Thermostable Alanine Racemase from *Bacillus stearothermophilus*: DNA and Protein Sequence Determination and Secondary Structure Prediction†

Katsuyuki Tanizawa, Atsushi Ohshima,‡ Alfred Scheidegger,§ Kenji Inagaki,‖ Hidehiko Tanaka, and Kenji Soda*
Laboratory of Microbial Biochemistry, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611, Japan
Received June 30, 1987; Revised Manuscript Received October 8, 1987

ABSTRACT: The nucleotide sequence of the alanine racemase (EC 5.1.1.1) gene from a thermophile, *Bacillus stearothermophilus*, was determined by the dideoxy chain termination method with universal and synthetic site-specific primers. The amino acid sequence of the enzyme predicted from the nucleotide sequence was confirmed by peptide sequence information derived from the N-terminal amino acid residues and several tryptic fragments. The alanine racemase gene consists of 1158 base pairs encoding a protein of 386 amino acid residues; the molecular weight of the apoenzyme is estimated as 43 341. The racemase gene of *B. stearothermophilus* has a closely similar size (1158 vs 1167 base pairs) to that of the gene of a mesophile, *B. subtilis*, but shows a higher preference for codons ending in G or C. A comparison of the amino acid sequence with those of *Bacillus subtilis* and *Salmonella typhimurium dadB* and *alr* enzymes revealed overall sequence homologies of 31–54%, including an identical octapeptide bearing the pyridoxal 5′-phosphate binding site. Although the residues common in the four racemases are not continuously arrayed, these constitute distinct domains and their hydropathy profiles are very similar. The secondary structure of *B. stearothermophilus* alanine racemase was predicted from the results obtained by theoretical analysis and circular dichroism measurement.

Alanine racemase (EC 5.1.1.1) is a pyridoxal 5′-phosphate (pyridoxal-P) dependent enzyme catalyzing the interconversion of L- and D-alanine (Adams, 1976; Soda et al., 1986). The enzyme occurs widely in prokaryotes and is responsible for the formation of the D-enantiomer used in the construction of the peptidoglycan layer of bacterial cell walls. Recent studies from the laboratories of Walsh and his co-workers (Badet & Walsh, 1985; Badet et al., 1986) demonstrated that (1-aminoethyl)-phosphonate (Ala-P), an alanine analogue designed for the specific inactivation of alanine racemases, acts differentially on the enzymes from Gram-positive and -negative bacteria. Thus, alanine racemases from *Streptococcus faecalis* and *Bacillus stearothermophilus* are irreversibly inactivated in a time-dependent manner by D- and L-isomers of Ala-P (Badet et al., 1986), whereas those from *Escherichia coli* and *Salmonella typhimurium* are inhibited reversibly (Atherton et al., 1979).

To elucidate the molecular basis of the reaction and inactivation mechanisms of alanine racemases, further information on the molecular structure of these enzymes including the entire amino acid sequence and the three-dimensional structure is needed. The protein sequences of alanine racemases from *S. typhimurium* (Wasserman et al., 1984, for the *dadB* enzyme, and Galakatos et al., 1986, for the *alr* enzyme) and *Bacillus subtilis* (Ferrari et al., 1985) have been determined from the DNA sequences of their genes. We have cloned the alanine racemase gene of *B. stearothermophilus*, purified the enzyme to homogeneity, and characterized it enzymologically (Inagaki et al., 1986). In this paper, we describe the nucleotide sequence of the *B. stearothermophilus* alanine racemase gene and the encoded polypeptide sequence. We also compared sequences of four alanine racemases now available and predicted a possible secondary structure. These data would facilitate the fine structure analysis in the crystallographic study of the enzyme, which is currently in progress (Neidhart et al., 1987).

Materials and Methods

Materials. The plasmid pICR4 containing the alanine racemase and alanine dehydrogenase genes of *B. stearothermophilus* was prepared as described previously (Inagaki et al., 1986). *E. coli* strain JM109, M13mp18, and M13mp19 RF DNAs, the universal primers for sequencing with M13 RF vectors (M1, 5′-AGTCACGACGTTGTA-3′) and for supercoil sequencing with pBR322 (R, 5′-GTATCACGAGGCCCTT-3′; and S2, 5′-GTTGAAGGCT-3′), deoxy- and dideoxynucleoside triphosphates, restriction enzymes, T4 DNA ligase, and the Klenow fragment of DNA polymerase I (2 units/μL) were purchased from Takara Shuzo, Kyoto. Deoxycytidine 5′-(α-[^{32}P]phospho)triphosphate (>400 mCi/mmol) was obtained from Amersham; Nensorb 20 prefilters and columns were from Du Pont; and Seppak C_{18} columns were from Waters. 5′-(4′,4′-Dimethoxytrityl)-deoxynucleotide triphosphoramidites, solid-phase resins, and other chemicals for oligonucleotide synthesis were obtained from Applied Biosystems. Isopropyl β-D-thiogalactoside (IPTG) and 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal) were the products of Sigma.

Subcloning of the Alanine Racemase Gene. After the 4.3-kb *Sal*I fragment containing the alanine dehydrogenase gene was excised from pICR4, the remaining 7.7-kb fragment was religated to yield pICR401 containing the alanine racemase gene (see Figure 1). The 3.9-kb *Hin*dIII–*Sal*I fragment of pICR401 was isolated by agarose gel electrophoresis (0.7%) followed by electroelution and was purified by the Nensorb 20 procedure of Du Pont. Three DNA fragments (0.4, 1.5,

† This work was supported in part by Grant-in-Aid for Scientific Research 60760083 from the Ministry of Education, Science and Culture of Japan.
* Author to whom correspondence should be addressed.
‡ On study leave from the Central Research Institute of Takara Shuzo, Ohtsu, Japan.
§ Present address: International Research Laboratories, Ciba-Geigy (Japan), Takarazuka, Japan.
‖ Present address: Division of Biological Function and Genetic Resources Science, Faculty of Agriculture, Okayama University, Okayama 700, Japan.

