T Y R L T T C I H S N W Q L K A L
Common            F       L E E A     L R     G       P I L   L E G F F   A   D L                 L   T     H     Q L   A L

                111       115       120       125       130       135       140       145       150       155       160       165
ALR             E A V E L A E P V T V W M K L D T G M H R L G V R P E E A E A F Y Q R L T H C K N V R Q P V N I V S H F A R A
DADB            Q N A R L N A P L D I Y V K V N S G M N R L G F Q P E R A Q T V W Q Q L R A M R N V G E - M T L M S H F A Q A
Common                L   P             K       G M   R L G     P E   A           Q   L           N V               S H F A   A

                166       170       175       180       185       190       195       200       205       210       215       220
ALR             D E P E - C G A T E H Q L D I F N A F C Q G K P G Q R S I A A S G G I L L W - P Q S H F D W A R P G I I L Y G
DADB            D H P E G I G E A M R R I A L A T - - - E G L Q C A Y S L S N S A A T L - W H P Q A H Y D W V R P G I I L Y G
Common          D   P E     G                     G               S       S       L   W   P Q   H   D W   R P G I I L Y G

                221       225       230       235       240       245       250       255       260       265       270       275
ALR             V S P L E H K P W G - - P D F G F Q P V M S L T S S L I A V R D H K A G E P V G Y G G T W V S E R D T R L G V
DADB            A S P S G Q - - W R D I A D T G L K P V M T L S S E I I G V Q T L S A G E R V G Y G G G Y S V T Q E Q R I G I
Common            S P         W       D   G     P V M   L   S       I   V           A G E   V G Y G G                 R   G

                276       280       285       290       295       300       305       310       315       320       325       330
ALR             V A M G Y G D G Y P R A A P S G T P V L V N G R E V P I V G R V A M D M I C V D L G P N A Q D N A G D P V V L
DADB            V A A G Y A D G Y P R H A P T G T P V L V D G I R T R T V G T V S M D M L A V D L T P C P Q A G I G T P V E L
Common          V A   G Y   D G Y P R   A P   G T P V L V   G           V G   V   M D M     V D L   P     Q     G   P V   L

