Nucleotide Sequence of the Overlapping Genes for the Subunits of Bacillus subtilis Aspartokinase II and Their Control Regions*

(Received for publication, February 18, 1987)

Nai-Yong Chen, Fu-Mei Hu, and Henry Paulus

From the Department of Metabolic Regulation, Boston Biomedical Research Institute, Boston, Massachusetts 02114 and the Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts 02115

The nucleotide sequence of a 2.9-kilobase Bacillus subtilis DNA fragment containing the entire coding region of aspartokinase II and adjacent chromosomal regions (Bondaryk, R., and Paulus, H. (1985a) J. Biol. Chem. 260, 585–591) has been determined. The results confirm the earlier prediction that the two subunits of aspartokinase II,  and , are encoded by in-phase overlapping genes. The nucleotide sequence showed strong ribosome binding sites before the translation initiation codons of the and subunits. Deletion of most of the coding region unique to the subunit had no effect on the synthesis of the smaller subunit, demonstrating that the subunit is indeed the product of independent translation. The site of transcription initiation of the aspartokinase genes was found to be located 300 nucleotides upstream from the translation start of the subunit. The intervening region contained a short reading frame capable of encoding a 24-residue lysine-rich polypeptide, which overlaps a region of extensive dyad symmetry culminating in a -independent transcription terminator. This region may be an attenuator control element that regulates the expression of the aspartokinase gene in response to the availability of lysine, the end product of the pathway. The coding sequence of the aspartokinase II subunits was immediately followed by a -independent transcription terminator. This termination site has an unusual symmetry, which allows it also to serve as a transcription terminator for a gene that converges on the aspartokinase II gene from the opposite direction, an interesting example of genetic economy. The deduced amino acid sequence of Bacillus subtilis aspartokinase II was compared with the sequences of the three aspartokinases from Escherichia coli (Cassan, M., Parsot, C., Cohen, G. N., and Patte, J. C. (1986) J. Biol. Chem. 261, 1052–1057). Significant sequence similarities suggest a close evolutionary relationship between the four enzymes.

The coding region specifying the two subunits of aspartokinase II (ATP-L-aspartate 4-phosphotransferase, EC 2.7.2.4) from Bacillus subtilis has recently been cloned in a bacterial plasmid (Bondaryk and Paulus, 1985a). Characterization by specific cleavage with restriction endonucleases suggested a map for the aspartokinase II gene in which the coding sequence for the smaller subunit overlaps the promoter-distal portion of the coding sequence for the subunit in the same reading frame (Bondaryk and Paulus, 1985a). Studies of the expression of aspartokinase in Escherichia coli transformed with a recombinant plasmid carrying the complete coding region showed the product to be indistinguishable in its molecular and regulatory properties from the aspartokinase II isolated from Bacillus subtilis, indicating that the cloned DNA fragment contained all the information necessary for the structure and synthesis of the enzyme (Bondaryk and Paulus, 1985a, 1985b).

In this paper, we present the nucleotide sequence of the entire aspartokinase II gene and the adjacent regions on the Bacillus subtilis chromosome. Our data allow the tentative identification of potential control regions and support the earlier proposal (Bondaryk and Paulus, 1985b) that the two aspartokinase subunits are the products of independent translation of in-phase overlapping genes. Comparison of the deduced amino acid sequence of Bacillus subtilis aspartokinase II with that of the three Escherichia coli aspartokinases provides interesting insights into the evolutionary relationships between these enzymes.

EXPERIMENTAL PROCEDURES AND RESULTS

The nucleotide sequence of the Aspartokinase ZZ Gene—The nucleotide sequence of the entire 2.9-kb PstI fragment, known to contain the complete aspartokinase II coding region (Bondaryk and Paulus, 1985a), is shown in Fig. 2, together with the amino acid sequence defined by the major open reading frames. Our earlier studies had shown that the coding sequences of both the and subunits of aspartokinase straddle a unique BamHI site (Bondaryk and Paulus, 1985a). Accordingly, the BamHI site at position 1645 defines the open reading frame extending from residue 612 to residue 2411.

1 Portions of this paper (including "Experimental Procedures," part of "Results," part of "Discussion," Figs. 1, 3, and 9, and Tables IV–VII) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 87M-0520, cite the authors, and include a check or money order for $5.80 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

2 The abbreviation used is: kb, kilobase pairs.

3 Unless otherwise indicated, nucleotide residues are numbered according to the system used in Fig. 2.

* This investigation was supported by Grant DBM 84-14846 from the National Science Foundation and by Biomedical Research Support Group Grant RR 05711 from the National Institutes of Health Biomedical Research Support Program. Some of the computer resources used in these studies were provided by the BIONET National Computer Resource for Molecular Biology, whose funding is provided by the Biomedical Research Technology Program, National Institutes of Health Grant 1 U41 RR-01685. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank/EMBL Data Bank with accession number(s) J02732.

§ Permanent address: Institute of Microbiology, Academia Sinica, Beijing, China.

8787
Fig. 2. Nucleotide sequence of the B. subtilis aspartokinase I1 gene and surrounding regions. The coding region for the aspartokinase subunits (residues 612–1841), a preceding short open reading frame which may function in transcription attenuation (residues 362–436), and an unidentified open reading frame (residues 1–270) are translated. Regions of dyad symmetry are overlined with arrows and their centers indicated by a dot, potential ribosome binding sites are marked by triangles, and the -10 and -35 regions of the putative aspartokinase I1 promoter are enclosed in boxes, with the probable transcription start point shown by a star. The deduced amino acid sequence and putative control elements of a potential operon with opposite polarity are shown in a separate figure (Fig. 9, miniprint).
FIG. 2—continued