0006-2960/88/0427-1311$01.50/0

Table I: Sequences of Synthetic Primers

no.	sequence	orientation/ locus[a]
I_1 (17-mer)	5′-GCAAACAGCTGCCCGCT-3′	(−) 366–382
I_2 (17-mer)	5′-CCCGGTTGTTGGCTCAG-3′	(+) 1110–1126
T_1 (17-mer)	5′-TCCGCAGTCGCAAAATG-3′	(−) 2218–2234
T_2 (18-mer)	5′-CATGAAGTGGAAGCAATG-3′	(+) 1481–1498
T_3 (17-mer)	5′-GCGATGAAAGTCGTTCA-3′	(−) 2709–2725

[a] Orientation: 5′-*Hin*dIII–*Sal*I-3′ = (+); reverse complement = (−). Locus: in base pairs measured in the (−) orientation from the *Pst*I site at about 250 bp apart from the *Sal*I site (see Figure 1).

and 1.9 kb) were then obtained by complete digestion of the 3.9-kb fragment with *Pst*I. These fragments were inserted by T4 DNA ligase in separate reactions and under standard conditions into M13mp18 or M13mp19 RF vectors that had been treated with *Pst*I (for the 0.4- and 1.5-kb fragments) or with *Pst*I and *Hin*dIII (for the 1.9-kb fragment) (Messing & Vieira, 1982). The ligated RF DNAs were used directly to transform competent *E. coli* strain JM109. In each transfection, more than 10 colorless plaques were selected from YT plates containing 2% X-Gal and 0.6% IPTG (Messing, 1983).

DNA Sequencing. 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 deoxycytidine 5′-(α-[^{32}P]phospho)triphosphate as the label. To expedite the process, each set of single-stranded recombinants was prescreened by size on agarose gel electrophoresis (0.7%).

The sequencing of extended regions and the bridging of nonoverlapping contiguous regions of the sequence were accomplished by priming of the single-stranded recombinants with the synthetic primers and by supercoil sequencing of the double-stranded plasmid with pBR322 R and S2 and synthetic primers (Wallace et al., 1981). 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 primers (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 GC content or apparent G stacking, the Sanger reaction was also run with the 2′-deoxy-7-deazaguanosine triphosphate sequencing kit (Takara Shuzo, Kyoto) instead of deoxyguanosine triphosphate (Mizusawa et al., 1986).

Synthesis of Sequencing Primers. Oligonucleotide synthesis was performed on an Applied Biosystems DNA synthesizer 381A by the phosphoramidite method (Broido et al., 1984). At the end of the solid-phase synthesis, the fully blocked, resin-bound oligomer was subjected to an additional dimethoxytrityl-deblocking cycle and was then treated successively with thiophenol (25 °C, 1 h) and 28% ammonium hydroxide (55 °C, 5 h) to afford phosphate deprotection and release from the support, respectively. The synthesized products were desalted through Seppak C_{18} columns following the manufacturer's directions.

Trypsin Digestion and Isolation of Tryptic Fragments. Alanine racemase was purified to homogeneity from cell extracts of *E. coli* C600 carrying the plasmid pICR4 as described previously (Inagaki et al., 1986). After precipitation of the enzyme (about 3 mg) with 5% trichloroacetic acid and washing of the precipitate with water, tryptic digestion was performed at 37 °C for 12 h in 0.1 M ammonium bicarbonate (pH 8.5) with 0.03 mg (1:100 w/w) of bovine pancreas trypsin (TPCK-treated, Cooper). The clear solution obtained was acidified to stop digestion by addition of trichloroacetic acid (final concentration 1%) and dried by centrifugal evaporation. The tryptic peptides were separated with an ULTRON N-C_{18} reversed-phase column (4.6 × 150 mm, Shinwa Kako Inc., Kyoto, Japan) on an LKB Ultrochrome GTi HPLC system (Bromma, Sweden). A 40-min linear gradient from 0 to 56% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid was used to elute peptides at a flow rate of 0.6 mL/min. Elution of peptides was monitored by absorbance at 220 nm. Five peptides that were well separated from other small peptides were further purified by HPLC under the chromatographic conditions similar to those above except that 0.1% (v/v) heptafluorobutyric acid was used instead of 0.1% trifluoroacetic acid and a 30-min linear gradient from 16 to 48% acetonitrile was employed.

Amino Acid Sequencing. The purified alanine racemase (about 2 nmol in 1 mM potassium phosphate buffer, pH 7.2) and the tryptic peptides isolated (0.6–1.5 nmol of T-1 to T-5) were subjected directly to automated Edman degradation with an Applied Biosystems 470A gas–liquid-phase protein sequencer (Strickler et al., 1984). The phenylthiohydantoin (PTH) amino acid derivatives were separated and identified with an on-line PTH analyzer Model 120A (Applied Biosystems) with a PTH-C_{18} column. The initial and averaged repetitive yields of the Edman degradation were about 42–75 and 97%, respectively.

Circular Dichroism Analysis. CD measurements were carried out with a Jasco J-600 recording spectropolarimeter at 25 °C with a 1-mm light path length cell. The instrument was calibrated with (+)-10-camphorsulfonic acid, $\Delta\epsilon$ = +2.37 M^{-1} cm^{-1} at 290.5 nm. In the calculation of the mean residue ellipticity, $[\theta]$, the mean residue weight was taken to be 112.0 for the enzyme protein. The CD spectra were obtained at protein concentrations of 0.02–0.3 mg/mL in the far-UV region (180–240 nm) under nitrogen atmosphere.

