Characterization of the Genes Encoding D-Amino Acid Transaminase and Glutamate Racemase, Two D-Glutamate Biosynthetic Enzymes of Bacillus sphaericus ATCC 10208

IAN G. FOTHERINGHAM,* STEFAN A. BLEDIG, AND PAUL P. TAYLOR

Biosciences Laboratory, NSC Technologies, Mt. Prospect, Illinois 60056

Received 3 April 1998/Accepted 27 May 1998

In Bacillus sphaericus and other Bacillus spp., D-amino acid transaminase has been considered solely responsible for biosynthesis of D-glutamate, an essential component of cell wall peptidoglycan, in contrast to the glutamate racemase employed by many other bacteria. We report here the cloning of the dat gene encoding D-amino acid transaminase and the glr gene encoding a glutamate racemase from B. sphaericus ATCC 10208. The glr gene encodes a 28.8-kDa protein with 40 to 50% sequence identity to the glutamate racemases of Lactobacillus, Pediococcus, and Staphylococcus species. The dat gene encodes a 31.4-kDa peptide with 67% primary sequence homology to the D-amino acid transaminase of the thermophilic Bacillus sp. strain YM1.

Bacteria normally synthesize D-glutamate and D-alanine as essential components of cell wall peptidoglycan (24). In addition, certain species such as the bacilli synthesize a number of other D-amino acids, including D-phenylalnine, D-asparagine, and D-ornithine, as intermediates in the production of secondary metabolites such as peptide-based antibiotics (2, 13, 36). D-Alanine biosynthesis from L-alanine by alanine racemase appears to be ubiquitous in bacteria (1, 34), but two distinct enzymatic routes have been identified for bacterial D-glutamate biosynthesis. In organisms such as Escherichia coli, Lactobacillus spp., and Pediococcus spp., specific racemases are responsible for the direct biosynthesis of D-glutamate and D-alanine from their L-amino acids (6, 9, 20). Conversely, Bacillus spp., such as Bacillus sphaericus and Bacillus licheniformis, possess a D-amino acid transaminase capable of synthesizing D-glutamate among a broad range of D-amino acids from keto acid precursors, using D-alanine as the amino donor (13, 35). The D-amino acid transaminases of B. sphaericus ATCC 10208 and Bacillus sp. strain YM1 have been particularly well studied (18, 27, 31, 35). The bacilli have generally been considered to synthesize only D-alanine through the action of a racemase, with D-glutamate biosynthesis attributed to the D-amino acid transaminase using α-ketoglutarate and D-alanine as substrates (5, 20), although a strain of Bacillus pumilus which lacks D-amino acid transaminase does possess glutamate racemase activity (14). Only one organism, Staphylococcus haemolyticus, has been shown to possess both D-amino acid transaminase and glutamate racemase activities (22).

We have cloned and characterized genes encoding two distinct D-amino acid biosynthetic enzymes from B. sphaericus ATCC 10208 by using a complementation screen with E. coli WM335 (15). This mutant, dependent on exogenously supplied D-glutamate has been used to previously isolate genes encoding D-amino acid transaminase and glutamate racemase (14, 22, 32). Enzymatic assays indicate that the genes isolated in this work encode the D-amino acid transaminase purified from this strain earlier and a previously undetected glutamate racemase. The presence of the two enzymes in B. sphaericus sug-

* Corresponding author. Mailing address: Biosciences Laboratory, NSC Technologies, 601 East Kensington Rd., Mt. Prospect, IL 60056. Phone: (847) 506-2839. Fax: (847) 506-4270. E-mail: igfoth@ccmail.monsanto.com.
bacillus fermentus (37%), Lactobacillus brevis (40%), and to a lesser extent E. coli (26%). Strong sequence identity (50 and 30%, respectively) was also observed between B. sphaericus glutamate racemase and the putative products of the racE (GenBank accession no. Z75208) and yrpC (GenBank accession no. U93875) genes identified in the B. subtilis genomic sequence. The sequence alignment is shown in Fig. 1. The peptide also displayed a low level of primary sequence identity (21%) to the aspartate racemases of Streptococcus thermophilus and Desulfurococcaceae sp., but no significant identity was observed with alanine racemase of B. subtilis or Bacillus stearothermophilus or with B. subtilis amino acid racemase (GenBank accession no. Z94943). Identification of the ATG start codon of the protein was strongly supported by the sequence homology between the protein’s N terminus and the other known glutamate racemase sequences, the presence of a consensus Bacillus ribosome binding site GAGG 7 bp upstream of the ATG codon, and the absence of upstream ATG, TTG, and GTG start codons capable of initiating the same open reading frame. Downstream of the coding sequence, there is a region of strong dyad symmetry typical of a bacterial transcription terminator.