1361 1376 1391 1406
AAT TTA ATT GTC AGA GGC ATT GCA TTT GAA GAT CAA ATC ACA AGA GTA ACC ATT
An Leu Ile Val Arg Gly Ile Ala Phe Glu Asp Gln Thr Val Thr Ile
1421 1436 1451
TAG GGC CTO ACT AGC GCC CTO ACA ACT TGG CTT ACT ATT TTT ACA ACA CTT GCC
Tyr Gly Leu Thr Ser Gly Thr Thr Ser Thr Val Thr Thr Ala Lys Arg Asn Ile Asn Val Asp Ile Ile Gln Thr Glu Ala Glu Asp Lys Thr
1466 1481 1496 1511
AAA AGA AAC ATA AAC CTG GAT ATC ATT ATC CAA AGC GAC GGC GAG GAC AAG ACT
Lys Arg Asn Ile Asn Ile Val Arg Ala Phe Thr Thr Asp Gly Leu Glu Thr Glu Ser Ala
1526 1541 1556 1571
GCA ATT CTO TCT CTO GCA AAA ACA GCT GCA GAC GCT GGG GCT CTT
Glu Ile Ser Phe Gly Thr Val Lys Glu Ala Thr Ala Ser Gly Thr Val Ala
1586 1601 1616 1631
GAA GAT TAT AAA GAC GCG CTG GAA TTT GAG AAA ATC GAG ACA GAA ACC AAA TGG
Glu Glu Tyr Asp Ala Arg Val Asp Gln Ile Thr Arg Ala AAA Arg AAA TGG
1636 1651 1666 1681
GCT AAA GAA TCT ATT GCT GGA TCC GTC AAA ACT TTG TCT ACT ATT TTT ACA ACA CTT GCC
Ala Lys Val Ser Ile Val Gly Ser Gly Thr Ile Phe Ser Thr Thr Thr Ala
1706 1721 1736 1751 1766 1781
TCT GAA ATC AAA GTC GCA ACA ACA AAA ATG TTA GAC GGC GCT GGG GCT CTT
Glu Ser Arg Ile Arg Val Arg Arg AAA ATG Val Asp Gly Ala Gly Gly
1796 1811 1826 1841
GAC GGG CTC GAA TTT GGC TAA AAC GAC CCT GCT GAA TTA TGA
Asp Gly Val Glu Phe Gly Leu Tyr Ala Asp Gln CTT CCT GAA TTA
1851 1866 1881
CAATCAAAA GGCGGGACTA AGTCAGAACC ACTTTATCGG TGAATTTGTT CAGCGATAAA
CTCTCTTTTT CACTTGTTCA TATGATTCGG CAATCTGATC ATTCATCAAT
1921 1936 1951 1966 1981
AGTCAGAACC ACTTTATCGG TGAATTTGTT CAGCGATAAA
CTCTCTTTTT CACTTGTTCA TATGATTCGG CAATCTGATC ATTCATCAAT
1991 2006 2021 2036 2051
TAAATTCTT CCAACATGAA ACCAGGCTCC ACCGAGAAGC AAGCTTCTTT CTCGCTTTTCA ACGCGCTTGG
2061 2076 2091 2106 2121
AAATCAGGCC CCGCTCAACT TCAATTCGAA CCTCCTGTTTT TTTCGCTTCT CACCGCTGCA GCCCTGCCG
2131 2146 2161 2176 2191
CCCGGCTTCT TCAAGAGCG GCGGATCAT CTCCCTCTAG TCAAGAGCAA ACTTACGCG GAGATTTT
2201 2216 2231 2246 2261
CCTGCCCCGCA CTACATGGG AAGATATAT TGGCCGATCC TATGCTATCT GTATCTAGT
2271 2286 2301 2316 2331
GCTAGGCTTACGGTTT CTAATTTCTT TTTCGCTTCT CAAATTTTCT TGGCTATCTC CTATCAGCTC
2341 2356 2371 2386 2401
TGTCGATCGT TAAATCCTGAA CAAATGAGCT TCAAGAGGG GAGATCTT
2411 2426 2441 2456 2471
TTATGAAAA TTGAAGAGAA AATATATAT CAAATTTTAA CAAAATTCG AATTGAAGAAA
2481 2496 2511 2526 2541
ATTATATTT ATATACGCCT TTTTCTGAG CCGTTTGGC GGGAGAATC AAGTAAATG TGAAATAGC
2551 2566 2581 2596 2611
TTTAAGAAC TGGTACAAAG CAAAGAGAG GGGAGAAGAA TTAGGCTGAA CTACCTGCA TCAACAGGG
2621 2636 2651 2666 2681
GGAAAAAGA GTCTAGAGAG TATTTTCTAT CAAAATGGAG CAGTCTTTT GCCTCTGGC
2691 2706 2721 2736 2751
TGCTGCTCT GTGTTATCAG ATTTATGCC TCAACACTG TGGCTTGAGG GGGCCGCGG CATTGCAAA
2761 2776 2791 2806 2821
GGCTGCGT CATATAGAGA GGCTTCTGTAT TCAATTTTCT CAAATTTTCT TGGCTATCTC CTATCAGCTC
2831 2846 2861 2876 2891
ATTATATTT ATATACGCCT TTTTCTGAG CCGTTTGGC GGGAGAATC AAGTAAATG TGAAATAGC
2901 2916
TGGGATCTTCT GCAATGCTTCAGCGGAGG
The DNA fragments described in Table I, two additional fragments generated from one of these by internal BglII cleavage at position 251, and an EcoRV fragment (residues 957-1484) derived from the entire 2.9-kb PstI segment were provided with suitable cohesive ends using synthetic DNA linkers and ligated into the BamHI and SalI sites of the polynucleotide region of pSL100, except for segment 1-260, which was ligated into the BamHI and HindIII sites. Cultures of E. coli HB101 were transformed with the recombinant plasmids, 10^6 cells were plated at various concentrations of chloramphenicol as indicated, and the number of colonies was scored.

### Table II

| Fragment inserted into pSL100 | Plating efficiency in presence of chloramphenicol at concentration of |
|-----------------------------|----------------------------------------------------------|
|                            | 10 µg/ml | 20 µg/ml | 50 µg/ml | 100 µg/ml |
| Residues 1-598              | 100      | 100      | 100      | 100       |
| Residues 1-577              | 100      | 50       | 0        | 0         |
| Residues 1-489              | 100      | 100      | 100      | 100       |
| Residues 1-380              | 100      | 100      | 100      | 100       |
| Residues 1-243              | 0        | 0        | 0        | 0         |
| Residues 1-260              | 0        | 0        | 0        | 0         |
| Residues 261-598            | 0        | 0        | 0        | 0         |
| Residues 957-1484           | 0        | 0        | 0        | 0         |

*Corrected for loss of insert in 6% of transformed cells.