Results and Discussion

Subcloning and DNA Sequencing of the Alanine Racemase Gene. The original plasmid pICR4 (12.0 kb) that we constructed (Inagaki et al., 1986) contained the genes of both alanine racemase and alanine dehydrogenase of *B. stearothermophilus* for convenience in assaying the racemase. Therefore, the 4.3-kb *Sal*I fragment containing the dehydrogenase gene was excised from pICR4, and the remaining 7.7-kb fragment was religated to yield pICR401 containing only the alanine racemase gene (Figure 1). Dideoxy chain termination sequence analysis (Sanger et al., 1977) was performed in both strands of DNA with the M13 recombinants from the *Pst*I digests (0.4, 1.5, and 1.9 kb) of the 3.9-kb *Hin*dIII–*Sal*I fragment of pICR401. The entire continuous sequence of the 3.9-kb fragment was then obtained by specific priming of the *Pst*I fragments subcloned into M13 vectors with synthesized oligonucleotides (Table I) and by supercoil sequencing of pICR401 (Figure 1). Artifacts in the reading of sequencing gels (e.g., G stacking) were successfully identified by the use of 2′-deoxy-7-deazaguanosine triphosphate instead of deoxyguanosine triphosphate in the Sanger reaction. The alanine racemase gene was located and oriented by means of the N-terminal amino acid sequence of the enzyme determined by automated Edman degradation. The complete sequence of the racemase gene is shown along with the translated protein sequence in Figure 2.

Structure of the Alanine Racemase Gene. The entire DNA sequence of the 3.9-kb fragment contains an open reading frame of 1158 nucleotides starting with initiation codon ATG

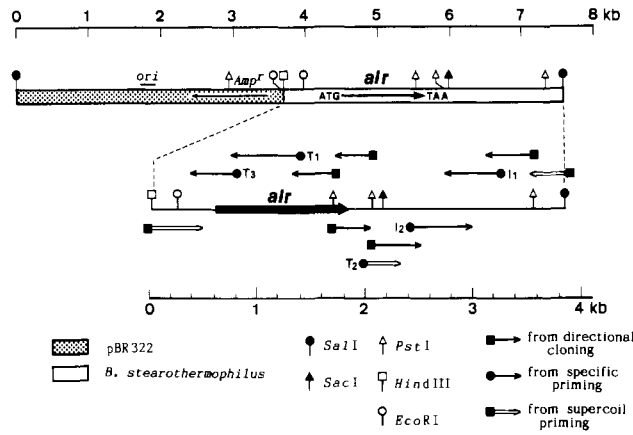

FIGURE 1: Restriction map and DNA sequencing scheme for the 3.9-kb *Hin*dIII–*Sal*I fragment of pICR401. The composition of pICR401 (Materials and Methods) is shown at the top. The isolated 3.9-kb *Hin*dIII–*Sal*I fragment was subcloned into M13 vectors by directional cloning, and the resulting single-stranded templates were sequenced by the dideoxy chain termination method with either the universal or site-specific (see Table I) sequencing primers. Three regions were also determined by the supercoil sequencing of pICR401.

and ending with termination codon TAA at position 1159. The ATG codon was chosen as the translation initiation site because of its location close to the possible ribosome binding site (Shine & Dalgarno, 1974; Gold et al., 1981). Thus, seven bases upstream from the ATG codon there is a 10-base sequence AAGGCAGCGA (−16 to −7) that is considerably complementary with the 3′-end of 16S rRNA (Shine & Dalgarno, 1974). The A+T content of the region upstream of the initiation codon is 57 mol %, which is higher than those in the total *B. stearothermophilus* chromosomal DNA (42–48 mol %) and in the reading frame of the racemase gene (44 mol %). This region contains a putative promoter (Figure 2) that displays some sequence homology to the canonical *E. coli* promoters TTGACA (−35) and TATAAT (−10) (Pribnow, 1975). Downstream from the TAA stop codon there is a G+C-rich region of dyad symmetry capable of forming a stem and loop structure. However, the sequence is not followed by a stretch of T residues, unlike the *E. coli* ρ-independent transcription terminators (Rosenberg & Court, 1979).

The G+C content of the coding region for alanine racemase is 56.2 mol %. This value is within the range of the genomic G+C content of *B. stearothermophilus* (52–58 mol %) and is higher than that (47.8 mol %) of the alanine racemase gene from mesophile *B. subtilis* (Ferrari et al., 1985). In particular, the *B. stearothermophilus* gene shows a high preference for G or C residues at the third base (the wobble position) of the codons; the G+C content at that position is 63.5 mol %, whereas it is 50.7 mol % in the *B. subtilis* gene. Changes from A/T to G/C in DNA sequences, particularly in the wobble position of each codon, are thought to be one of the mechansims of gene stabilization at high temperatures (Kagawa et al., 1984).

