J. Biochem. 120, 425-432 (1996)

Aspartate Aminotransferase from an Alkalophilic *Bacillus* Contains an Additional 20-Amino Acid Extension at Its Functionally Important N-Terminus[1]

Natalia Battchikova,*,[2] Marianne Koivulehto,* Alexander Denesyuk,* Leonid Ptitsyn,† Yuri Boretsky,† Jukka Hellman,‡ and Timo Korpela*

**Finnish-Russian Joint Biotechnology Laboratory, University of Turku, BioCity 6A, FIN-20520 Turku, Finland; †Institute of Biotechnology, Nauchny pr. 8, Moscow 117246, Russia; and ‡Centre for Biotechnology, University of Turku, BioCity 5B, FIN-20520 Turku, Finland*

Received for publication, April 19, 1996

Aspartate aminotransferase (AspAT), responsible for a minor part of the total AspAT enzymic activity in alkalophilic *Bacillus circulans*, was purified, its N-terminal amino acid sequence was determined, and its gene was cloned as two separate fragments. DNA sequencing showed an open reading frame of 432 amino acids (M_r 47,439) exhibiting moderately low homology with AspATs from other sources. Sequence alignment of the enzyme with chicken mitochondrial, chicken cytoplasmic and *Escherichia coli* AspATs was performed with the MULTALIN program and further optimized assuming that the three-dimensional structures of the proteins were conserved. The primary structure of the studied AspAT diverged markedly from the others in the catalytically important small domain and in a segment of 31 amino acids in the large domain. The functional N-terminal arm was about two times longer than those of AspATs from other sources. According to the molecular model, the unique regions of *B. circulans* AspAT are all located together, forming a continuous network of contacts. Additional contacts formed by the elongated N-terminal arm may result in some limitation of domain movements in the alkalophilic enzyme in comparison to in other known AspATs.

Key words: aspartate aminotransferase, alkalophile, *Bacillus circulans*, cloning, protein sequence analysis.

During microbial growth under extremely alkaline conditions (pH 9-12), nutrient amino compounds can evaporate from solution, decompose, or react chemically with other biomolecules (*1, 2*). Thus, it is conceivable that nitrogen sources and their metabolism within alkalophilic microbes can be exceptional. Also, the structural characteristics of alkaline proteins making them stable and functional have also been studied only scantily, and they could have various scientific and applied consequences (*3, 4*).

Aspartate aminotransferase (AspAT) [EC 2.6.1.1] is a key enzyme in the nitrogen metabolism of all organisms. AspAT belongs to the α-family of pyridoxal 5'-phosphate (PLP)-dependent enzymes and catalyzes the reversible transfer of the amino group of aspartate or glutamate to oxoacids, 2-oxoglutarate or oxaloacetate, feeding through these reactions various metabolic pathways involving the amino functions (*5*). AspATs are α_2-dimeric proteins, with monomers comprising a large domain, a small domain and an amino-terminal arm (*6*). The active site for PLP is located near the subunit interface and can be covered or uncovered by the small domain ("open" and "closed" conformations of the active site). The N-terminus of the amino acid chain from one subunit protrudes toward the large domain of another one and has an important role in catalysis (*7, 8*). It seems likely that the enzymes of the α-family are organized according to the same basic three-dimensional structure, while the substrate specificity is acquired through rather small variations of the active site (*9*). AspAT can serve as an excellent model for studying specific structural features of proteins from alkalophilic organisms since AspATs have been isolated from different sources and their molecular structures are known in detail.

The present study is one of a series on an alkalophilic *Bacillus circulans* produced in our laboratory. Earlier (*10*) we showed that in this strain the majority of apparent AspAT activity in cells was due to the side activity of phosphoserine aminotransferase (PSAT) [EC 2.6.1.52]. The aim of the present study was to explore the fundamentals of the nitrogen metabolism of alkalophilic microbes and to find possible specific structural elements of the proteins from alkalophiles in general.

The overall deduced secondary and tertiary structures of the AspAT from the alkalophile apparently resemble those of other AspATs, but seem to be more rigid because of additional contacts formed by two α-helixes of the elongat-

[1] The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number, X94433.

[2] To whom correspondence should be addressed. Tel: +358-21-3338048, Fax: +358-21-3338080, e-mail: natbat@utu.fi

Abbreviations: AspAT, aspartate aminotransferase; PLP, pyridoxal 5'-phosphate; PSAT, phosphoserine aminotransferase; ORF, open reading frame; PCR, polymerase chain reaction.

ed N-terminal arm. Further structure-function studies will clarify their significance.