Glutamate racemase activity of pIF1001. Cultures of strains to be assayed were inoculated from single colonies into 10 ml of LB and grown with shaking at 37°C. These were used to inoculate 50-ml LB cultures which were grown to an initial optical density at 600 nm of 0.05. Cells for glutamate racemase assay were washed, resuspended in 50 mM Tris HCl buffer (pH 8.0), and lysed in a French pressure cell at 1,000 lb/in². Supernatant fluid was recovered for the assay following centrifugation at 14,000 × g for 30 min at 4°C. Glutamate racemase activity was assayed by an N-glutamate dehydrogenase-coupled assay as described previously (22). Assay results were corrected for nonspecific reduction of NADP, which was determined from control assays carried out in the absence of D-glutamate. One unit is defined as 1 μmol of NADP reduced per min at 37°C. The results are...
Table 1. D-Amino acid transaminase and glutamate racemase activity of dat and glr clones

<table>
<thead>
<tr>
<th>Strain</th>
<th>D-Amino acid transaminase sp act (U/mg of protein)</th>
<th>Glutamate racemase sp act (U/mg of protein)</th>
<th>Complementation of WM355</th>
</tr>
</thead>
<tbody>
<tr>
<td>WM355/pIF306</td>
<td>0.0080</td>
<td>0.007</td>
<td>-</td>
</tr>
<tr>
<td>WM355/pIFF1001 (glr)</td>
<td>0.0065</td>
<td>0.025</td>
<td>+</td>
</tr>
<tr>
<td>WM355/pIF1002 (dat)</td>
<td>7.09</td>
<td>0.005</td>
<td>+</td>
</tr>
<tr>
<td>WM355/pIF372 (dat)</td>
<td>0.3</td>
<td>ND*</td>
<td>+</td>
</tr>
</tbody>
</table>

* Results represent averages of three assays corrected for endogenous rates.
+ One unit of activity defined as 1 mmol of pyruvate (NADH) produced per min at 37°C.
+ One unit of activity defined as 1 mmol of NADPH produced per min at 37°C.
* Complementation being normal overnight growth on LB agar. Symbols ‘-’, no growth; ‘+’, growth.

measured in Table 1. Measurements of glutamate racemase activity in pIF1001-bearing cells were consistently three to four times higher than those of the control cells (WM335 carrying pIF306), while the activity of WM335 carrying pIF1002 was not higher than that of the control. The protein encoded on pIF1001 was concluded to be a glutamate racemase of B. sphaericus ATCC 10208, and the gene was designated glr. The glr gene does not appear to be part of an operon, as there are no significant open reading frames upstream or downstream of the gene. Moreover, glr is oriented on pIF1001 such that expression is convergent with the upstream pheA-derived promoter region. This suggests that the native glr promoter is present on the fragment cloned in pIF1001, although we have not yet identified the transcription start site in B. sphaericus. Analysis of the DNA sequence upstream of the coding region did not reveal sequences with significant homology to the consensus Bacillus σ+ -35 (TTGACA) and -10 (TATAAT) regions for several B. sphaericus genes expressed during vegetative growth (4, 33). The codon usage of glr is generally consistent with that of moderately expressed E. coli genes (7).

Characterization of the B. sphaericus dat gene. Plasmid pIF1002 contains a 2.5-kb insert, and sequence analysis revealed an open reading frame of 848 bp encoding a protein of 31,392 Da (GenBank accession no. U26732). The N-terminal coding sequence was subsequently determined by using additional primers and was found to agree entirely with that of the purified enzyme (data not shown). This confirmed that the gene cloned in pIF1002 encodes d-amino acid transaminase, and we have named it dat. The C-terminal amino acid sequence of the B. sphaericus ATCC 10208 d-amino acid transaminase has been previously reported to consist of the peptide sequence VI (FY)LAL-COOH (5). The VI(FY)LAL peptide does not occur anywhere in the peptide sequence, and the origin of this stretch of sequence is unknown. The coding sequence is preceded by a strong B. subtilis ribosome binding site (AAAGGA) located 9 bp upstream of the ATG start codon. The gene is followed by a region of extensive dyad symmetry typical of bacterial transcription terminators.