Amino Acid Sequence and Composition of Alanine Racemase. The alanine racemase gene encodes a protein of 386 amino acids. The amino acid sequences of the N-terminal region and of five peptide fragments obtained by trypsin digestion of the purified gene product and subsequent peptide sequencing are in excellent agreement with those predicted from the DNA sequence (Figure 2). The two residues (-Val-Val-) adjacent to the lysyl residue involved in the binding of pyridoxal-P (see below), however, are not identical with those (-Pro-Pro-) determined previously (Badet et al., 1986). Since the residue of cycle B (Pro) of the octapeptide containing

```
-120                                                                -61
TCA.CTG.TCT.TTG.TAC.GCT.TTC.CCC.TAT.TCC.TTC.TTC.CGC.AAT.GGC.ATG.GTA.AAC.TAA.AAA
        -35                                                 -10
-60                                                                 -1
AGG.ACA.AAA.ATG.ATT.TCG.AGC.GCT.TTC.AAG.TTT.GCG.CAA.CAG.GAA.AGG.CAG.CGA.GAA.ACC
                                                        SD
 1                                                                  60
ATG.AAC.GAC.TTT.CAT.CGC.GAT.ACG.TGG.GCG.GAA.GTG.GAT.TTG.GAC.GCC.ATT.TAC.GAC.AAT
MET-Asn-Asp-Phe-His-Arg-Asp-Thr-Trp-Ala-Glu-Val-Asp-Leu-Asp-Ala-Ile-Tyr-Asp-Asn
61                                                                 *  120
GTG.GAG.AAT.TTG.CGC.CGT.TTG.CTG.CCG.GAC.GAC.ACG.CAC.ATT.ATG.GCG.GTC.GTG.AAA.GCG
Val-Glu-Asn-Leu-Arg-Arg-Leu-Leu-Pro-Asp-Asp-Thr-His-Ile-Met-Ala-Val-Val-Lys-Ala
121                                                                180
AAC.GCC.TAT.GGA.CAT.GGG.GAT.GTG.CAG.GTG.GCA.AGG.ACA.GCG.CTC.GAA.CGG.GGG.CCT.CCG
Asn-Ala-Tyr-Gly-His-Gly-Asp-Val-Gln-Val-Ala-Arg-Thr-Ala-Leu-Glu-Arg-Gly-Pro-Pro
181                                                                240
CCT.GCG.GTT.GCC.TTT.TTG.GAT.GAG.GCG.CTC.GCT.TTA.AGG.GAA.AAA.GGA.ATC.GAA.GCG.CCG
Pro-Ala-Val-Ala-Phe-Leu-Asp-Glu-Ala-Leu-Ala-Leu-Arg-Glu-Lys-Gly-Ile-Glu-Ala-Pro
241                                                                300
ATT.CTA.GTT.CTC.GGG.GCT.TCC.CGT.CCA.GCT.GAT.GCG.GCG.CTG.GCC.GCC.CAG.CAG.CGC.ATT
Ile-Leu-Val-Leu-Gly-Ala-Ser-Arg-Pro-Ala-Asp-Ala-Ala-Leu-Ala-Ala-Gln-Gln-Arg-Ile
301                                                                360
GCC.CTG.ACC.GTG.TTC.CGC.TCC.GAC.TGG.TTG.GAA.GAA.GCG.TCC.GCC.CTT.TAC.AGC.GGC.CCT
Ala-Leu-Thr-Val-Phe-Arg-Ser-Asp-Trp-Leu-Glu-Glu-Ala-Ser-Ala-Leu-Tyr-Ser-Gly-Pro
                        T-1
361                                                                420
TTT.CCT.ATT.CAT.TTC.CAT.TTG.AAA.ATG.GAC.ACC.GGC.ATG.GGA.CGG.CTT.GGA.GTG.AAA.GAC
Phe-Pro-Ile-His-Phe-His-Leu-Lys-Met-Asp-Thr-Gly-Met-Gly-Arg-Leu-Gly-Val-Lys-Asp
421                                                                480
GAG.GAA.GAG.ACG.AAA.CGA.ATC.GTA.GCG.CTG.ATT.GAG.CGC.CAT.CCG.CAT.TTT.GTG.CTT.GAA
Glu-Glu-Glu-Thr-Lys-Arg-Ile-Val-Ala-Leu-Ile-Glu-Arg-His-Pro-His-Phe-Val-Leu-Glu
481                                                                540
GGG.TTG.TAC.ACG.CAT.TTT.GCG.ACT.GCG.GAT.GAG.GTG.AAC.ACC.GAT.TAT.TTT.TCC.TAT.CAG
Gly-Leu-Tyr-Thr-His-Phe-Ala-Thr-Ala-Asp-Glu-Val-Asn-Thr-Asp-Tyr-Phe-Ser-Tyr-Gln
541                                                                600
TAT.ACC.CGT.TTT.TTG.CAC.ATG.CTC.GAA.TGG.CTG.CCG.TCG.CGC.CCG.CCG.CTC.GTC.CAT.TGC
Tyr-Thr-Arg-Phe-Leu-His-Met-Leu-Glu-Trp-Leu-Pro-Ser-Arg-Pro-Pro-Leu-Val-His-Cys
            T-2
601                                                                660
GCC.AAC.AGC.GCA.GCG.TCG.CTC.CGT.TTC.CCT.GAC.CGG.ACG.TTC.AAT.ATG.GTC.CGC.TTC.GGC
Ala-Asn-Ser-Ala-Ala-Ser-Leu-Arg-Phe-Pro-Asp-Arg-Thr-Phe-Asn-Met-Val-Arg-Phe-Gly
                                                        T-4
661                                                                720
ATT.GCC.ATG.TAT.GGG.CTT.GCC.CCG.TCG.CCC.GGC.ATC.AAG.CCG.CTG.CTG.CCG.TAT.CCA.TTA
Ile-Ala-Met-Tyr-Gly-Leu-Ala-Pro-Ser-Pro-Gly-Ile-Lys-Pro-Leu-Leu-Pro-Tyr-Pro-Leu
721                                                                780
AAA.GAA.GCA.TTT.TCG.CTC.CAT.AGC.CGC.CTC.GTA.CAC.GTC.AAA.AAA.CTG.CAA.CCA.GGC.GAA
Lys-Glu-Ala-Phe-Ser-Leu-His-Ser-Arg-Leu-Val-His-Val-Lys-Lys-Leu-Gln-Pro-Gly-Glu
781                                                                840
AAG.GTG.AGC.TAT.GGT.GCG.ACG.TAC.ACT.GCG.CAG.ACG.GAG.GAG.TGG.ATC.GGG.ACG.ATT.CCG
Lys-Val-Ser-Tyr-Gly-Ala-Thr-Tyr-Thr-Ala-Gln-Thr-Glu-Glu-Trp-Ile-Gly-Thr-Ile-Pro
                                T-5
841                                                                900
ATC.GGC.TAT.GCG.GAC.GGC.GTC.CGC.CGC.CTG.CAG.CAC.TTT.CAT.GTC.CTT.GTT.GAC.GGA.CAA
Ile-Gly-Tyr-Ala-Asp-Gly-Val-Arg-Arg-Leu-Gln-His-Phe-His-Val-Leu-Val-Asp-Gly-Gln
901                                                                960
AAG.GCG.CCG.ATT.GTC.GGC.CGC.ATT.TGC.ATG.GAC.CAG.TGC.ATG.ATC.CGC.CTG.CCT.GGT.CCG
Lys-Ala-Pro-Ile-Val-Gly-Arg-Ile-Cys-Met-Asp-Gln-Cys-Met-Ile-Arg-Leu-Pro-Gly-Pro
961                                                                1020
CTG.CCG.GTC.GGC.ACG.AAG.GTG.ACA.CTG.ATT.GGT.CGC.CAA.GGG.GAC.GAG.GTA.ATT.TCC.ATT
Leu-Pro-Val-Gly-Thr-Lys-Val-Thr-Leu-Ile-Gly-Arg-Gln-Gly-Asp-Glu-Val-Ile-Ser-Ile
                                T-3
1021                                                               1080
GAT.GAT.GTC.GCT.CGC.CAT.TTG.GAA.ACG.ATC.AAC.TAC.GAA.GTG.CCT.TGC.ACG.ATC.AGT.TAT
Asp-Asp-Val-Ala-Arg-His-Leu-Glu-Thr-Ile-Asn-Tyr-Glu-Val-Pro-Cys-Thr-Ile-Ser-Tyr
1081                                                               1140
CGA.GTG.CCC.CGT.ATT.TTT.TTC.CGC.CAT.AAG.CGT.ATA.ATG.GAA.GTG.AGA.AAC.GCC.ATT.GGC
Arg-Val-Pro-Arg-Ile-Phe-Phe-Arg-His-Lys-Arg-Ile-Met-Glu-Val-Arg-Asn-Ala-Ile-Gly
1141                                1158                           1200
CGC.GGG.GAA.AGC.AGT.GCA.TAA.TCT.GCT.TCA.AAA.GAC.AAA.GGA.TAG.AAA.ATG.GAG.AAA.TGA
Arg-Gly-Glu-Ser-Ser-Ala-***
1201                                                               1260
GCT.TTG.CAG.CGC.CGT.TGT.TTA.ATG.GTA.TCA.TGA.AGT.GGA.AGC.AAT.GAA.TGG.CTG.CAG.TTC
```