MATERIALS AND METHODS

Bacterial Strains and Plasmids—*B. circulans* subsp. *alkalophilus* (ATCC 21783) was used for the isolation of AspAT and chromosomal DNA for cloning of the AspAT gene. Cells were grown in medium, comprising 0.2% polypeptone, 0.5% yeast extract, 0.13% $K_2HPO_4 \cdot 3H_2O$, 0.02% $MgSO_4 \cdot 7H_2O$, 0.3% L-aspartate, 1% potato starch, and 1% Na_2CO_3 (*11*). Plasmids pUC18, pUC19 (Pharmacia, Sweden), pBC SK$^+$ (Stratagene, USA), and *Escherichia coli* XL1-blue [*recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac*, [F′ *proAB, lacI*^{q} *ZDM15, Tn10(tet*^{r})]] (Stratagene, USA) were used for cloning. *E. coli* cells were grown in LB medium (*12*) containing 12.5 μg/ml of tetracycline and, when necessary, ampicillin (50 μg/ml) or chloramphenicol (30 μg/ml) was added to the medium.

Isolation of AspAT—*B. circulans* cells were grown in a 10-liter fermentor and harvested in the late logarithmic phase, *i.e.* when the pH of the growth medium had decreased to 8.5–8.7. Washed cells (about 60 g wet weight) were suspended in 320 ml of buffer A (20 mM Tris-HCl, pH 7.6, 2 mM EDTA, 20 μM PLP, 0.1 mM DTT), followed by sonication and centrifugation at 20,000 × *g* for 20 min. The 35–75% ammonium sulphate fraction, containing more than 90% of the total AspAT activity, was dissolved in 40 ml of buffer B (50 mM Tris-HCl, pH 8.5, 20 μM PLP, 1 mM DTT), dialyzed against buffer B containing 0.35 M NaCl, and then applied to a DEAE-cellulose column (2.5 × 7 cm) equilibrated with the same buffer. The fractions containing AspAT activity were pooled, dialyzed against buffer B and then applied to a DEAE-Sepharose CL-6B column (2.5 × 25 cm) equilibrated with buffer B containing 0.08 M NaCl. The column was washed with 150 ml of the same buffer and the bound material was eluted with a 1.5-liter linear gradient of 0.08 to 0.5 M NaCl in buffer B. The fractions comprising the second AspAT active peak (0.25–0.33 M NaCl) were pooled and dialyzed against buffer A containing 0.15 M NaCl. The proteins were then loaded on a Q-Sepharose anion exchange column (1.6 × 12 cm) and eluted with a 400-ml linear gradient of 0.15 to 0.55 M NaCl in buffer A. The active fractions were pooled, and ammonium sulphate was added to the solution to 40% saturation. The proteins were loaded on a Butyl-Toyopearl HIC column (1.6 × 12 cm) and eluted with a 400-ml linear gradient of 40 to 0% of ammonium sulphate in buffer A. The active fractions were pooled, dialyzed against buffer C (5 mM K-phosphate buffer, pH 7.6, 5 μM PLP, 0.1 mM DTT) containing 20 mM NaCl, and then concentrated by ultrafiltration (PM-10 filter; Amicon). The enzyme solution was applied to a Sephadex G-75 gel filtration column (1.6 × 80 cm) equilibrated with buffer C. The active fractions were pooled and dialyzed against buffer C. The proteins were applied to a hydroxyapatite column (0.9 × 2 cm, Bio-Rad) and eluted with a 100-ml linear gradient of 5 to 200 mM K-phosphate buffer, pH 7.6, containing 0.3 mM $CaCl_2$. The enzyme thus purified was concentrated by ultrafiltration, dialyzed against buffer C and then stored at −20°C. AspAT activity was measured as described previously (*10*). Aromatic aminotransferase activity was measured as described by Hayashi *et al.* (*13*).