### Table III

| Fragment inserted into pL7003 | Plating efficiency in presence of chloramphenicol at concentration of |
|-------------------------------|----------------------------------------------------------|
|                               | 5 µg/ml | 10 µg/ml | 20 µg/ml | 50 µg/ml |
| Residues 1-598                | 100     | 10       | 2        | 0        |
| Residues 1-577                | 100     | 50       | 8        | 0        |
| Residues 1-489                | 100     | 100      | 100      | 80       |
| Residues 1-380                | 100     | 100      | 100      | 100      |
| Residues 1-243                | 0       | 0        | 0        | 0        |
| Residues 1-260                | 0       | 0        | 0        | 0        |
| Residues 261-598              | 0       | 0        | 0        | 0        |
| Residues 957-1484             | 0       | 0        | 0        | 0        |

Beyond residue 489 have significantly less promoter activity than those terminating at residue 380 or 489. This effect was especially clear in *B. subtilis* (Table III), suggesting that the region between residue 489 and the translation start site represents a negative control element. Although the quantitative aspects of these results must be interpreted with caution, since it was not established that the copy numbers of the various plasmids were the same, the smaller effect of the intervening region seen in *E. coli* (Tables I and II) suggested that the full expression of the negative control requires the physiological context of the *B. subtilis* cytoplasm. As discussed later, the intervening region probably functions in the attenuation of transcription.

The same promoter probe plasmids were also used both in *B. subtilis* and in *E. coli* to examine the region near the beginning of the β subunit. No promoter activity was found associated with residues 957–1484 (Tables II and III). As this includes nearly 400 base pairs upstream from the beginning of the β subunit, it seems likely that the latter is translated from the same primary transcript as the α subunit.
The site of transcription initiation was also determined by the direct examination or RNA transcripts by hybridization mapping of mRNA isolated from E. coli JM101 transformed with a plasmid carrying the B. subtilis aspartokinase II gene or from B. subtilis VB217 derepressed for aspartokinase II. The procedure involved the extension of a radiolabeled primer, annealed to a single-stranded DNA template, by bacteriophage T4 DNA polymerase in the presence of mRNA hybridized to the same template (Hu and Davidson, 1986). Since T4 DNA polymerase cannot displace a hybridized RNA moiety from its DNA template, primer extension should stop at the 5'-terminus of the hybridized mRNA, the 3'-end of the growing DNA chain thus marking its position. Experiments were carried out with three different templates that had been inserted into M13mp18 or M13mp19 at the BglII site (residue 261), the DdeI site (residue 212), or the Smal site (residue 99) of the 5'-flanking region of the B. subtilis aspartokinase II gene and which extended into the coding region for the \( \alpha \) subunit. As shown in the Fig. 3, the site of the first significant termination of primer extension in the presence of mRNA was at residues 280, 281, and 282, respectively, with the three DNA templates, whereas no significant termination was seen in that region in the absence of added RNA. The first DNA residue hybridized with mRNA was thus at position 281, suggesting that this might represent the transcription start site. Considerable primer extension proceeded beyond residue 280, probably because of the presence of partially degraded mRNA molecules and perhaps also because of limited RNA strand displacement by T4 DNA polymerase. The latter was suggested by the progressive increase in length of the extended primer as the distance between the priming site and the transcription start site increased with the three templates employed (Fig. 3). On the other hand, the possibility cannot be excluded that the 3-residue range of primer extension observed in these experiments was due to limited exonuclease degradation of the 5'-ends of the transcripts or to heterogeneity in the transcription start site. Almost identical results were obtained with mRNA isolated from an aspartokinase II overproducing strain of B. subtilis and from E. coli transformed with a plasmid carrying the aspartokinase II gene.

Sites of Translation Initiation—The translation initiation site of the \( \alpha \) subunit of aspartokinase II is unambiguously defined as the ATG sequence at position 612, the amino-terminal nonapeptide of the \( \alpha \) subunit being encoded by residues 615–641 and the in-phase TAA at position 606 preceding the use of other potential formyl-methionine codons upstream. On the other hand, the fact that residues 1347–1394 code for the amino terminus of the \( \beta \) subunit does not necessarily designate the ATG sequence at residue 1347 as the translation initiation site, since the isolated \( \beta \) subunit might instead be the product of proteolytic processing of the

---

**Fig. 3. Mapping of the start point of B. subtilis aspartokinase II mRNA.** The transcription start point of the aspartokinase II gene was identified by primer extension with T4 DNA polymerase as described under "Experimental Procedures." The single-stranded DNA templates consisted either of residues 261–1163 (Panel A), residues 212–656 (Panel B), or residues 99–1163 (Panel C) of the cloned B. subtilis DNA fragment and were hybridized with mRNA isolated from B. subtilis VB217 (lane 1) or from E. coli JM101 transformed with a plasmid carrying the aspartokinase II gene (lane 3) or were left unhybridized as controls (lane 2). After annealing with and extension of \( ^{32} \)P-labeled sequencing primer, the samples were subjected to polyacrylamide gel electrophoresis under denaturing conditions in parallel with dideoxynucleotide sequencing mixtures prepared with the same template-primer combinations (lanes 4–7). Part of the nucleotide sequence deduced from the sequencing lanes is shown on the right of each panel, with the putative -10 region of the promoter indicated by a box and the shortest extended primer segment by an asterisk.
Overlapping Genes for *B. subtilis* Aspartokinase II Subunits

**Fig. 4.** Effect of deletion of proximal portion of the *B. subtilis* aspartokinase II α subunit on the synthesis of β subunit. The production of aspartokinase II subunits was measured by Western blotting as described under "Experimental Procedures" in *E. coli* HB101 transformed with a pUC18 plasmid carrying the entire aspar- tokinase II coding region (lane 1), transformed with a similar plasmid from which a portion of the coding sequence of the α subunit (residues 566-1234) had been deleted (lanes 3 and 4), or untransformed (lane 5). Lane 2 is a sample of purified *B. subtilis* aspartokinase II, with the α and β subunits identified on the left.

α subunit, as had been originally proposed (Moir and Paulus, 1977b). To distinguish between these possibilities, the production of β subunit was examined in *E. coli* HB101 transformed with a plasmid carrying a deletion of residues 566-1234, which included the translation start and a large portion of the coding sequence for the α subunit. Extracts from the transformed cells were subjected to polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate and analyzed for the presence of aspartokinase II subunits by immunoblotting with antiserum against the β subunit, which cross-reacts with both aspartokinase subunits (Moir and Paulus, 1977b). The results showed that the deletion of residues 566-1234 prevented the production of the aspartokinase II subunit but not of the β subunit (Fig. 4). The observation that, under these conditions, β subunit could be produced in the absence of α subunit clearly demonstrated that β subunit is not derived from the latter but is the product of independent translation. Since the deletion used in the experiment encompassed the entire translation initiation region of the α subunit, it is most likely that the translation of β subunit starts at the ATG sequence at residue 1347, which corresponds to the amino-terminal methionine residue of the β subunit and is preceded by a strong ribosome binding site (residues 1333-1338). The nature of the additional band of molecular weight about 30,000 seen in these experiments (Fig. 4, lane 1) is not clear. The question of whether it is a degradation product of α subunit or the result of translation initiation at an additional site is under investigation.