FIGURE 2: DNA sequence and translation of the alanine racemase gene. Coding region starts at position 1 and is terminated at position 1159. The six peptides (25 N-terminal residues and 5 tryptic fragments, some of which ended at Tyr due to the contaminated chymotryptic activity) for which sequences were determined from the purified gene product are underlined. The lysyl residue involved in the binding of pyridoxal-P is indicated by an asterisk. The −35 and −10 regions of the putative promoter and the proposed Shine–Dalgarno sequence (SD) for the ribosome binding site are also underlined. In the 3′-flanking region from the coding sequence, a sequence capable of forming a stem and loop structure, which may be involved in transcription termination, is indicated by arrows.

the active-site lysyl residue had been identified with some uncertainty (Badet et al., 1986), we assume that the -Val-Val-Lys- sequence predicted from the DNA sequence is plausible. On the basis of the predicted amino acid sequence, the molecular weight of alanine racemase is 43 341, which corresponds with the value (39 000) determined by sodium dodecyl sulfate gel electrophoresis (Inagaki et al., 1986).

Comparison of the experimentally determined (Inagaki et al., 1986) and predicted overall amino acid compositions of the racemase protein revealed good correlation for all residues except Trp (Table II). The discrepancy may be due to inaccurate spectrophotometric determination of the Trp content. By a repeated determination using the method of hydrolysis in 4 N methanesulfonic acid (Simpson et al., 1976), we obtained an amino acid composition much closer to the predicted one not only in Trp content but also in contents of other residues (see Table II).

Sequence Comparison of Four Alanine Racemases. The complete amino acid sequences of alanine racemase from four

Table II: Amino Acid Composition of Alanine Racemase from *B. stearothermophilus*[a]

amino acid	predicted	obsd H^b	obsd M^c	amino acid	predicted	obsd H^b	obsd M^c
Ala	37	38	39	Ile	24	19	24
Arg	30	27	28	Leu	37	34	40
Asn	9			Lys	13	13	14
Asx	31	32	31	Met	10	8	10
Asp	22			Phe	16	15	17
Cys	4	2	3	Pro	26	22	25
Gln	10			Ser	16	18	14
Glx	34	33	29	Thr	19	22	19
Glu	24			Trp	4	16	3
Gly	26	25	27	Tyr	14	11	14
His	16	14	16	Val	29	25	29

[a] The number of amino acid residues is presented as mole per mole of subunit. Predicted values were derived from the translation of the alanine racemase gene sequence. [b] Observed values (H) were recalculated from the results obtained by hydrolysis in 6 N HCl (Inagaki et al., 1986) with the corrected subunit molecular weight (43 341). [c] Obtained by hydrolysis in 4 N methanesulfonic acid containing 0.2% 3-(2-aminoethyl)indole (Pierce) at 115 °C for 22 h (M).