N-Terminal Sequence Analysis—The purified AspAT was desalted using a ProSpin cartridge (Applied Biosystems, USA), and the sample disc was placed on a Polybrene-coated and precycled glass fiber filter, followed by sequencing with an Applied Biosystems Model 477A protein sequencer equipped with an on-line model 120A PTH-amino acid analyzer. For internal sequencing, AspAT from the final purification step was electrophoresed on a 12.5% SDS-PAGE according to Moos *et al.* (*14*), and then transferred to a polyvinylidene difluoride (PVDF) membrane (ProBlott; Applied Biosystems, USA) by electroblotting. The membrane was stained with Coomassie Brilliant Blue, and then the desired band was excised and subjected to digestion with trypsin (sequencing grade; Boehringer Mannheim GmbH, Germany) as described by Fernandez *et al.* (*15*). The peptides were separated with a HPLC equipped with a Vydac C-18 column (2.1 mm × 150 mm). The tryptic peptides were applied to glass fiber filters and sequenced as described above.

General DNA Techniques—DNA manipulations and transformation of *E. coli* were accomplished according to Maniatis *et al.* (*12*). Restriction enzymes and T4 DNA ligase were purchased from Promega (USA). *Eco*RI, *Bam*HI, *Sac*I, *Cla*I, *Pst*I, *Hin*dIII, and *Sph*I were used for the digestion of chromosomal DNA from alkalophilic *B. circulans* cells. DNA restriction fragments were separated by agarose gel electrophoresis and transferred to a Hybond-N membrane (Amersham, UK). Oligonucleotides A1 [5′-ATGAA(C/T)CCI(C/T)TIGCIGGICA(A/G)(C/T)TIAA(C/T)GA(A/G)AA-3′] containing inosine (I), A2 (5′-GGAGCTGCGCTGTGTATGGC-3′), and *Bam*HI adapter (5′-CGGGATCCCG-3′) were synthesized by KEBO Lab (Finland). The A1 and A2 oligonucleotides were labeled with [γ-^{32}P]ATP and used as probes in Southern hybridization experiments carried out overnight at 55°C. β-Agarase (New England Biolabs, USA) was used for the isolation of DNA from agarose gels under the conditions recommended by the manufacturer. *Taq* DNA polymerase (Hytest, Finland) was used for polymerase chain reaction (PCR) experiments. Nucleotide sequencing was carried out by the method of Sanger *et al.* (*16*) using Sequenase 2.0 (USB, USA) and a TaqTrack sequencing kit (Promega, USA).

Computer Analysis—The BLAST network service at the National Center for Biotechnology Information was used for the search for sequence homology (*17*). The MULTALIN program, version 4.1 (*18*), and a principle of conservation of the secondary structures (α-helices and β-strands) were applied for multiple sequence comparison and alignment. The primary structures and the localization of the secondary structural elements of the chicken mitochondrial (7AAT) and cytosolic (2CST), and *E. coli* (1ASN) AspATs were obtained from the Brookhaven Protein Data Bank (*19*). Homology molecular modeling of *B. circulans* AspAT was performed using the Biopolymer module of software package SYBYL (Tripos Associates, USA) on a Silicon Graphics Iris Crimson workstation.

RESULTS AND DISCUSSION

Purification of AspAT—The proteins obtained on ammonium sulphate precipitation of a crude extract of *B. circulans* subsp. *alkalophilus* cells were subjected to anion exchange chromatography, followed by monitoring of

AspAT activity by a colorimetric method (*20*). Total AspAT activity was resolved into a major peak (60%), a minor peak (28%), and a fraction unbound to the DEAE-Sepharose column (12%). In the previous study (*10*) we showed that the major peak eluted with 0.15-0.22 M NaCl contained PSAT with considerably high AspAT activity. AspAT was purified from the minor peak eluted with 0.25-0.33 M NaCl. The purification procedure included ion exchange, hydrophobic interaction, size exclusion, and hydroxyapatite chromatography (Table I). The M_r of the protein was about 43,000, as judged on SDS-PAGE (data not shown). Sequencing was performed to analyze the primary structures of some AspAT fragments. The N-terminal amino acid sequence of the enzyme was determined to be MNPLAGQLNEKLQAG. The amino acid sequences of internal fragments obtained on trypsin digestion were ELYFP, EGILSQ, SLWR, LADLHPR, TLGIIR, and DILDSG.

Powell and Morrison demonstrated that the aromatic aminotransferase from *E. coli*, having prominent activity for transaminating aromatic amino and oxo acid substrates, shows side activity towards aspartate and oxaloacetate comparable to that of AspAT (*21*). The enzyme isolated from alkalophilic *B. circulans* was checked for activity towards aromatic amino acids by means of the tyrosine aminotransferase-specific assay (*13*). However, no tyrosine aminotransferase activity was detected.

