J. Indian Inst. Sci., Jan.–Feb 1989, **69**, 55–64.

Structure and expression of the overlapping genes for the subunits of *Bacillus subtilis* aspartokinase II

HENRY PAULUS* AND NAI-YONG CHEN†

*Department of Metabolic Regulation, Boston Biomedical Research Institute, Boston, Massachusetts 02114.
†Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, U.S.A.

Received on February 3, 1988.

Abstract

Aspartokinase II of *Bacillus subtilis* is composed of two nonidentical subunits, α and β, arranged in an $\alpha_2\beta_2$ structure. The two subunits are encoded by in-phase overlapping genes, which constitute a single operon. The promoter overlaps an adjacent open reading frame and is followed by an extensive leader sequence that appears to function as a transcription attenuator. The transcription terminator of the aspartokinase II operon is shared with a converging transcription unit. The translation initiation sites of the aspartokinase II α and β subunits were defined by sequence analysis and site-directed mutagenesis. The study of specific deletion, frame-shift, and missense mutations clearly demonstrated that the aspartokinase II subunits are translated independently from overlapping genes.

Key words: DNA sequence, transcription initiation, promoter, transcription attenuator, transcription terminator, translation initiation, site-directed mutagenesis, enzyme subunits.

1. Introduction

Earlier structural studies on aspartokinase II from *B. subtilis* showed that the enzyme is composed of two nonidentical subunits, termed α and β, with molecular weights 43,000 and 17,000, respectively (for review, see ref. 1). The subunits are arranged in an $\alpha_2\beta_2$ structure and have a high degree of homology, the smaller β subunit corresponding to the C-terminal portion of the α subunit. In order to understand the biochemical basis for the homology between the aspartokinase subunits, we undertook the examination of the DNA sequences encoding these polypeptides. Upon screening a λ Charon 4A library of *B. subtilis* DNA for polypeptides cross-reacting with antisera against aspartokinase II, we identified a 5.8-kb EcoRI fragment that encoded both aspartokinase subunits and which could be further reduced by PstI cleavage to an active 2.9-kb fragment. Additional cleavage of the 2.9-kb fragment by BamHI led to the loss of the ability to direct the synthesis of the aspartokinase subunits. Instead, two cross-reactive polypeptides of molecular weights 40,000 and 14,000

*Author for correspondence.

were produced, each 3,000 daltons shorter than the corresponding aspartokinase subunit. The simplest interpretation of this result was that the two aspartokinase subunits are encoded by a single DNA sequence, and that the deletion produced by BamHI cleavage removed the portion that encodes the C-terminal sequence of both the polypeptides. This provided an explanation for the homology between the β subunit and the C-terminal portions of the α subunit. It also allowed the alignment of the coding sequence for aspartokinase II with the restriction map of the 2.9-kb PstI fragment, thus preparing the way for the elucidation of its nucleotide sequence.

2. Nucleotide sequence of the aspartokinase II coding region

The nucleotide sequence of the entire cloned 2.9-kb PstI fragment containing the coding region for the aspartokinase II subunits was determined by the dideoxynucleotide chain-termination method[2], using sequencing vectors that were progressively shortened either by random DNase I cleavage or treatment with exonuclease III. Analysis of the sequence[3] revealed three major open reading frames: residues 1-264; residues 612-1835; and (with opposite polarity) residues 2328-1885 (fig. 1). The coding sequence for the aspartokinase II subunits could be assigned to the central open reading frame (residues 612-1835) on the basis of the following evidence: The BamHI site, whose cleavage leads to the shortening of both aspartokinase subunits[4], is located at position 1645; the N-terminal nonapeptide of the aspartokinase II α subunit corresponds to the codons at positions 615-641; the N-terminal hexadecapeptide of the β subunit corresponds to residues 1347-1394; the carboxyl-terminal sequence alanyl-valine, common to both aspartokinase subunits, corresponds to the codons at residues 1830-1835, which are immediately followed by tandem peptide chain-termination codons; and the molecular weights of the aspartokinase II subunits (43,000 and 17,000) are in good agreement with those of the polypeptides defined by the putative coding sequence (43,710 and 17,728, respectively). The sequence data thus confirmed our preliminary conclusion, based on restriction mapping, that the two aspartokinase subunits are translated from in-phase overlapping coding regions[4].