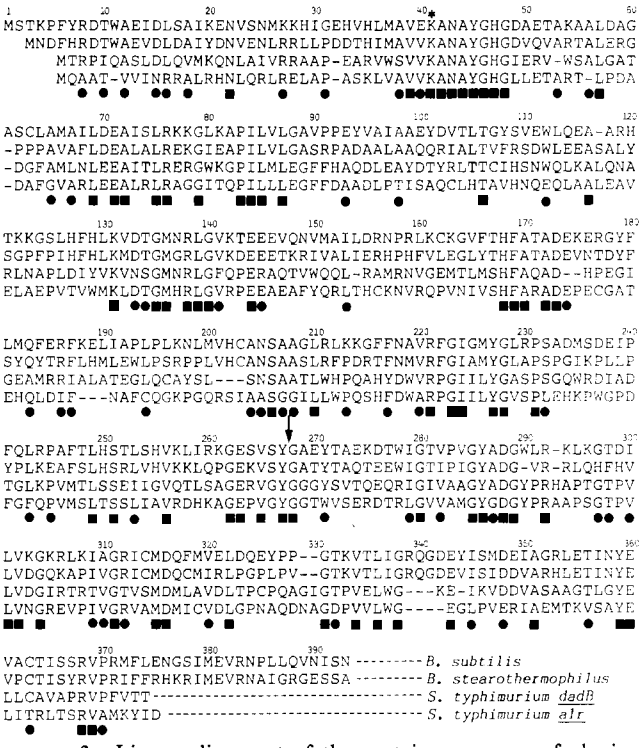

FIGURE 3: Linear alignment of the protein sequences of alanine racemases from *B. subtilis*, *B. stearothermophilus*, *S. typhimurium dadB*, and *S. typhimurium alr*. The four sequences were aligned by introducing gaps (hyphens) to maximize identities according to the method of Dayhoff et al. (1983). The residue numbers are given on the basis of the common sequence. Identical residues among the four (■) and three (●) sequences are shown below. The active-site lysyl residue is indicated by an asterisk. The vertical arrow shows the position where the limited proteolysis occurs.

different sources [*B. stearothermophilus*, this paper; *B. subtilis*, Ferrari et al. (1985); *S. typhimurium*, the *dadB* gene, Wasserman et al. (1984); and *S. typhimurium*, the *alr* gene, Galakatos et al. (1986)] are now available, and the four sequences were linearly aligned essentially according to a mutation data scoring matrix (Dayhoff et al., 1983) as shown in Figure 3. Despite the fact that *Bacillus* and *Salmonella* belong to Gram-positive and Gram-negative bacteria, respectively, and that *dadB* (minute 36) and *alr* (minute 91) genes map at two distinctly different regions on the *S. typhimurium* chromosome (Galakatos et al., 1986), the four

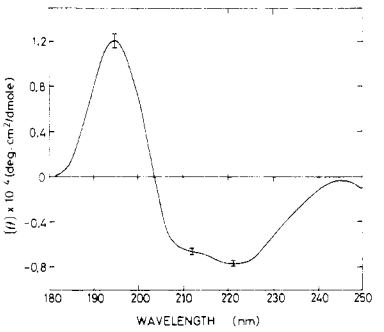

FIGURE 4: Far-ultraviolet CD spectrum of *B. stearothermophilus* alanine racemase (0.02–0.3 mg/mL 10 mM potassium phosphate buffer, pH 7.2, at 25 °C).

racemase sequences display considerable homology; a total of 74 residues matches in the four and a total of 53 residues matches in the three sequences compared (about 33% homology on average of the four sequences). The sequence homologies between the two racemases are calculated as 54% (*B. stearothermophilus:B. subtilis*), 43% (*dadB:alr*), 35% (*B. stearothermophilus:dadB*), 34% (*B. subtilis:dadB*), 31% (*B. stearothermophilus:alr*), and 30% (*B. subtilis:alr*). The longest continuous sequence of identical residues is an octapeptide containing the active-site lysyl residue that binds the cofactor pyridoxal-P. The other identical residues are not so much continuously arrayed, but with the three-residue matches these constitute obviously at least 9–10 regions distributed evenly in the entire sequence (see Figure 3). Each of these regions is characteristically rich in residues with similar chemical properties such as basic amino acids, hydrophobic amino acids, and sulfur amino acids. Although no residues essential for the enzyme catalysis except for the pyridoxal-P binding lysyl residue have been identified, those that participate in binding of a substrate and the cofactor may be found in residues which are identical among the four sequences. Limited proteolysis of the enzyme occurs also in one of the common sequences (-^{264}Tyr-^{265}Gly- of the *B. stearothermophilus* sequence; Galakatos et al., unpublished results), suggesting that the protein of the four racemases has a common structure composed of large and small domains connected by the hinge sequence.

Secondary Structure Prediction. The far-ultraviolet CD spectrum depicted in Figure 4 is an average of five measurements with the purified alanine racemase from *B. stearothermophilus*. The spectrum is characterized by two negative minima around 210 and 223 nm and a positive peak at 195 nm. The α-helix and β-structure contents of alanine racemase were calculated to be about 34 and 26%, respectively, from CD data by the least-squares method using ellipticity values of 19 points in the region between 190 and 243 nm (Chen et al., 1974).

Hydropathy profiles, which are of value in predicting secondary structural features, of the four alanine racemase sequences are compared in Figure 5. In contrast to the rather low sequence homology (30–54%), the hydropathy profiles of the four enzymes are very similar with each other, suggesting that they have essentially the same secondary structures. The alanine racemase from *B. stearothermophilus* is more thermostable than the other three enzymes (Inagaki et al., 1986). This distinctive thermostability presumably reflects the structural differences in nonhomologous regions of the proteins. One of the properties that might affect the stability of protein molecules is the hydropathy of each stretch of the amino acid sequence. Although the profiles are similar as a whole, several portions in the *B. stearothermophilus* enzyme sequence, which