Cloning of the AspAT Gene—The A1 oligonucleotide encoding eleven N-terminal amino acids of the protein was used for Southern hybridization. A restriction map of the AspAT gene constructed from data obtained is shown in Fig. 1. According to the map, the *Bam*HI-*Sph*I (4 kb) fragment was sufficiently long enough to contain the full gene of the 43 kDa protein independently of its orientation. Chromosomal DNA was digested with *Bam*HI and *Sph*I, and DNA fragments of about 4 kb were isolated from an agarose gel. They were ligated to the corresponding site of pUC18 or pUC19, and then *E. coli* XL1-blue was transformed with the ligation mixture. Hybridization with the radioactive A1 oligonucleotide showed that colonies containing the AspAT gene were not obtained. In earlier studies (*10*) we had a similar problem in attempts to clone the full gene of PSAT from the same microorganism. We only succeeded in cloning the gene as two fragments, with intermediate sequencing of a PCR product. The same approach was used here. *Bam*HI-*Sph*I fragments of the chromosomal DNA described above were further digested with *Hin*dIII and *Sac*I, and DNA fragments of 400-700 bp were isolated from an agarose gel, followed by ligation into pUC18 and transformation. Hybridization of colonies with the radioactive A1 oligonucleotide resulted in seven positive colonies among about 3,000. Only two of the seven colonies remained positive when hybridization was repeated. Plasmids isolated from them had about 2 times higher molecular weights than expected. Restriction with *Hin*dIII and *Sac*I revealed that the plasmids contained two *Hin*dIII-*Sac*I fragments (about 500 and 700 bp in the first plasmid, and about 500 and 350 in the second one) per molecule. Fragments of about 500 bp from both initial colonies were isolated from an agarose gel and cloned into pUC18. The resulting colonies grew slowly and yielded a small amount of plasmid DNA. The reasons for this remain obscure. Sequencing of inserts showed that the *Hin*dIII-*Sac*I fragments from the two initial colonies were identical and comprised 460 bp. Sequence analysis of the fragments revealed the beginning of an open reading frame (ORF) for AspAT interrupted by a *Sac*I site.

The *Sac*I-*Cla*I fragment encoding the rest of the protein was cloned as follows. A ligation mixture of *Bam*HI-*Sph*I chromosomal fragments and the pUC18/*Bam*HI-*Sph*I vector was used as a template for PCR carried out in the presence of the A1 oligonucleotide and the M13 universal sequence primer. The reaction resulted in a PCR product of about 2.5 kb. Restriction analysis indicated the correlation between the PCR product and the AspAT gene (data not shown). The PCR product was trimmed with T4 DNA polymerase and the annealed self-complementary *Bam*HI adapter was ligated to its ends. The fragment was digested with *Bam*HI and *Cla*I, and then cloned in pBC SK$^+$. Partial sequencing of inserts from four colonies revealed differences in their structures apparently caused by the insufficient accuracy of the *Taq* DNA polymerase. However, it became evident that the rest of the AspAT chromosomal

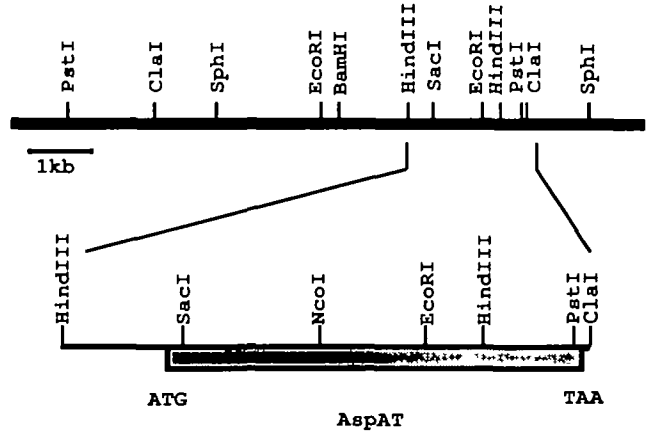

Fig. 1. **Restriction map of the chromosomal fragment containing the alkalophilic *B. circulans* AspAT gene.** The region for which the nucleotide sequence was determined is shown on a larger scale. The box denotes the position of the ORF of AspAT.