**DISCUSSION**

**Translation Initiation Sites for the Aspartokinase Subunits**—The cloned 2.9-kb *PstI* fragment of *B. subtilis* DNA contains three major open reading frames: residues 1-264; residues 612-1835; and (with opposite polarity) residues 2328-1885. The aspartokinase II α subunit is encoded by the central reading frame, its amino-terminal nonapeptide being specified by residues 615-641 and its carboxyl-terminal dipeptide by residues 1830-1835. The amino-terminal hexadecapeptide of the β subunit corresponds to residues 1347-1394 within the α subunit coding sequence, in agreement with our earlier suggestion (Bondaryk and Paulus, 1985a) that the two aspartokinase subunits are specified by in-phase overlapping reading frames. The coding regions for the amino termini of the aspartokinase subunits either start with ATG or are directly preceded by ATG, consistent with the notions that they represent translation start sites. Just upstream of the putative formyl-methionine codons of the α and β subunits are the sequences AAAGG (residues 597-601) and AGGAGG (residues 1333-1338), respectively, both complementary to the 3′-end of *B. subtilis* 16 S ribosomal RNA (5′-CCTCCTTCT-3′) and, therefore, potential ribosome binding sites (McLaughlin et al., 1981). It is very likely that the synthesis of the α subunit starts with the ATG sequence at residue 612 and that the terminal formyl-methionine codon is removed from the nascent α subunit post-translationally. The translation start of the β subunit is more problematical, since the sequence data alone cannot rule out the possibility that the β subunit is derived post-translationally from α subunit by proteolytic processing. However, that possibility could be eliminated by the use of a mutant plasmid from which the translation initiation site and the amino-terminal half of the α subunit had been deleted. Strains transformed with such a deletion plasmid were unable to produce α subunit but did produce normal β subunit, a clear demonstration that under these conditions the β subunit of aspartokinase II is not derived from α subunit but rather is translated independently.

The observation that the aspartokinase II subunits are translated independently raises the interesting question of how their synthetic rates are coordinated. In exponentially growing cells of *B. subtilis*, the α and β subunits are synthesized at equivalent rates, whereas α subunit is produced in 2-fold excess in germinating *B. subtilis* spores and β subunit is overproduced nearly 4-fold in *E. coli* transformed with a recombinant plasmid carrying the *B. subtilis* aspartokinase II gene (Bondaryk and Paulus, 1985b). This suggests that in growing *B. subtilis* a balance has evolved between the various factors that determine the rate of peptide chain initiation to assure equimolar intracellular levels of the aspartokinase subunits and that this balance is perturbed under non-steady state conditions (germinating spores) or in a foreign cytoplasm. Among the factors that could influence the relative rates of initiation of α and β subunits are the relative affinities of ribosomes for the respective ribosome binding sites, the competition between ribosomes actively translating the α subunit mRNA and those binding to the mRNA at the internal site for the initiation of β subunit, and the rate of degradation of the 5′-terminal portion of the aspartokinase mRNA.

The sequence data allow us to evaluate only the first factor and indicate that the ribosome binding site at the beginning of the α subunit post-translationally. The trans-
ysis showed that an effective promoter region was located between residues 243 and 380 of the cloned \textit{B. subtilis} \textit{psII} segment but that the promoter was inactivated by cleavage with \textit{BglII} endonuclease at position 261. Examination of the nucleotide sequence near the unique \textit{BglII} site revealed the presence of an AT-rich region, followed by hexanucleotide sequences (TTGTCG and TAAAT) homologous to those frequently found about 35 and 10 nucleotides upstream of bacterial transcription start sites (TGGACA and TATAAT, respectively; Hawley and McClure, 1983). The -35 and -10 consensus sequences are separated by 17 base pairs, the preferred spacing for bacterial promoters (Hawley and McClure, 1983). The presence of the \textit{BglII} endonuclease recognition site in this spacer region explains the inactivation of the promoter by cleavage with \textit{BglII}. Mapping experiments of mRNA isolated from \textit{E. coli} JM101 transformed with a recombinant plasmid carrying the \textit{B. subtilis} aspartokinase gene or from \textit{B. subtilis} VB217, an overproducer of aspartokinase II, indicated that its 5' terminus corresponded to residues 281–283 in the DNA sequence, consistent with the proposed promoter location. Analysis of other polynucleotide sequences including more than 300 residues just upstream from the \textit{a} subunit translation initiation site (residues 281–598) and 400 residues upstream from the \textit{b} subunit start (residues 957–1484) failed to reveal any other promoter elements, suggesting that the two aspartokinase subunits are encoded by a single mRNA species initiated near residue 281. The aspartokinase II mRNA might thus be considered a polycistronic mRNA but with the unusual property that the two cistrons are overlapping.

Another unusual property of the aspartokinase II mRNA is that more than 300 nucleotides intervene between its 5'-end and the site of translation initiation of the \textit{a} subunit. Examination of this leader sequence reveals two interesting features (Fig. 5). One of these is an open reading frame (residues 362–433) which encodes a 24-residue polypeptide and is preceded by a strong ribosome binding site (residues 347–352). The other is the presence of four regions with extensive self-complementarity and thus with the potential to form hairpin loops. The first of these palindromic regions overlaps the open reading frame, whereas the last has the characteristic of a \( p \)-independent terminator by ending with seven consecutive uridylic residues (Adhya and Gottesman, 1978). Furthermore, sequence complementarity exists between the first, second, and fourth palindromic regions such as to allow an alternate pattern of secondary structure (indicated by the lines in Fig. 5). This structural pattern is characteristic of the transcription attenuator elements found in many biosynthetic operons (Kolter and Yanofsky, 1982), the first, second, and fourth hairpin loops corresponding to the promoter, pre-premier, and terminator elements, respectively (Nargang et al., 1980). The putative 24-residue leader peptide contains 4 lysine residues in the region that precedes or overlaps the protector loop, so that lysine deficiency would cause the stalling of ribosomes in this region and thereby destabilize the protector structure and favor the alternate base pairing pattern involving two pre-premier loops without a functional transcription terminator (Nargang et al., 1980). According to this model, limitation of lysine, a major end product of the aspartate pathway, would allow transcription to proceed beyond the attenuator site into the structural gene for aspartokinase, thus enhancing the rate of synthesis of the lysine biosynthetic enzyme. Experiments are in progress to test this model by mapping the transcripts produced in \textit{B. subtilis} under conditions of lysine deficiency and excess. It should be noted that the structure of the putative aspartokinase II attenuator is much more complex than that of the only other \textit{B. subtilis} attenuator described (Shimotsu et al., 1986), which does not seem to involve the synthesis of a leader peptide but rather the interaction with a regulatory protein. Undoubtedly, the control of transcription of the aspartokinase gene will be more complex than outlined above and may involve additional factors, such as a putative negative control element defined by one class of mutants resistant to S-(2-aminoethyl)-L-cysteine (Yeh and Steinberg, 1978; Mattiiöö et al., 1979). It may be that the attenuation control of the aspartokinase II operon combines the elements involved in the regulation of \textit{E. coli} and of \textit{B. subtilis} transcription attenuators, \textit{i.e.} the synthesis of a leader peptide as well as the interaction with a regulatory protein. Extensive studies on transcription both \textit{in vitro} and \textit{in vivo} will be necessary to test this possibility.