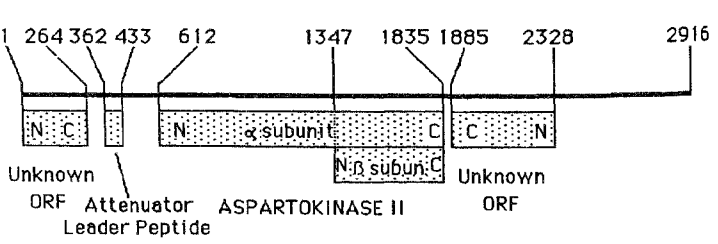


FIG. 1. Disposition of open reading frames in the 2.9-kb PstI fragment of *B. subtilis* DNA containing the coding region for aspartokinase II. The numbers indicate the distance in base pairs from one of the PstI sites, and the amino- and carboxyl-termini of the reading frames are indicated by N and C, respectively. The nucleotide residues are numbered as in Chen *et al*[3].

3. Control of transcription

Various DNA segments upstream of the translation start sites of the aspartokinase II subunits were introduced into promoter probe plasmids and tested for promoter activity[3]. As shown in Table I, the shortest segment to yield full promoter activity included residues 1-380, while cleavage at the unique BglII site (residue 261) yielded two inactive fragments. These observations suggested that the promoter is relatively distant from the translation start and involves essential elements on both sides of residue 261. No promoter activity was seen within 0.4 kb of the translation start site of the β subunit, indicating that the two aspartokinase subunits were translated from a single transcript.

The transcription initiation site was defined more precisely by transcript mapping using a primer elongation procedure[5]. The annealing of mRNA, isolated from a strain of *B. subtilis* that overproduced aspartokinase II or from *E. coli* transformed with a recombinant aspartokinase II plasmid, to M13mp18 or M13mp19 carrying appropriate regions of the cloned *B. subtilis* DNA partially inhibited primer extension beyond residue 281 ± 1[3]. This indicated that the mRNA population contained molecules complementary to residues 282 ± 1 and above, thus defining residue 282 ± 1 as the 5'-end of the aspartokinase II transcript. Examination of the DNA sequence 10 and 35 nucleotides upstream from the transcription start revealed sequences homologous to those seen in promoters recognized by the *E. coli* and the *B. subtilis* σ^{43} RNA polymerase holoenzyme (fig. 2). The unique BglII site at residue 261 is thus between the -10 and -35 regions of the promoter, consistent with the absence of promoter activity in either fragment produced by BglII cleavage. It is

Table I

Fragment inserted into pPL703	Plating efficiency in the presence of chloramphenicol at a 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 1-498	100	100	100	100
Residues 1-380	0	0	0	0
Residues 1-243	0	0	0	0
Residues 1-260	0	0	0	0
Residues 262-598	0	0	0	0
Residues 957-1484	0	0	0	0

Analysis of the promoter activity of various *B. subtilis* DNA fragments inserted into a promoter-probe vector and expressed in *B. subtilis*. DNA fragments were ligated into the BamHI and SalI sites of the polylinker region of pPL703 adjacent to the chloramphenicol acetyltransferase (CAT) gene and introduced into *B. subtilis* BR151. Cells (10^4) were plated at the chloramphenicol concentrations indicated and the number of colonies was compared to the number obtained in the absence of antibiotic. Nucleotides are numbered as in fig. 1 (From Chen *et al*[3]).

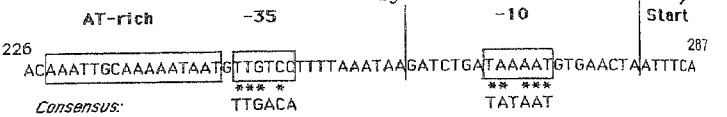


FIG 2. Promoter region of the *B. subtilis* aspartokinase II operon. Nucleotides are numbered as in fig. 1. The consensus sequence shown is that of the promoter for the *B. subtilis* σ^{34} RNA-polymerase holoenzyme.

of interest that the −35 promoter element overlaps the open reading frame of the unidentified gene that precedes aspartokinase II.