TABLE I. **Purification of AspAT from *B. circulans* subsp. *alkalophilus*.** Total AspAT activity was resolved into a major peak (60%, PSAT), a minor peak (28%, AspAT), and a fraction unbound to a DEAE-Sepharose column. The minor peak containing AspAT is considered in the table as the starting point of the purification procedure. AspAT activity was measured as described previously (*10*).

Step	Protein (mg)	Total activity (U)	Specific activity (U/mg)	Yield (%)	Purification factor
DEAE-Sepharose, minor peak	180	6.1	0.033	100	1
Q-Sepharose	38.2	5.0	0.13	82	3.9
Butyl-Toyopearl	4.2	3.4	0.81	56	24.5
Sephadex G-75	2.0	3.2	1.6	52	48.5
Hydroxyapatite	0.1	2.0	20.0	33	606

```
      HindIII
1     AAGCTTCAGCCGGAAGCGGCTATGCTACGATTGATGCACAGGACCCGAAGTATGCCAAAA   60

61    GGCTGAAAGGGGTCGCCGTCATCAAGCTGGTGCCCCATCGGCTGAGCGCAAAATTCAAAT  120

121   TCGGGCAAAATCTGACACCGGAGCGTCTGGAGAGCATCGCGCAGGCGCTCGAAAGCCGGG  180

181   ATCAGGAGCAGGATCGCGAAACCGCAAAACTAATGCGAAAATATTTCACGGGATCGCATT  240

241   AAAGCGACGCGAATTGATCGTGACGGTTGATGTGCAAGGTGGTATAATGTTAGGCAATGA  300

301   ATCGGAGACATTACATCCAATGTCGTCATAATATCCAATAGAATTGCGGAAGGAAGGAGC  360

        A1 ------------------------------->
361   CAATTCATGAACCCATTGGCCGGACAATTGAATGAGAAGCTGCAAGCCGGTAACCCGCAT  420
           M  N  P  L  A  G  Q  L  N  E  K  L  Q  A  G  N  P  H

                                  SacI
421   GTATTTGACATGCTCTCCACACTCGGCAAGGAGCTCTATTTCCCGAAGGAAGGCATTCTG  480
      V  F  D  M  L  S  T  L  G  K  E  L  Y  F  P  K  E  G  I  L

481   AGCCAATCCGCTGAGGCAGCAGCCCATGCGAAAAAATATAATGCAACGATTGGAATCGCA  540
      S  Q  S  A  E  A  A  A  H  A  K  K  Y  N  A  T  I  G  I  A

541   ACCGAAGGCGGCATTCCTATGCATCTGGGGGTCATTCAGGAGAAGCTGTCGGCTTTCGCT  600
      T  E  G  G  I  P  M  H  L  G  V  I  Q  E  K  L  S  A  F  A

                                        A2 -------------------
601   CCGAAGGACCTGTACCCGTACGCCCCTCCAGCCGGCAAGCCGGAGCTGCGCTCTCTATGG  660
      P  K  D  L  Y  P  Y  A  P  P  A  G  K  P  E  L  R  S  L  W

      >
661   CGGGACAAGATGCTGGAAGAGAACCCTTCGCTGCAAGGGAAGAGCTTCGGCAATCCGATT  720
      R  D  K  M  L  E  E  N  P  S  L  Q  G  K  S  F  G  N  P  I

721   GCAACCAATGCACTGACCCACGGTCTCAGCATCGTCGCTGACCTGTTCGTAGATGAAGGC  780
      A  T  N  A  L  T  H  G  L  S  I  V  A  D  L  F  V  D  E  G