D-Amino acid transaminase activity of pIF1002. Cells for d-amino acid transaminase assay were prepared as described above for the glutamate racemase assay but were washed and resuspended in 1 ml of 50 mM potassium phosphate buffer (pH 8.5). Extracts of E. coli WM335 carrying pIF1002 were assayed for D-amino acid transaminase activity by the lactate dehydrogenase-coupled assay as described previously (11). Assay results were corrected for nonspecific oxidation of NADH, which was determined by assaying in the absence of D-alanine. One unit of activity is defined as that which produces 1 μmol of pyruvate (NAD) per min at 37°C. The results shown in Table 1 indicated levels of D-alanine-dependent NADH oxidation approximately 1,000-fold higher than those of control cultures carrying only the vector pIF306. Similar assays conducted with WM335 transformed with pIF1001 had shown no increase in D-alanine-dependent NADH oxidation over the levels in the controls. The specific activity of the purified B. sphaericus D-amino acid transaminase has been reported to be 116 to 160 U/mg, indicating that the WM335/pIF1002 cells express Dat to approximately 5% of releasable cell protein. The high level of dat expression may be due in part to the colinear orientation of the dat coding sequence with the constitutive pheA promoter located upstream on the vector. To determine if transcription occurred in the absence of a heterologous promoter, the dat gene and 400 bp of upstream untranslated sequence were isolated by PCR and subcloned onto the low-copy-number vector pLG338 (28) between the unique SalI and EcoRI sites. Oligonucleotides used to isolate the B. sphaericus dat gene by PCR had the sequences 5'-CATGTCGAGCTTAATCCAGCGT TAGC-3' and 5'-GACGAAATCTTTTAGTGGCTTTTT AATC-3'. The resulting plasmid, pIF372, was able to confer wild-type growth upon WM335 in the absence of exogenous D-glutamate and showed D-amino acid transaminase activity of 0.3 U/mg. These results indicated that dat expression could originate from the 400-bp upstream sequence. However, neither this transcription start site nor that used in B. sphaericus has been identified.

Primary sequence conservation among known d-amino acid transaminase enzymes. Genes encoding d-amino acid transaminases have previously been isolated from the thermophilic Bacillus sp. strain YM1 (30) and from S. haemolyticus (22). In addition, we have previously identified and characterized the dat gene of B. licheniformis (32) (GenBank accession no. U26947). The sequence identity between the enzymes encoded by these genes and the d-amino acid transaminase of B. sphaericus was examined as shown in Fig. 2. The B. sphaericus d-amino acid transaminase shows 67% identity to that of the thermophilic Bacillus sp. strain YM1, 48% identity to the D-amino acid transaminase of S. haemolyticus, and 42% identity to the B. licheniformis d-amino acid transaminase. The sequence also shows 42% identity to the putative D-amino acid transaminase encoded by the B. subtilis yheM gene (GenBank accession no. Y14082) identified by the sequencing of the genome. The crystal structure of the YM1 enzyme has been determined to 1.9-A resolution, identifying a novel enzyme fold for a transaminase and enabling the assignment of residues believed to be important to the catalytic mechanism (29). Alignment of the primary sequence of B. sphaericus dat and the other d-amino acid transaminases with those of the YM1 enzyme and the other d-amino acid transaminases not surprisingly shows complete conservation of K145 (important for substrate cofactor proton transfer) and residues Y31, R50, E177, I204, and T205, which are all involved in cofactor ion pairing. There is also partial conservation of the serine residues S179 to S181, which contribute to cofactor positioning. Interestingly, the region from Ser240 to Ser243, which has been suggested to be important in substrate specificity through side chain discrimination, shows sequence variations in the known d-amino acid transaminase enzymes.