                331       335       340       345       350       355       360       365       370
ALR             W G E G L P V E R I A E M T K V S A Y E L I T R L T S R V A M K Y I D -
DADB            W G K E I K V D D V A S A A G T L G Y E L L C A V A P R V P - - F V T T
Common          W G         V       A             Y E L             R V
```

FIGURE 5: Linear alignment of the protein sequences of *alr* and *dadB* alanine racemases. Identities were located by the method of Dayhoff (1983) and are shown below the two sequences. The active site decapeptide that contains the PLP-binding lysyl residue at position 35 is underlined.

with the aid of the DBCOMP, DBUTIL, and DBAUTO programs of Roger Staden (Staden, 1982a) modified by the MIT Seq Group. As shown in Figure 2, artifacts in the reading of sequencing gels (e.g., G stacking) were successfully identified by the use of a method based on codon preference as a means of predicting protein coding regions (Staden & McLachlan, 1982). Coding probabilities were calculated on the basis of a standard codon usage table constructed from the 6498-bp coding region of the *trp* operon of *S. typhimurium* (Crawford et al., 1980) and were plotted with the ANALYSEQ program (Staden, 1982a). The extent of linear alignment (Dayhoff et al., 1983) and diagonal conservative correlation (Staden, 1982b) of the Alr and DadB protein sequences was computed on the basis of mutation data scoring matrices; for the respective DNA calculations, a unitary scoring matrix was employed.

RESULTS

Subcloning and DNA Sequencing of alr Region. Dideoxy chain termination sequence analysis (Sanger et al., 1977) of the M13 recombinants from the *Hae*III, *Alu*I, *Hin*cII, and *Sau*3AI digests provided an initial set of seven nonoverlapping, contiguous sequences encompassing *ca.* 86% of the total length of the *Nco*I–*Bst*EII fragment. Following the sequence determination of the four directional clones, 95% of the total length was established over three regions. Also at that stage, the *alr* gene was located and oriented by means of its *N*-terminal and active site protein sequences (Esaki & Walsh, 1986). Finally, specific priming generated a continuous sequence and allowed the completion of the analysis in both directions.

Table II: Predicted and Observed *alr* Amino Acid Composition

amino acid	predicted[a]	observed[b]	amino acid	predicted[a]	observed[b]
Ala	42	46	Ile	16	16
Arg	25	25	Leu	36	28
Asn	10		Lys	9	13
Asx	27	29	Met	9	9
Asp	17		Phe	10	11
Cys	5	4	Pro	23	24
Gln	14		Ser	14	19
Glx	37	41	Thr	17	17
Glu	23		Trp	6	7
Gly	32	35	Tyr	8	7
His	11	10	Val	32	31

[a] Predicted values were derived from the translation of the *alr* gene sequence. [b] Observed values are as described in the following paper (Esaki & Walsh, 1986).

Amino Acid Sequence and Composition of Alr Protein. The coding region of the *alr* gene comprises 1077 bp encoding a protein of 359 amino acids (Figure 3). The predicted N-terminal and active site sequences are in excellent agreement with those determined by protein sequencing of the purified gene product (Figure 3). On the basis of the predicted amino acid sequence, the molecular weight of the Alr protein is 39071, which falls within 10% of the value determined by SDS disc gel electrophoresis (Esaki & Walsh, 1986).

Comparison of the experimentally determined (Esaki & Walsh, 1986) and predicted overall amino acid compositions of the Alr protein revealed good correlation for all residues except for Lys and Leu (Table II). The discrepancy may be due to impurities in the protein sample analyzed.

Comparison of alr and dadB Gene and Protein Sequences. Despite the fact that *alr* (minute 91) and *dadB* (minute 36) map at two distinctly different regions on the *S. typhimurium* chromosome (E. Daub et al., unpublished results), considerable homology is displayed at the DNA and the gene-product level: both genes are of very similar size (*alr*, 1077 bp; *dadB*, 1068 bp). Linear alignment of the two nucleotide sequences on the basis of a unitary scoring matrix (Dayhoff et al., 1983) revealed a 52% identity dispersed throughout the length of the genes. The diagonal representation of this correlation is illustrated in Figure 4.

As expected, similar homology is displayed by the deduced amino acid sequence of the Alr and DadB proteins. In addition to size parity (Alr, 359 amino acid residues; DadB, 356 amino acid residues), both enzymes bind the PLP cofactor at equivalently positioned lysine residues (Alr, ^{34}Lys; DadB,^{35}Lys) in the amino-proximal 10% of the protein. Linear alignment (Dayhoff et al., 1983) yielded 151 identities out of 349 possible residue matches, including the active site decapeptide and one octa-, one hexa-, and three pentapeptides (Figure 5). The use of diagonal matrices (Staden, 1982b) highlighted additional regions of conservative homology and indicated that the main region of difference is centered around the middle portion of the two primary sequences (Figure 4).

Discussion

Using the dideoxy chain termination method (Sanger et al., 1977), we have determined the DNA sequence of the *S. typhimurium* gene encoding the Alr alanine racemase. The Alr racemase is one of two isoenzymes responsible for D-alanine production in *S. typhimurium* and is believed to satisfy the biosynthetic requirements of the bacterium (Wasserman et al., 1983; E. Daub et al., unpublished results). For catabolic purposes, D-alanine is generated by the *dadB* racemase, a protein of very similar size as Alr (DadB, 356 amino acids; Alr, 359 amino acids) but of increased activity (Esaki & Walsh, 1986; k_{cat}(DadB)/k_{cat}(Alr) = 60/1) and distinctly different gene locus (E. Daub et al., unpublished results). We used the known DadB protein sequence (Wasserman et al., 1984) to compute the absolute and conservative homology between the two isozymes. We find 43% identity and much greater conservative correlation. Most striking is the fact that both protein sequences contain an identical active site decapeptide, which includes the lysyl PLP binding site.

This active site sequence is highly conserved in the broad-specificity alanine racemase from *Pseudomonas striata* (Roise et al., 1984) and in the Gram-positive *Bacillus stearothermophilus* alanine racemase (Badet et al., 1986) and thus appears to be a stable feature of PLP racemase active sites. The extensive homology throughout the primary structures of DadB and Alr disclosed by the diagonal plots suggests a strong evolutionary relationship and a gene duplication event in the origin of the two racemases. The *alr*-encoded alanine racemase appears to be the biosynthetically consequential racemase, and this structural information may aid in the design of novel antibacterial agents targeted against this crucial enzyme.

Acknowledgments

We thank Dr. Steve Wasserman for communication of unpublished results. We are grateful to Drs. Peter Schulz, Oliver Peoples, and Keith Vass for helpful suggestions.

Registry No. PLP, 54-47-7; alanine racemase, 9024-06-0; DNA (*Salmonella thyphimurium* alanine racemase gene *alr*), 101418-60-4; alanine racemase (*Salmonella typhimurium* reduced), 101418-61-5.

References

Badet, B., Roise, D., & Walsh, C. T. (1984) *Biochemistry 23*, 5188.

Badet, B., Inagaki, K., Soda, K., & Walsh, C. (1986) *Biochemistry* (fourth paper of four in this issue).

Crawford, I. P., Nichols, B. P., & Yanofsky, C. (1980) *J. Mol. Biol. 142*, 489.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524.

Esaki, N., & Walsh, C. T. (1986) *Biochemistry* (second paper of four in this issue).

Kunkel, T. A., Eckstein, F., Mildvan, A. S., Koplitz, P. M., & Loeb, L. A. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 6734.

Lo, K.-M., Jones, S., Hackett, N., & Khorana, G. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2285.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Messing, J. (1983) *Methods Enzymol. 101C*, 20.

Messing, J., & Vieira, J. (1982) *Gene 19*, 269.

Mills, D. R., & Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 2232.

Remaut, E., Tsao, H., & Fiers, W. (1983) *Gene 21*, 103.

Roise, D., Soda, K., Yagi, T., & Walsh, C. T. (1984) *Biochemistry 23*, 5195.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463.

Staden, R. (1982a) *Nucleic Acids Res. 10*, 4731.

Staden, R. (1982b) *Nucleic Acids Res. 10*, 2951.

Staden, R., & McLachlan, A. D. (1982) *Nucleic Acids Res. 10*, 141.

Tan, Z. K., Ikuta, S., Huang, T., Dugaiczyk, A., & Itakura, K. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 382.

Wasserman, S. A., Walsh, C. T., & Botstein, D. (1983) *J. Bacteriol. 153*, 1439.

Wasserman, S. A., Daub, E., Grisafi, P., Botstein, D., & Walsh, C. T. (1984) *Biochemistry 23*, 5182.