781   GATGCCGTCATCTATCCCGACAAGAACTGGGAAAACTATGACCTGACATTCGGCGTCCGC  840
      D  A  V  I  Y  P  D  K  N  W  E  N  Y  D  L  T  F  G  V  R

      NcoI
841   CGCCATGGGGTGAACGTGAATTTCCCTCTTTTCACATCGGAGATGACCTTTAATGCCGCA  900
      R  H  G  V  N  V  N  F  P  L  F  T  S  E  M  T  F  N  A  A

901   GGTTTGCGCGACGCCCTTCTGGCCCAGAAGGATAAGGGGAAAGCCGTGGTCATCCTCAAC  960
      G  L  R  D  A  L  L  A  Q  K  D  K  G  K  A  V  V  I  L  N

961   TTCCCGAACAACCCGACAGGCTACACACCTGGGGCAGCCGAGGGAGACGCCATCGTCGCA 1020
      F  P  N  N  P  T  G  Y  T  P  G  A  A  E  G  D  A  I  V  A

1021  GCGATCAAGGATGCGGCCGAAGCGGGAATCAATGTCGTTGTGGTTACGGATGATGCGTAT 1080
      A  I  K  D  A  A  E  A  G  I  N  V  V  V  V  T  D  D  A  Y

1081  TTCGGCCTGTTCTTTGAAGATTCGCTCAAGGAATCGCTGTTTGGCCGCCTGGCCGATCTG 1140
      F  G  L  F  F  E  D  S  L  K  E  S  L  F  G  R  L  A  D  L

                                                EcoRI
1141  CATCCGCGCGTGCTGGCGGTGAAGGTGGATGGAGCCACGAAGGAGGAATTCGTCTGGGGA 1200
      H  P  R  V  L  A  V  K  V  D  G  A  T  K  E  E  F  V  W  G

1201  TTCCGTGTCGGCTTCATTACCTATGCTTCGGAGAACAAGGACGTGCTGGATGCCCTGGAG 1260
      F  R  V  G  F  I  T  Y  A  S  E  N  K  D  V  L  D  A  L  E

1261  CAGAAGACGCTCGGCATCATCCGCGCAACCATCTCCAGCGGTCCCCACCCTTCGCAGACC 1320
      Q  K  T  L  G  I  I  R  A  T  I  S  S  G  P  H  P  S  Q  T

                                                      HindIII
1321  TTTGTGCTGGATGCGTTGAAGGCACCGGGATTCAAGGAGCAGAAGGAAGAGAAGCTTCAG 1380
      F  V  L  D  A  L  K  A  P  G  F  K  E  Q  K  E  E  K  L  Q

1381  ATTATGAAAGGCCGCGCCAATAAAGTGAAGGACATTCTGGACAGCGGCAAATATAGCGAA 1440
      I  M  K  G  R  A  N  K  V  K  D  I  L  D  S  G  K  Y  S  E

1441  GCATGGgACTACTACCCGTTCAACTCCGGATACTTCATGTGTCTGAAGCTCAAGACGGTC 1500
      A  W  D  Y  Y  P  F  N  S  G  Y  F  M  C  L  K  L  K  T  V

1501  CAGGCTGAAGCGCTCCGCTCCCATCTGATTCACGAGTACGGGATCGGCACCATTGCCTTG 1560
      Q  A  E  A  L  R  S  H  L  I  H  E  Y  G  I  G  T  I  A  L

1561  GGCGAACATGATCTGCGGATCGCATTCTCTTGCATTGAGGAGCCATATTTGGAGGATCTG 1620
      G  E  H  D  L  R  I  A  F  S  C  I  E  E  P  Y  L  E  D  L

                                        PstI
1621  TTTGATTTGATTCACCAGGGTGTTCAGGATCTGCAGCAGGCGTAATCACGCAGTGACCGG 1680
      F  D  L  I  H  Q  G  V  Q  D  L  Q  Q  A  *

                  ClaI
1681  ACACACGCGGTGATGATCGAT  1701
```