The termination of the aspartokinase II transcript appears to occur right after the translation termination site at a region of dyad symmetry, which in its latter half contains a run of five thymidylate residues. The symmetrical region resembles a \( p \)-independent terminator site (Adhya and Gottesman, 1978) except that the consecutive thymidylates are within the palindromic sequence. This type of structure has also been observed in the bidirectional transcription terminator of the
converging tonB and P14 genes of *E. coli* (Grundstrom and Jaurin, 1982). Transcription of either DNA strand would give rise to an RNA molecule which could assume a hairpin structure with five uridylates in its distal stem as required for a ρ-independent terminator site (Fig. 6). As mentioned earlier, the reading frame for the aspartokinase II subunits converges on a large open reading frame of opposite polarity (residues 2228–1885), and the transcription of both coding regions is presumably terminated at the same bidirectional ρ-independent terminator element.

Relation to Adjacent Open Reading Frames—Extensive open reading frames both precede and follow that of the aspartokinase II subunits. As discussed above, the reading frame just downstream from the aspartokinase gene is of opposite polarity and appears to share with it a bidirectional transcription terminator. It potentially encodes a 148-residue polypeptide of unknown function and is preceded by a strong ribosome binding site and a potential transcription promoter (see the Miniprint), consistent with the idea that it represents a converging monocistronic operon.

The cloned 2.9-kb *B. subtilis* DNA fragment begins with an open reading frame that encodes 88 amino acid residues and is followed by tandem termination codons. If this indeed represents a fragment of a potential coding region, one can ask the interesting question of whether it may be functionally related to the aspartokinase gene. Such a relationship is suggested by the fact that the putative promoter of the aspartokinase gene overlaps the carboxyl-terminal portion of the open reading frame, a situation that has also been observed elsewhere, e.g. the *ampC* and *frd* operons of *E. coli* (Postle and Good, 1985). A consequence of such an arrangement is that the terminator of the upstream operon is interposed between the transcription and translation start sites of the second operon and thus serves as an attenuator of its transcription. In the case at hand, it is the attenuator element discussed earlier, postulated to control the transcription of the aspartokinase II gene in response to the availability of lysine, which would serve as transcription terminator of the earlier operon. Consequently, under conditions of lysine limitation, termination of transcription of the upstream operon would be incomplete and transcriptional read-through into the aspartokinase operon would yield a polycistronic mRNA in addition to the transcript initiated at the aspartokinase II promoter. Such a situation would of course not occur with the recombinant plasmid under study, which carries only a terminal promoterless fragment of the upstream operon, but would obtain with the genes on the *B. subtilis* chromosome. The identity of the gene adjacent to the aspartokinase II locus

Fig. 7. Alignment of the deduced amino acid sequences of *B. subtilis* aspartokinase (AK) II and the three *E. coli* aspartokinases. The deduced amino acid sequences of the α subunit of *B. subtilis* aspartokinase II has been aligned with that of *E. coli* aspartokinase III (Cassan et al., 1986) with the aid of a BIONEP Intelligenetics program. The latter, in turn, was aligned with the other *E. coli* aspartokinases as proposed by Cassan et al. (1986). The dotted vertical line indicates the start of the *B. subtilis* aspartokinase II β subunit as well as the site of partial trypsin digestion of *E. coli* aspartokinase-homoserine dehydrogenase (HSDH) I (Sibilli et al., 1981).
would thus be of considerable interest. Inspection of the amino acid sequence of the sequenced fragment between nucleotide residues 112 and 189 reveals striking clusters of lysine residues, perhaps a clue to a possible relationship to lysine biosynthesis.

**Amino Acid Sequence Comparison with the E. coli Aspartokinases**—The nucleotide sequences of the genes encoding the three E. coli aspartokinases have been elucidated by Cohen and co-workers (Katiaka et al., 1980; Zakin et al., 1983; Cassan et al., 1986). A comparison of the corresponding amino acid sequences has revealed considerable similarity between aspartokinase III, the product of the lysC gene, and the aspartokinase domains of aspartokinase I-homoserine dehydrogenase I (thra) and aspartokinase II-homoserine dehydrogenase II (metL) (Cassan et al., 1986). In Fig. 7 these sequences, aligned as proposed by Cassan et al. (1986), are compared with the deduced amino acid sequence of B. subtilis aspartokinase II.[...]

**Aromatic Amino Acid Sequence Comparison with the E. coli Aspartokinases**—The nucleotide sequences of the genes encoding the three E. coli aspartokinases, 31, 26, and 22% of its amino acid residues being identical with those of E. coli aspartokinase III, I, and II, respectively. The regions of major interspecies similarity correspond to the regions of major similarity between the three E. coli aspartokinases (Cassan et al., 1986), which lie between residues 75–55 and residues 136–238 of B. subtilis aspartokinase II. Especially extensive similarity is seen in the 145–191 region, suggesting that it may perhaps represent the catalytic center of aspartokinase. An indication of similarity on the level of tertiary structure is provided by the observation that the start of the subunit (residue 247) is in a position homologous to the site of partial trypsin cleavage of E. coli aspartokinase I-homoserine dehydrogenase I (Sibilli et al., 1981). Partial proteolysis experiments with B. subtilis aspartokinase II have indicated that the subunit constitutes a discrete globular domain linked by a protease-sensitive hinge to the catalytic domain (Paulus, 1984), analogous to the hinge linking the aspartokinase domain to one of the homoserine dehydrogenase domains in the bifunctional E. coli enzyme (Fazel et al., 1983).