Another striking aspect of the results shown in Table I is the much lower promoter activity in some of the longer DNA fragments tested. It is interesting that this negative effect was not seen with DNA fragments that lack either a short open reading frame or one of four regions of dyad symmetry between residues 362 and 544. Closer examination of these features revealed a structural pattern characteristic of the transcription attenuator elements found in many biosynthetic operons of enterobacteria[6], with a coding sequence for a lysine-rich 24-residue peptide, preceded by a strong ribosome binding site and overlapping with the first of the four regions of dyad symmetry, the last of which resembles a ρ-independent transcription terminator. When transcribed into RNA, three of the palindromic regions can assume an alternative secondary structure, which would be favoured by the arrest of ribosomes near the lysine codons of the putative leader peptide, in which the ρ-independent terminator is non-functional (fig. 3), thus providing a mechanism for modulating premature termination of

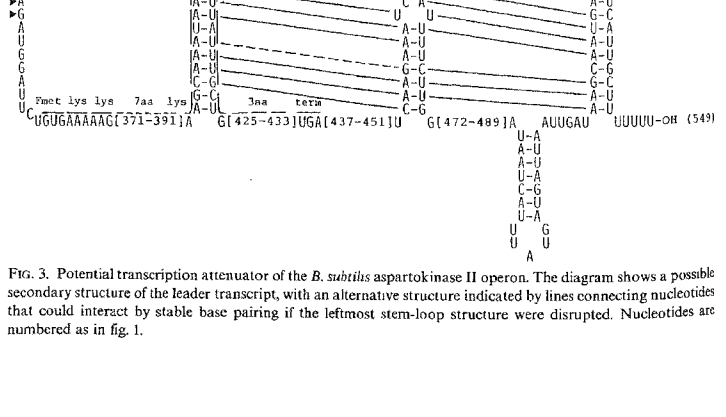


FIG. 3. Potential transcription attenuator of the *B. subtilis* aspartokinase II operon. The diagram shows a possible secondary structure of the leader transcript, with an alternative structure indicated by lines connecting nucleotides that could interact by stable base pairing if the leftmost stem-loop structure were disrupted. Nucleotides are numbered as in fig. 1.

transcription in response to the availability of lysine, an end produce of the aspartate pathway.

4. Termination of transcription

The open-reading frame at the beginning of the cloned 2.9-kb PstI fragment is not immediately followed by an identifiable transcription termination element and it is possible that its termination occurs at the putative aspartokinase II transcription attenuator. If this were indeed the case, one would expect that, when lysine is limiting so that premature transcription termination at the attenuator is suppressed, a run-through transcript encoding the aspartokinase subunits as well as the upstream gene would occur in addition to the mRNA initiated at the aspartokinase promoter. A similar overlap has been seen between the ampC and frd operons in *E. coli*[7].

The end of the aspartokinase II coding region is immediately followed by a region of dyad symmetry, which resembles a ρ-independent transcription terminator except that the run of thymidylate residues is within the palindromic region. A similar structure has been found to terminate the converging tonB and P14 genes in *E. coli*, where it seems to function as a bidirectional transcription terminator[8]. In view of the fact that a large open-reading frame