Fig. 2. **Nucleotide and deduced amino acid sequences of the *B. circulans* AspAT gene and its flanking regions.** The putative ribosome binding site is underlined. The underlined amino acids were determined by protein sequencing of the N-terminus and tryptic fragments of the purified AspAT. The positions of the A1 and A2 oligonucleotides used for hybridization and/or PCR are shown by dashed lines.

gene can be cloned as a *Sac*I-*Cla*I fragment since the *Sac*I site was unique and the *Hin*dIII-*Cla*I fragment contained stop codons in all three frames. Moreover, the partial sequence of the *Sac*I-*Cla*I fragment provided information for the synthesis of a hybridization probe (A2) for this fragment. The *Bam*HI-*Sph*I fragments of chromosomal DNA described above were digested with *Sac*I and *Cla*I, and fragments of about 1.3 kb were isolated from an

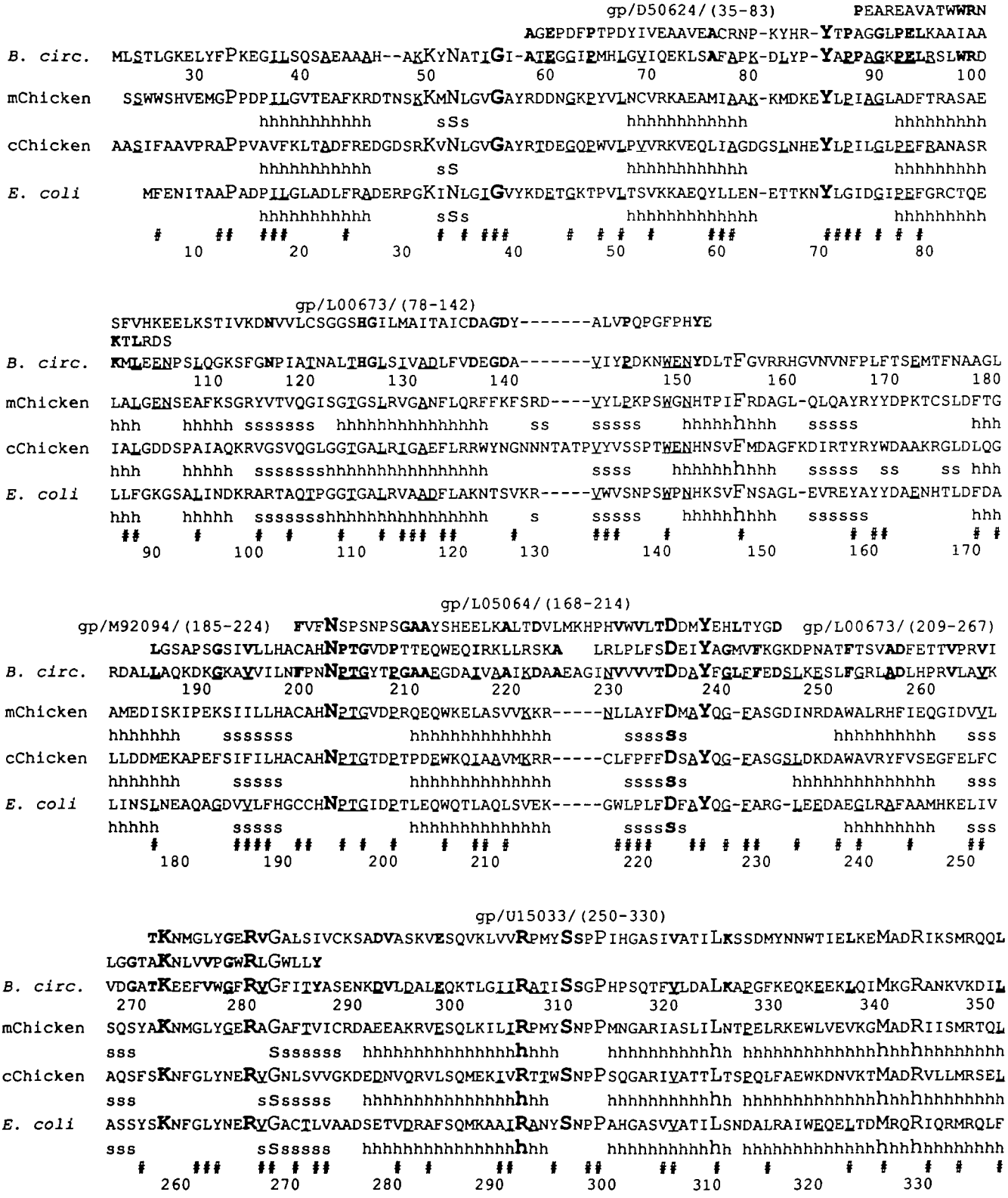

Fig. 3 (continued on next page)

agarose gel and cloned into pBC SK$^+$. More than 30% of the colonies obtained were positive as to hybridization with the A2 probe. The *Sac*I-*Cla*I insert from one of the plasmids was sequenced. The data obtained allowed us to establish the full nucleotide sequence of the AspAT gene and to deduce the AspAT amino acid sequence (Fig. 2).