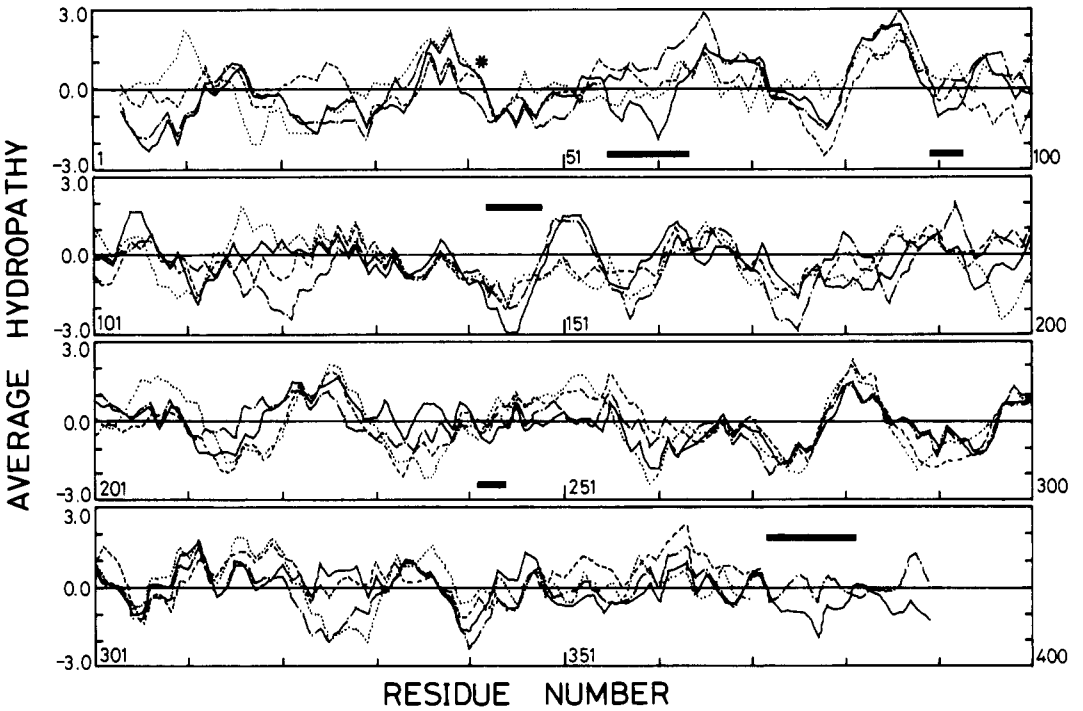

FIGURE 5: Comparison of hydropathy profiles of the four alanine racemases. Consecutive hydropathy averages are plotted for a seven-residue window advancing from the N- to the C-terminus. Relative hydrophilicity and hydrophobicity were recorded in the range +3.0 to −3.0 (Kyte & Loolittle, 1982) for each of the four sequences (—, *B. stearothermophilus* enzyme; ·–·, *B. subtilis* enzyme; - - -, *S. typhimurium dadB* enzyme; and ···, the *S. typhimurium alr* enzyme), which had been aligned by introducing gaps to maximize identities (Figure 3). The position of the active-site lysyl residue (asterisk) and eminently hydrophobic regions in the *B. stearothermophilus* enzyme (heavy lines) are indicated.

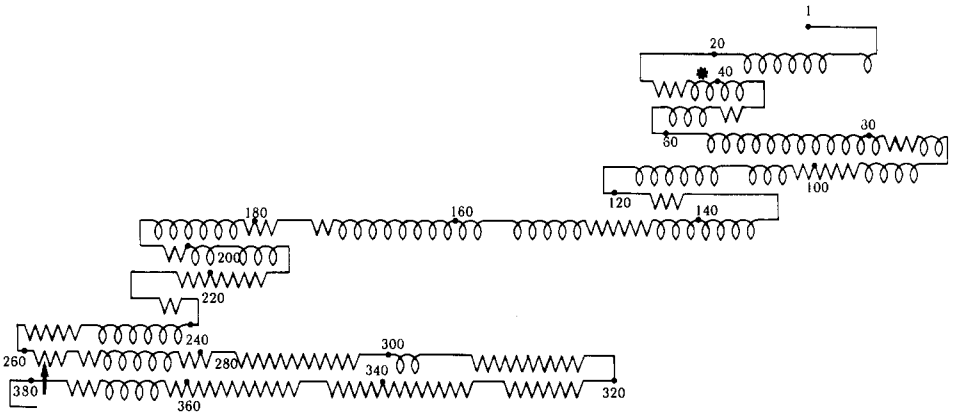

FIGURE 6: Schematic diagram of the predicted secondary structure in alanine racemase. α-Helices (coiled), β-strands (jagged), and β-turns are from the Chou and Fasman method (1978). The lysyl residue that binds pyridoxal-P (asterisk) and the probable proteolysis site (arrow) are indicated.

correspond to nonhomologous regions, are clearly more hydrophobic than the other three enzyme sequences (indicated by heavy lines in Figure 5).

Prediction of secondary structure by theoretical methods (Chou & Fasman, 1978; Garnier et al., 1978) also indicated that the four alanine racemase sequences display similar profiles of the α-helix and β-structure distribution in the entire sequence (data not shown). Figure 6 shows a schematic representation of the predicted secondary structure of the *B. stearothermophilus* alanine racemase. The contents of α-helix and β-sheet structure were estimated as about 30 and 35%, respectively. Similar results were obtained by using the method of Garnier et al. (1978). It is noteworthy that the helical content in one-third region from the C-terminus (residues 277–379) is extremely low (see Figure 6), whereas two-thirds from the N-terminus is a mixture of α-helices and β-strands, which occur alternately as observed in α/β type proteins (Levitt & Chothia, 1976). This suggests that the enzyme protein is composed of two structurally dissimilar domains connected by a short polypeptide (residues 258–266), which corresponds to a region that suffers from the limited proteolysis (see above). The observation that the enzyme exists as a stable intermediate during denaturation with guanidine hydrochloride (at about 4.2 M; Tanizawa et al., unpublished results) also supports this proposal.

ACKNOWLEDGMENTS

We thank Dr. N. G. Galakatos and Prof. C. T. Walsh for providing the unpublished information on limited proteolysis.

Registry No. DNA (*Bacillus stearothermophilus* alanine racemase gene derived), 112347-40-7; alanine racemase (*Bacillus stearothermophilus* reduced), 112347-41-8; alanine racemase, 9024-06-0.