The close structural relationship between B. subtilis aspartokinase II and the E. coli aspartokinases suggests a common evolutionary origin. A closer relationship exists between B. subtilis aspartokinase II and E. coli aspartokinase III (31% of residues identical) than between the three E. coli aspartokinases (24.5–29.7% of residues identical; Cassan et al., 1986). Cassan et al. (1986) argue, on the basis of similarity of the carboxyl-terminal portion of E. coli aspartokinase III to the homoserine dehydrogenase domain of the two bifunctional aspartokinases, that fusion of aspartokinase with homoserine dehydrogenase to yield aspartokinase-homoserine dehydrogenases I and II must have occurred before the separation of aspartokinase III from the latter. According to this argument, B. subtilis aspartokinase II diverged from E. coli aspartokinase III after the latter diverged from the other E. coli enzymes. If this were indeed the evolutionary pathway of B. subtilis aspartokinase II, one could conclude that the β subunit of that enzyme originated from the homoserine dehydrogenase domain of the bifunctional E. coli aspartokinases or, more specifically, from the globular 25-kDa domain (Iα) postulated to link the aspartokinase and homoserine dehydrogenase catalytic domains in E. coli aspartokinase I-homoserine dehydrogenase I (Fazel et al., 1983). In view of the fact that the function of the β subunit of B. subtilis aspartokinase II is not yet understood (Paulus, 1984), such a possibility would be of considerable interest.

**Conclusions**—Our results show that the B. subtilis aspartokinase operon has several unusual features. It is composed of two overlapping cistrons, a situation originally described only in viral and plasmid genomes (Nasmark et al., 1983) but now being increasingly recognized also on the bacterial chromosome (Smith and Parkinson, 1980; Plumbridge et al., 1985; Mackman et al., 1985; Flower and McHenry, 1986) and even in eukaryotes (Kozak, 1986). The structural genes are preceded by an exceptionally long leader sequence that appears to function as a transcription attenuator but differs from the only such regulatory element described in B. subtilis (Shimosu et al., 1986) by encoding a leader peptide analogous to those found in E. coli (Koler and Yanofsky, 1982). The transcription attenuator seems to function also as a transcription terminator of an adjacent operon, whereas the transcription termination site of the aspartokinase II operon is shared with a converging operon, interesting examples of genetic economy or regulatory subtlety. In order to understand the functioning of these unusual elements, it will be necessary to modify them by deletion or site-directed mutagenesis and study their expression when reintegrated (Haldenwang et al., 1980) into the B. subtilis chromosome. It is hoped that experiments of this type will not only advance our understanding of the control of lysine biosynthesis in B. subtilis and of structure-function relationships in aspartokinase II but also provide interesting insights into the evolution of regulatory strategies.

**Acknowledgments**—We thank Yu-Mei Chen and A. L. Sonenshein for gifts of bacterial strains and plasmids and Jen-Shiang Hong and Rene Lu for helpful advice. The analysis of the amino-terminal sequences of the aspartokinase II subunits was performed by Anna Wong, and preliminary experiments on the mapping of transcription start sites were done by David Fessell. Special thanks are due to Lee M. Graves and Robert L. Switzer (University of Illinois) for useful discussions that alerted us to a sequence error and for sharing with us their unpublished data on the amino-terminal sequence of the aspartokinase II α subunit.

**REFERENCES**

Adhya, S. & Gottesman, M. (1978) Annu. Rev. Biochem. 47, 967–996 Bondaryk, R. P. & Paulus, H. (1983a) J. Biol. Chem. 258, 586–591 Bondaryk, R. P. & Paulus, H. (1983b) J. Biol. Chem. 260, 592–597 Brosius, J. (1984) Gene 27, 57–68 Cassan, M., Brosius, J., Paulus, H. & Patte, J.-C. (1986) J. Biol. Chem. 261, 1052–1057 Choe, P. & Fasman, G. D. (1978) Adv. Enzymol. Relat. Areas Mol. Biol. 47, 45–147 Donnelly, C. E. & Sonenshein, A. L. (1984) J. Bacteriol. 157, 965–986 Fazel, A., Muller, K., Le Bras, G., Garel, J. R., Veron, M. & Cohen, G. N. (1983) Biochemistry 22, 158–165 Flower, A. M. & McHenry, C. S. (1986) Nucleic Acids Res. 14, 8091–8101 Freier, S. M., Kierzek, R., Jaeger, J. A., Sugino, N., Caruthers, M. H., Nelson, T. & Turner, D. H. (1986) Proc. Natl. Acad. Sci. U. S. A. 83, 9573–9577 Grundstrom, T. & Jaurin, B. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 1111–1115 Haldenwang, W. G., Banner, C. D. B., Ollington, J. F., Losick, R., Hoch, J. A., O'Connor, M. B. & Sonenshein, A. L. (1986) J. Bacteriol. 162, 90–98 Hawley, D. K. & McClure, W. R. (1983) Nucleic Acids Res. 11, 2237–2255 Henikoff, S. (1984) Gene 28, 351–359 Hu, M. C. T. & Davidson, N. (1986) Gene 42, 21–29 Katinka, M., Cozzart, P., Sibilli, L., Saint-Girons, I., Chalvignac, M. A., Le Bras, G., Cohen, G. N. & Yaniv, M. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5730–5733 Kolbs, H. V., J., Lu, R. C. & Wohlrab, H. (1985) J. Chromatogr. 327, 195–200 Koler, R. & Yanofsky, C. (1982) Annu. Rev. Genet. 16, 113–134 Kozak, M. (1986) Cell 47, 481–483 Lin, H. C., Lei, S. P. & Wilcox, G. (1985) Anal. Biochem. 147, 114–119
Overlapping Genes for B. subtilis Aspartokinase II Subunits

Lu, R. C. & Wong, A. (1985) *J. Biol. Chem.* 260, 3456-3461

Mackman, N., Nicaud, J. M., Gray, L. & Holland, I. B. (1985) *Mol. Gen. Genet.* 205, 529-536

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning*, p. 196, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Mongkolsuk, S., Ambulos, N. P. & Lovett, P. S. (1984) *J. Bacteriol.* 160, 1-8

Moore, S. & Stein, W. H. (1957) *Methods Enzymol.* 6, 819-831

Nargang, F. E., Subrahmanyan, C. S. & Umbarger, H. E. (1980) *Proc. Natl. Acad. Sci. U. S. A.* 77, 1823-1827

Normark, S., Bergstrom, S., Edlund, T., Grundstrom, T., Jaurin, B., Lindberg, F. P. & Olsson, O. (1983) *Annu. Rev. Genet.* 17, 499-525