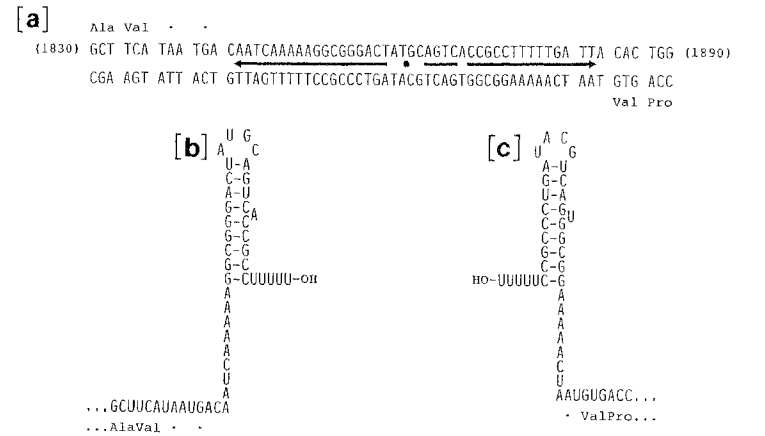


FIG 4. Bifunctional ρ-independent transcription terminator at the end of the *B. subtilis* aspartokinase II operon (a) The DNA sequence following the aspartokinase II coding region. Nucleotides are numbered as in fig. 1 and regions of dyad symmetry are indicated by the arrows (b), (c) The putative 3'-ends of the mRNAs that would result from the operation of the bidirectional transcription terminator shown in (a).

(residues 2328-1885) converges on the aspartokinase II gene (fig. 1), the symmetrical element intervening between the reading frames could serve as terminator for both. As illustrated in fig. 4, transcription of either DNA strand would give rise to an RNA molecule which could assume a hairpin structure with five consecutive uridylate residues in its distal end as required for a ρ-independent transcription termination site[9].

5. Initiation of translation

The N-terminal nonapeptide of the aspartokinase II α subunit corresponds to residues 615-641 of the cloned 2.9-kb PstI fragment[3]. Directly upstream from this sequence is a methionine codon and a potential ribosome-binding site (AAAGG, residues 597-601). Since

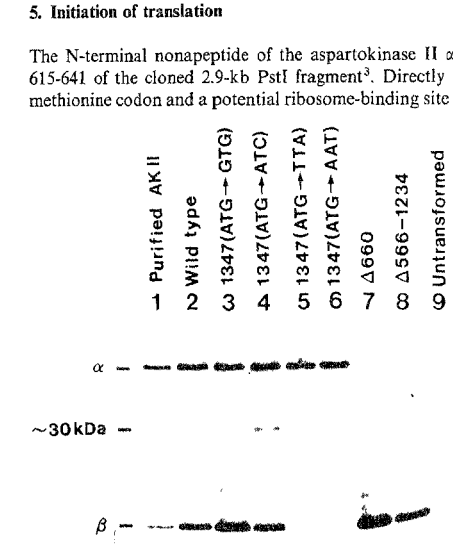


FIG. 5. Immunological analysis of aspartokinase II subunits produced in *E coli* Gif106M1 transformed with multicopy plasmids carrying the wild-type and mutant forms of the *B. subtilis* aspartokinase II gene. Western blot analysis was carried out on French press extracts of transformed *E. coli* Gif106M1[10] as described by Chen *et al*[3]. The samples analyzed were purified aspartokinase II (lane 1), untransformed *E. coli* Gif106M1 (lane 9), and *E. coli* Gif106M1 transformed with pNC8641 (lane 2), pNC1-55 (lane 3), pNC2-44 (lane 4), pNC3-6 (lane 5), pNC4-9 (lane 6), pNC860 (lane 7), and pNC800 (lane 8). α and β designates the position of the corresponding aspartokinase subunits in the purified aspartokinase standard. The mutational alteration of each of the plasmids used is indicated over each lane and is summarized in fig 6.