REFERENCES

Adams, E. (1976) *Adv. Enzymol. Relat. Areas Mol. Biol. 44*, 69–138.

Atherton, F. R., Hall, M. J., Hassall, C. H., Lambert, R. W., Lloyd, W. J., & Ringrose, P. S. (1979) *Antimicrob. Agents Chemother. 15*, 696-705.
Badet, B., & Walsh, C. (1985) *Biochemistry 24*, 1333-1341.
Badet, B., Inagaki, K., Soda, K., & Walsh, C. T. (1986) *Biochemistry 25*, 3275-3282.
Broido, M. S., Zon, G., & James, T. L. (1984) *Biochem. Biophys. Res. Commun. 119*, 663-670.
Chen, Y.-H., Yang, J. T., & Chau, K. H. (1974) *Biochemistry 13*, 3350-3359.
Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45-148.
Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524-545.
Ferrari, E., Henner, D. J., & Yang, M. Y. (1985) *Biotechnology 3*, 1003-1007.
Galakatos, N. G., Daub, E., Botstein, D., & Walsh, C. T. (1986) *Biochemistry 25*, 3255-3260.
Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97-120.
Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S., & Stormo, G. (1981) *Annu. Rev. Microbiol. 5*, 365-403.
Inagaki, K., Tanizawa, K., Badet, B., Walsh, C. T., Tanaka, H., & Soda, K. (1986) *Biochemistry 25*, 3268-3274.
Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T., & Oshima, T. (1984) *J. Biol. Chem. 259*, 2956-2960.
Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105-132.
Levitt, M., & Chothia, C. (1976) *Nature (London) 261*, 552-557.
Messing, J. (1983) *Methods Enzymol. 101*, 20-78.
Messing, J., & Vieira, J. (1982) *Gene 19*, 269-276.
Mizusawa S., Nishimura, S., & Seela, F. (1986) *Nucleic Acids Res. 14*, 1319-1324.
Neidhart, D. J., Distefano, M. D., Tanizawa, K., Soda, K., Walsh, C. T., & Petsko, G. A. (1987) *J. Biol. Chem. 262*, 15323-15326.
Pribnow, D. (1975) *J. Mol. Biol. 99*, 419-443.
Rosenberg, M., & Court, P. (1979) *Annu. Rev. Genet. 13*, 319-353.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Shine, J., & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1342-1346.
Simpson, R. J., Neuberger, M. R., & Liu, T.-Y. (1976) *J. Biol. Chem. 251*, 1936-1940.
Soda, K., Tanaka, H., & Tanizawa, K. (1986) in *Vitamin B_6: Pyridoxal Phosphate: Chemical, Biochemical and Medical Aspects* (Dolphin, D., Poulson, R., & Avramović, O., Eds.) Vol. 1B, pp 223-251, Wiley, New York.
Strickler, J. E., Hunkapiller, M. W., & Wilson, K. J. (1984) *Anal. Biochem. 140*, 553-566.
Wallace, R. B., Johnson, M. J., Suggs, S. V., Miyoshi, K., & Itakura, K. (1981) *Gene 16*, 21-26.
Wasserman, S. A., Daub, E., Grisafi, P., Botstein, D., & Walsh, C. T. (1984) *Biochemistry 23*, 5182-5187.

Structure of Phenylalanine-Accepting Transfer Ribonucleic Acid and of Its Environment in Aqueous Solvents with Different Salts

Giuseppe Zaccai* and Su-Ya Xian‡
Institut Laue-Langevin, 156X, 38042 Grenoble Cedex, France
Received January 28, 1987; Revised Manuscript Received October 1, 1987

ABSTRACT: Yeast $tRNA^{Phe}$ was studied in different salt-containing solvents by UV absorbance and small-angle neutron scattering (SANS). This extends results obtained previously in NaCl and KCl solutions [Li, Z.-Q., Giegé, R., Jacrot, B., Oberthür, R., Thierry, J. C., & Zaccai, G. (1983) *Biochemistry 22*, 4380-4388]. As expected, at low concentrations of all salts studied, the tRNA molecule is unfolded. The importance of specific counterion interactions and the flexibility of the macromolecule are emphasized by the observation that it cannot take up its folded structure in $N(CH_3)_4Cl$ solvents, even when that salt concentration is increased to 1 M, in the absence of Mg ions. In CsCl solvents, on the other hand, the folded conformation is obtained in salt concentrations above about 0.2 M, similar to NaCl or KCl. By a comparison of SANS results in CsCl H_2O and CsCl ^2H_2O solvents with the data from NaCl and KCl solvents, thermodynamic and structural parameters were derived for the solvated macromolecule. All the data are accounted for, quantitatively, by a model for the particle in NaCl, KCl, or CsCl solution made up of $tRNA^{76-}$, closely associated with 76 positive hydrated counterions, surrounded by an aqueous solvent layer that excludes salt (and, therefore, of density different from that of bulk solvent). The mass of water in that layer depends on salt concentration, and the values found are consistent with those predicted by the Donnan effect.

Specific effects of salts on nucleic acid interactions have been known and studied for several decades [e.g., Chargaff et al. (1953), Latt and Sober (1967), review by Von Hippel and Schleich (1969), and Li et al. (1983)]. They remain incompletely understood, however, even though their importance is firmly established. Latt and Sober (1967a,b) studied the binding of synthetic oligopeptides to synthetic double-stranded polynucleotides as a function of salt and showed that there were specific effects associated with different cations. They interpreted their results in terms of competition between cation and oligopeptide for nucleic acid binding. Von Hippel and Schleich (1969) reviewed studies of salts and macromolecules

* Author to whom correspondence should be addressed.
‡ Present address: Department of Cell Biology, Beijing Medical University, Beijing, China.

0006-2960/88/0427-1316$01.50/0