Paulus, H. (1984) *J. Bacteriol.* 160, 403-418

Plumbidge, J. A., Deville, F., Sacerdot, C., Peterson, H. U., Cenamiyo, T., Cozzone, A. G., Grunberg-Manago, M. & Hershey, J. W. B. (1989) *EMBO J.* 8, 221-229

Pluskal, M. G., Przekop, M. B., Kovacevic, M. C. & Hicks, J. W. (1980) *J. Biol. Chem.* 255, 10228-10230

Postle, K. & Good, R. A. (1985) *Cell* 41, 577-585

Plumbridge, J. R., Murray, C. L. & Rabinowitz, J. C. (1981) *J. Biol. Chem.* 256, 11283-11291

Miller, J. H. (1972) *Experiments in Molecular Genetics*, pp. 352-355, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Maruyama, T., Gojobori, T., Aota, S. & Ikemura, T. (1986) *Nucleic Acids Res.* 14, 151-197

Mattioli, R., Bazzicalupo, M., Federici, G., Gallori, E. & Polsinelli, M. (1979) *J. Gen. Microbiol.* 114, 223-225

McLaughlin, J. R., Murray, C. L. & Rabinowitz, J. C. (1981) *J. Biol. Chem.* 256, 11283-11291

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning*, p. 196, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Moir, D. & Paulus, H. (1977a) *J. Biol. Chem.* 252, 4648-4654

Moir, D. & Paulus, H. (1977b) *J. Biol. Chem.* 252, 4655-4661

Marcos, A., Ambulos, N. P. & Lovett, P. S. (1984) *J. Bacteriol.* 160, 1-8

Nargang, F. E., Subrahmanyan, C. S. & Umbarger, H. E. (1980) *Proc. Natl. Acad. Sci. U. S. A.* 77, 5270-5274

Yeh, E. C. & Steinberg, W. (1978) *Mol. Gen. Genet.* 158, 287-296

Zakin, M. M., Duchange, N., Ferrara, P. & Cohen, G. N. (1983) *J. Biol. Chem.* 258, 3028-3031

**SUPPLEMENTAL MATERIAL TO**

**STRIKING THE GENE FOR THE SUBUNIT OF ASPARTOKINASE II FROM BACILLUS SUBTILIS**

**Nai-Yong Chen, Fu-Mei O'G., and Henry Paulus**

**EXPERIMENTAL PROCEDURES**

Materials - 125I- and 32P-labeled nucleotides were obtained from New England Nuclear, restriction endonucleases from New England Biolabs, and other enzymes from New England Biolabs. Bethesda Research Laboratories, or Pharmacon. Reagents for nucleotide sequence analysis, including the IN35 (35S)-labeled oligonucleotide primer, were from New England Biolabs. The lam0952H variant of lam0952 is in the subunit of aspartokinase II (Hori and Paulus, 1977a) was prepared as described by Hori and Paulus (1986a). Lam0952H-conjugated single-strand anti-sense virus was obtained from Missa Labs.

**Antiterminal Sequence Analysis** - Antiterminal analysis was performed by Ms. Anna Ving (Department of Medical Research, Protein Biomedical Research Institute) by automated Edman degradation. About 23 cycles of each of the α and β subunits of B. subtilis aspartokinase II, purified as described by Hori and Paulus (1977a), were applied in 50 mM sodium chloroform at 37 °C for 2 h. Priming was performed in a Beckman 80S DataColumn sequencing with primer G5427 (Lu and Ving, 1985). Phosphorylated Edman derivatives of amino acids were determined as described by Ralch et al. (1981).

**Partial Serological Analysis of Subunits and Species Differences** - No of the bacterial strains and plasmids used and the conditions of growth were described previously (Hori and Paulus, 1985). Bacillus subtilis 36168 and 36169 and plasmids pBS18 and pBS19 were obtained from New England Biolabs, plasmid pBS10 from Yu-Mei Chan, Microbiological School, plasmid pGEMH and its host, a S. marcescens strain, and pEM62 and EM22 (from A. L. Sussman, Tulac University, and B. subtilis 168 [pT570 in strain EM22]) from the Matsumoto Genetic Stock Center.

**Nucleotide Sequence Determination** - Nucleotide sequences were determined by the dideoxyribonucleotide chain-termination method of Sanger et al. (1977). using [α-32P]dATP and [γ-32P]ATP at high specific activity. The RNA was isolated by a method that employed sodium dodecyl sulfate (SDS) and phenol extraction with sodium dodecyl sulfate and phenol. The reaction mixtures were subjected to polyacrylamide gel electrophoresis along with a control from which RNA had been omitted and in parallel with corresponding sequence ladders as described by White and Paulus (1986).

**RESULTS AND DISCUSSION**

**Antiterminal Analysis of Subunits of aspartokinase II** - The antiterminal analyses of the amino acid sequences of the subunits of aspartokinase II were performed by the method of Sanger et al. (1977). using [α-32P]dATP and [γ-32P]ATP at high specific activity. The reaction mixtures were subjected to polyacrylamide gel electrophoresis along with a control from which RNA had been omitted and in parallel with corresponding sequence ladders as described by White and Paulus (1986).

**Analysis of Sequence Data** - The manipulation of nucleotide sequence data, such as searches for homologies and dyad symmetry, restriction analysis, translation into amino acid sequences, analysis of codon usage, and amino acid sequence comparisons, were done with the aid of the DNA STRAINAGE computer program.

**RESULTS AND DISCUSSION**

**Antiterminal Analysis of Subunits of aspartokinase II** - The antiterminal analyses of the amino acid sequences of the subunits of aspartokinase II were performed by the method of Sanger et al. (1977). using [α-32P]dATP and [γ-32P]ATP at high specific activity. The reaction mixtures were subjected to polyacrylamide gel electrophoresis along with a control from which RNA had been omitted and in parallel with corresponding sequence ladders as described by White and Paulus (1986).

**Analysis of Sequence Data** - The manipulation of nucleotide sequence data, such as searches for homologies and dyad symmetry, restriction analysis, translation into amino acid sequences, analysis of codon usage, and amino acid sequence comparisons, were done with the aid of the DNA STRAINAGE computer program.