the open-reading frame does not continue further upstream, these elements must represent the peptide chain-initiation site of the α subunit. On the other hand, the definition of the translation-initiation site for β subunit was more problematical. The N-terminal hexadecapeptide of the β subunit, which starts with a methionine residue, corresponds to residues 1347-1394 and is preceded by a potential ribosome-binding site (AGGAGG, residues 1333-1338). However, the reading frame continues upstream as the coding region for the α subunit and additional information is therefore required to identify the site of translation initiation of the β subunit. Indeed, although earlier pulse-labelling studies had indicated an independent translational origin of β subunit[10], it was important to establish unambiguously that the β subunit was not derived from the larger α subunit by proteolytic cleavage. This question could be addressed by site-directed mutagenesis of the cloned aspartokinase II gene and transformation of a strain of *E. coli* lacking aspartokinase with plasmids carrying the altered aspartokinase II genes, followed by Western blot analysis of the synthesis of the aspartokinase II subunits[10]. The deletion of most of the coding region for α subunit (residues 566-1234) or of a single nucleotide (residue 660) early in the coding region for α subunit (causing a shift in the reading frame) completely abolished the synthesis of α subunit but failed to interfere with the production of β subunit (fig. 5), clearly demonstrating that β subunit was not derived from α subunit. To identify the translation-initiation site of β subunit, the nucleotides corresponding to the putative-initiation codon for β subunit (residues 1347-1349) were subjected to oligonucleotide-directed mutagenesis[11] and changed to sequences that would lead to conservative amino-acid substitutions for the methionine residue at position 247 of the α subunit but differ in their ability to encode peptide chain initiation by N-formylmethionine (fig. 6). Replacing ATG with either TTA or AAT had no effect on the synthesis of α subunit but completely abolished the synthesis of β subunit (fig. 5), indicating that the ATG sequence at residues 1347-1349 is indeed the site of peptide chain initiation. As expected, replacement of ATG by GTG had no effect on peptide chain initiation. Surprisingly, however, mutation to ATC also had no effect, indicating that AUC could substitute for AUG in encoding N-formylmethionine. Analogous experiments were carried out with transformed *B. subtilis* cells, with identical results (Chen and Paulus, unpublished experiments). A similar lack of specificity in peptide chain initiation has been reported[12] in the analogous situation of the bacteriophage f1 overlapping genes, gene II and gene X. A mutation of the internal gene X initiation codon from ATG to TTG failed to surpress the synthesis of the gene X polypeptide[12]. It may be that the presence of the very strong ribosome-binding site (AGGAGG) allows a single base substitution in the initiation codon to be tolerated—such tolerance has been described in a number of other systems[13].

Another feature of the Western blots shown in fig. 5 that deserves comment is the presence, besides the aspartokinase α and β subunits, of an immunoreactive polypeptide of M_r ~ 30,000. The absence of this ~ 30-kDa component in cells transformed with the deletion mutant pNC800 (lane 8) and its presence in cells transformed with the α subunit-deficient frame-shift mutant pNC860 (lane 7) indicated that it was translated from the α subunit coding region, but independently of α subunit. Examination of the aspartokinase II coding sequence[3] shows an internal methionine codon (residues 1026-1028) preceded by the putative weak ribosome binding site AGGA (residues 1013-1016).

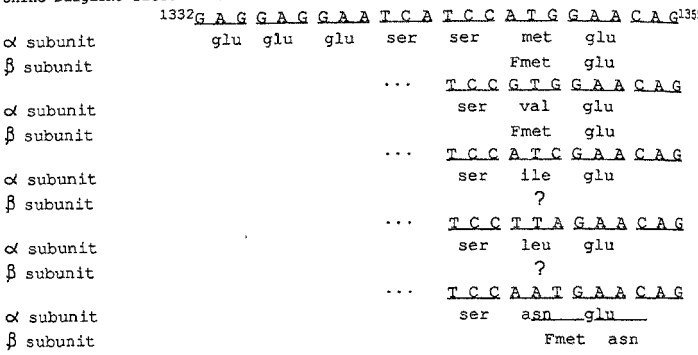
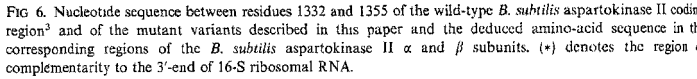


FIG 6. Nucleotide sequence between residues 1332 and 1355 of the wild-type *B. subtilis* aspartokinase II coding region[3] and of the mutant variants described in this paper and the deduced amino-acid sequence in the corresponding regions of the *B. subtilis* aspartokinase II α and β subunits. (*) denotes the region of complementarity to the 3'-end of 16-S ribosomal RNA.



A polypeptide initiated at that site would correspond to residues 139 through 408 of the α subunit and would have a molecular weight (28,983) and cross-reactivity with anti-β serum consistent with the properties of the unidentified ~ 30 kDa component. It is of interest that this polypeptide was not seen in analogous experiments with transformed *B. subtilis*. This suggests that the synthesis of the ~ 30 kDa component in *E. coli* may be an artifact due to the somewhat lower stringency of binding-site recognition by *E. coli* ribosomes compared to those of *B. subtilis*[14] and that this material is not produced when the aspartokinase II gene is expressed in *B. subtilis*, its physiological host.