Conclusions. In this report we describe the isolation from B. sphaericus ATCC 10208 of the dat gene encoding the d-amino acid transaminase and a second gene we have designated glr, which encodes a hitherto unknown glutamate racemase. This is the first clear report of a Bacillus strain possessing both D-glutamate biosynthetic enzymes, although the genomic sequence of B. subtilis indicates that it may also possess both...
enzymes. The _B. sphaericus_ enzymes display high primary sequence identity to other known D-amino acid transaminases and glutamate racemases. The physiological significance of the utilization of two independent types of D-amino acid biosynthetic enzymes and the relative contribution and regulation of the two _B. sphaericus_ enzymes in D-glutamate biosynthesis remain undetermined. Pucci et al. (22) have demonstrated that the D-amino acid transaminase and glutamate racemase of _S. haemolyticus_ can each complement the D-glutamate auxotrophy of _E. coli_ WM335 when these genes are present on low-copy-number plasmid vectors. Similarly, we have observed complementation of WM335 by both _B. sphaericus_ genes on the low-copy-number plasmid vector pLG338, suggesting that either enzyme may be able to synthesize sufficient D-glutamate to sustain cell growth. Definitive experiments to disable each gene independently in the chromosome of _B. sphaericus_ are required to determine the respective roles of the enzymes in D-glutamate biosynthesis. It is interesting to speculate that the glutamate racemase, by analogy to the alanine racemase, may be sufficient to provide the necessary D-glutamate for peptidoglycan synthesis, whereas the D-amino acid transaminase is required to provide a broader range of D-amino acids necessary for secondary metabolite biosynthesis. This would be consistent with the narrow substrate specificity typically displayed by glutamate and alanine racemases in contrast to the much broader substrate profile of the D-amino acid transaminases. Bacilli are known to produce a number of antibiotics, such as bacillomycins and bacitracins, which incorporate a variety of D-amino acids, such as D-phenylalanine, D-tyrosine, and D-asparagine (2, 13, 36).

The function of D-amino acid transaminase in the biosynthesis of D-glutamic acid (18) has been well documented, as has the ability of the enzyme to synthesize a wide range of D-amino acids from keto acid substrates (35). D-Amino acid transaminases from _B. subtilis_ and _B. licheniformis_ as well as the enzymes from _Bacillus_ sp. strain YM1 and _B. sphaericus_ have been shown to display diverse but distinct substrate specificities. The recent determination of the YM1 enzyme backbone fold and the identity of residues participating in the catalytic mechanism has provided a clearer understanding of the effects of the site-directed enzyme mutants created earlier (17, 19, 29). The isolation of the _B. sphaericus_ dat gene now facilitates structure-function analyses between two well characterized and highly homologous _Bacillus_ D-amino acid transaminases and potentiates three-dimensional molecular modeling studies. Molecular modeling of the _B. sphaericus_ D-amino acid transaminase sequence on the YM1 enzyme backbone may permit mutagenesis studies to probe differences in the properties of the enzymes. The 67% exact sequence identity between the YM1 and _B. sphaericus_ D-amino acid transaminases is considerably greater than the 43% identity used successfully to conduct earlier studies using _E. coli_ L-amino acid transaminases in which molecular models of the aromatic transaminase encoded by the _tyrB_ gene (8) were derived from the known three-dimensional structure of the aspartate transaminase (26) encoded by _aspC_. Site-directed mutagenesis studies of _aspC_ and _tyrB_ were then used to test predictions regarding the residues influencing the substrate specificity of those enzymes towards aromatic and dicarboxylic substrates (12).

FIG. 2. Primary sequence homology between the D-amino acid transaminase, other known microbial D-amino acid transaminases, and the gene product of the _B. subtilis_ ythB gene. The conserved sequence shows the amino acids that are conserved in three or more sequences. Individual percentages for sequence conservation with _B. sphaericus_ D-amino acid transaminase are as follows: _Bacillus_ sp. strain YM1, 67%; _S. haemolyticus_, 48%; _B. licheniformis_, 42%; and _B. subtilis_ ythB, 42%. Asterisks indicate residues assigned active functions in the YM1 enzyme (29): substrate-cofactor proton transfer (K145), cofactor ion pairing (Y31, R50, E177, S179 to S181, I204, and T205), and substrate side chain recognition (S240 to S243).
strain YM1 and B. sphaericus. Amino acids, such as D-methionine, D-phenylalanine, and D-norleucine, which are good substrates for the B. sphaericus enzyme, are poor substrates for the YM1 enzyme (31). Many of the D-amino acid transaminase residues implicated in the active-site architecture of the YM1 D-amino acid transaminase are conserved in the B. sphaericus D-amino acid transaminase, but there are differences in the Ser240-to-Ser243 region proposed as part of a side chain binding pocket for D-amino acid substrate side chains. It will be interesting to explore this observation and to probe the differences in D-amino acid transaminase substrate specificities through mutagenesis selection procedures, site-directed mutagenesis, and gene shuffling experiments.

We thank Anna Kootstra for excellent technical assistance and Dave Ager and David Pantalone for critical appraisal of this manuscript.

We are grateful to Monsanto Corporation for providing excellent research facilities for this work.

**REFERENCES**