**Nucleotide Sequence of the 7.8 kb B. subtilis DNA Fragment** - The sequence analysis strategy as well as a partial restriction map of the 7.8 kb B. subtilis DNA Fragment is detailed in Fig. 1. Except for residues 600-695 and a segment between residues 2630 and 2710, each residue of the B. subtilis DNA Fragment was analyzed. As a control point, the sequences of the two strands were in disagreement. This occurred at residue 334, but fortunately the discrepancy was resolved by the presence of a Fok I restriction endonuclease site at position 334. Another ambiguity at position 2499 has not been completely resolved but is free from the aspartokinase II sequence. The restriction map deduced from the nucleotide sequence was in good agreement with that determined experimentally (Hori and Paulus, 1985). except that one of the two closely related DNA sites and a distal Eco RI site has been overlooked in the earlier work.
Overlapping Genes for B. subtilis Aspartokinase II Subunits

| Cycle | α-subunit | β-subunit |
|-------|---------|---------|
| 1     | gly    | 4.3     |
| 2     | ser    | 4.6     |
| 3     | gly    | 4.9     |
| 4     | ser    | 5.0     |
| 5     | gly (gly) | 2.7 |
| 6     | ser    | 2.7     |
| 7     | gly    | 2.5     |
| 8     | ser    | 2.7     |
| 9     | gly    | 1.5     |
| 10    | ser    | 2.3     |
| 11    | gly    | 4.9     |
| 12    | ser    | 3.2     |
| 13    | gly    | 3.0     |
| 14    | ser    | 1.4     |
| 15    | gly (gly) | 1.7 |

*Sequences were using 27 amino acids of peptide

*Sequences were using 23 amino acids of peptide

*Omitted by HPC (Kosaruk et al., 1985)

FIG. 1: Restriction endonuclease site and sequence analysis strategy for the B. subtilis DNA fragment carrying the aspartokinase II gene. The numbers denote DNA length in base pairs (bp) and the arrows denote the direction of sequence analysis, the solid arrow denotes analysis with Mfin/MlvI, the dotted arrow analysis of the opposite strand in Mfin/MlvI. Restriction endonuclease cleavage sites are abbreviated as follows: B. subtilis: N: NotI; D: DstI; H: HincII; M: MboII; S: SacI; P: PstI; T: TspI; S: Sall; and H: HhaI.

Analysis of the Reduced amino Acid Sequence of B. subtilis Aspartokinase II - The derivation of the amino acid sequence from the nucleotide sequence of the aspartokinase II gene (Fig. 1) provided accurate values for the molecular weights of the enzyme and its subunits. As illustrated in Table IV, the molecular weights of 66,700 and 37,500 for the α and β subunits, respectively, are in good agreement with the values of 63,000 and 37,000, estimated from electrophoretic mobility in the presence of sodium dodecyl sulfate (Meier and Paulus, 1974). Tractol time agreement was near between calculated and estimated molecular weights of the α subunits from trypsin digestion by partial tryptic digestion of a subunit (Paulus, 1984) and the properties of fusion of the aspartokinase subunits with an adjacent region offlagatamine (Hammar and Fleischer, 1986). The value of 115,000 for the molecular weight of the α subunit from tryptic digestion of the α subunit, estimated by equilibrium ultracentrifugation (Meier and Paulus, 1974), is somewhat smaller than the calculated value of 122,000.

Table IV

| Peptide | Molecular Weight | Experimental | From mRNA Sequence |
|---------|------------------|--------------|-------------------|
| α subunit | 63,000 | 63,110 |
| β subunit | 37,500 | 37,120 |
| Metalloprotease product | 19,000 | 19,000 |
| Aspartokinase product | 40,000 | 40,170 |
| Aspartokinase product | 14,000 | 14,780 |

The amino acid composition of the aspartokinase II gene obtained by amino acid analysis of the purified enzyme (Meier and Paulus, 1974) is in fair agreement with that calculated from the nucleotide sequence (Table V), except for significant deviations of lysine, arginine, aspartate, glutamate, and serine and a high value for glycine. The low values for the basic amino acids could be a consequence of the two-column method (Moore and Stein, 1957) used at that time and that for maintaining a quaternary ensemble.

Table V

| Residue | α-Subunit | β-Subunit | Aspartokinase |
|---------|-----------|-----------|---------------|
| Ala     | 75        | 75        | 75            |
| Arg     | 5         | 2         | 2             |
| Asp     | 30        | 20        | 20            |
| Gln     | 30        | 30        | 30            |
| Gly     | 20        | 20        | 20            |
| His     | 2         | 2         | 2             |
| Ile     | 20        | 20        | 20            |
| Leu     | 25        | 25        | 25            |
| Lys     | 15        | 15        | 15            |
| Met     | 0         | 0         | 0             |
| Phe     | 15        | 15        | 15            |
| Ser     | 15        | 15        | 15            |
| Thr     | 15        | 15        | 15            |
| Trp     | 5         | 5         | 5             |
| Tyr     | 20        | 20        | 20            |
| Val     | 10        | 10        | 10            |

*From Meier and Paulus (1974), expressed as residues per 120,000 residues

The complete amino acid sequence of the aspartokinase II gene was obtained by amino acid analysis of the purified enzyme (Meier and Paulus, 1974) in fair agreement with that calculated from the nucleotide sequence (Table V), except for significant deviations of lysine, arginine, aspartate, glutamate, and serine and a high value for glycine. The low values for the basic amino acids could be a consequence of the two-column method (Moore and Stein, 1957) used at that time and that for maintaining a quaternary ensemble.
Overlapping Genes for *B. subtilis* Aspartokinase II Subunits

![Diagram of DNA sequence and protein structure](image)

**Key:**
- **AAA** Alpha helix
- **SS** Beta sheet
- **T** Turn

**Fig. 8. Predicted secondary structure of *B. subtilis* aspartokinase II.** The deduced amino acid sequence of the aspartokinase II subunit was aligned with that of the aspartokinase I subunit, using the algorithm of Chao and Johnson (1976) and the SCOR program. Alpha helices are denoted by **AAA**, beta sheet by **SS**, and turns by **T**.

**A Conserved Dihydroaminotransferase Unit of Unknown Function**

The structural gene for this dihydroaminotransferase I subunit is transcribed from a single-stranded DNA template and encodes a 121-amino acid polypeptide of unknown function. Its nucleotide and the corresponding amino acid sequence as well as the adjacent non-coding regions are shown in Fig. 9, which represents the DNA partial sequence. The number of nucleotides is indicated (Pybus and McClain, 1978). The nucleotide sequence is presented in the 5' to 3' orientation.

**Fig. 9. Potential operon of a conserved dihydroaminotransferase II gene.** The nucleotide sequence of the cloned *B. subtilis* DNA fragment (residues 270-308), showing the conserved complementarity to the gene given in Fig. 6. The number of residues is indicated, starting with 1 rather than 297. The major open reading frame has been translated and potential control elements are indicated with the same symbols as in Fig. 2.