6. Conclusions

Our results define the operon for *B. subtilis* aspartokinase II as a transcription unit with a number of unusual features. The operon is composed of two in-phase overlapping cistrons, a situation of special interest because the gene products are subunits of a single enzyme. 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*[15] by encoding a leader peptide analogous to those found in *E. coli*. The promoter overlaps in part with an adjacent operon, which in turn may use the transcription attenuator as its termination signal, whereas the transcription termination site of the aspartokinase II operon is shared with a converging operon. These unusual elements represent interesting examples of genetic economy and regulatory subtlety. In order to understand their functioning more fully, it will be necessary to modify them by deletion or site-directed mutagenesis and study their expression when reintegrated into the *B. subtilis* chromosome. Such experiments are now in progress in our laboratory.

Acknowledgements

This research was supported by grant DMB 84-14846 from the National Science Foundation.

References

1. PAULUS, H. — Regulation and structure of aspartokinase in the genus *Bacillus*, *J. Biosci.*, 1984, **6**, 403–418.

2. SANGER, F., COULSON, A. R., BARREL, B. G., SMITH, A. J. H. AND ROE, B. A. — Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing, *J. Mol. Biol.*, 1980, **143**, 161–178.

3. CHEN, N.-Y., HU, F.-M. AND PAULUS, H. — Nucleotide sequence of the overlapping gene for the subunits of *Bacillus subtilis* aspartokinase II and their control regions, *J. Biol. Chem.*, 1987, **262**, 8787–8798.

4. BONDARYK, R. P. AND PAULUS, H. — Cloning and structure of the gene for the subunits of aspartokinase II from *Bacillus subtilis*, *J. Biol. Chem.*, 1985, **260**, 585–591.

5. HU, M. C.-T. AND DAVIDSON, N. — Mapping transcription start points on cloned genomic DNA with T4 DNA polymerase: a precise and convenient technique, *Gene*, 1986, **42**, 21–29.

6. KOLTER, R. AND YANOFSKY, C. — Attenuation in amino acid biosynthetic operons, *Annu. Rev. Genet.*, 1982, **16**, 113–134.

7. GRUNDSTRÖM, T. AND JAURIN, B. — Overlap between *ampC* and *frd* operons on the *Escherichia coli* chromosome, *Proc. Natn. Acad. Sci. U.S.A.*, 1982, **79**, 1111–1115.

8. POSTLE, K. AND GOOD, R. F. — A bidirectional rho-independent transcription terminator between the *E. coli tonB* gene and an opposing gene, *Cell*, 1985, **41**, 577–585.

9. ADHYA, S. AND GOTTESMAN, M. — Control of transcription termination, *Annu. Rev. Biochem.*, 1978, **47**, 967–996.

10. BONDARYK, R. P., AND PAULUS, H. — Expression of the gene for *Bacillus subtilis* aspartokinase II in *Escherichia coli*, *J. Biol. Chem.*, 1985, **260**, 592–597.

11. ZOLLER, M. J. AND SMITH, M. — Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, *Meth. Enzymol.*, 1983, **100**, 468–500.

12. FULFORD, W. AND MODEL, P. — Gene X of bacteriophage f1 is required for phage DNA synthesis, *J. Mol. Biol.*, 1984, **178**, 137–153.

13. STORMO, G. D. Translation initiation, in *Maximizing gene expression* (W. Reznikoff and L. Gold, eds), Butterworths, 1986, pp. 195–224.

14. MACLAUGHLIN, J. R., MURRAY, C. L. AND RABINOWITZ, J. C. Unique features in the ribosome binding site sequence of the gram-positive *Staphylococcus aureus* β-lactamase gene, *J. Biol. Chem.*, 1981, **256**, 11283–11291.

15. SHIMOTSU, H., KURODA, M. I., YANOFSKY, C. AND HENNER, D. J. Novel form of transcription attenuation regulates expression of the *Bacillus subtilis* tryptophan operon, *J. Bacteriol.*, 1986, **166**, 461–471.