Nucleotide and Protein Sequence Analyses—The ORF for AspAT starts from ATG (position 363) and terminates with TAA (position 1659). All tryptic fragments determined by protein sequencing were found in the deduced amino acid sequence (Fig. 2). A putative ribosome-binding site, AAGGaAGG, is 10 bp upstream of the ATG start codon. This sequence can be bound to the 3′-end of 16S rRNA from *B. subtilis* with a free energy of −62 kJ/mol (*22*). The G+C content of the AspAT gene was 55%, which is about 6% higher than the total genomic G+C content (48.8%) of this alkalophile (*3*). The ORF consists of 1,296 bp and encodes a protein of 432 amino acids. The calculated M_r of this protein (47,439) is slightly higher than the value obtained on SDS-PAGE (about 43,000).

Comparison and Alignment of AspAT Sequences—The deduced amino acid sequence of *B. circulans* AspAT was used to search for sequence homology using the BLAST network service (*17*). The program revealed many segments of various aminotransferases similar to the central part of *B. circulans* AspAT. Six intersecting segments of lengths from 40 to 81 amino acids (from four different aspartate and one tyrosine aminotransferase) were chosen on the basis of the highest length and homology to distinct regions of the enzyme (Fig. 3, upper lines of the alignment). The identity and similarity of these six fragments with corresponding peptides of *B. circulans* AspAT were 16-31 and 40-50%, respectively. It is interesting that many of these segments belong to nitrogen-fixating plants, perhaps indicating an interchange of genetic materials.

B. circulans AspAT shows moderately low (19%) overall sequence homology with 33 AspATs in the Swiss-Prot Protein Sequence Data Bank, Release 32. Their sequence alignment is complicated since standard programs do not work correctly if the sequence identity is less than 30% (*26*). Previous studies on AspATs strongly suggested that AspATs resemble each other in their three-dimensional structures despite their low levels of sequence homology (*27*). Therefore, we included in the alignment only those AspATs for which three-dimensional and secondary structures are known based on the data in the Brookhaven Protein Data Bank (*19*). The alignment was made with the MULTALIN program (*18*) and further optimized assuming that the secondary structure of the proteins was conserved, which was a necessary condition to obtain a correlation to the three-dimensional structures.

In general, the secondary structure and its localization in the protein sequence depends on the arrangement of hydrophobic amino acids. The conservative hydrophobic positions and the secondary structure of AspATs are illustrated in Fig. 3. All these hydrophobic residues are located in either the regions containing α-helices and β-strands or the loops around functionally important amino acids. For example, clusters of four or five hydrophobic residues exist around Tyr70, Asn194, and Tyr225, and between Lys258 and Arg266 (here and hereafter the numeration corresponds to pig cytosolic AspAT; see Refs. *23* and *24*). In the α-helix regions, hydrophobic amino acids must be placed at positions i and $i+3$ or $i+4$, and in the β-strand at i and $i+1$ or $i+2$. Figure 3 shows that the conservative hydrophobic positions do, indeed, show such regularity and thus the *B. circulans* AspAT has a similar secondary structure to the other three AspATs. Practically all deletions or insertions in *B. circulans* AspAT are at sites corresponding to the loops in the other three AspATs. The 21 N-terminal amino acids of the enzyme form an extension, which is not shown in the Fig. 3.

The alignment of *B. circulans*, chicken mitochondrial, chicken cytosolic, and *E. coli* AspATs in Fig. 3 reveals 25 invariant amino acids, which include residues involved in the binding of PLP (Tyr70, Asn194, Asp222, Tyr225, Lys258, Arg266, and Ser296) and residues (Gly38, Asn194, Arg292, and Arg386) participating in substrate binding (*25, 28*). Thr109 and Trp140, providing hydrogen bonds to the phosphate group and distal carboxyl group of

```
B. circ.  DSGKYSEA----WDYYPFNSGYFMCLKLKTVQAEALRSHLIHEYGIGTIALGEHDLRIAFSCIEEPYLEDLFDLIHQGVQDLQQA
                    360       370       380       390       400       410       420       430
mChicken  VSNLKKEGSSHNWQHITDQIGMFCFTGLK----PEQVERLTKEFSIYMT----KDGRISVAGVASSNVGYLAHAIHQ-VTK
          hhhhhhh     hhhhh   sSsss      hhhhhhhhhhhsss      ssSsss     hhhhhhhhhhh hh
cChicken  RSRLESLGTPGTWNHITDQIGMFSFTGLN----PKQVEYMIKEKHIYLM----ASGRINMCGLTTKNLDYVAKSIHEAVTKIQ
          hhhhhhh     hhhhh    Sss       hhhhhhhhhhh ss       Sss       hhhhhhhhhhhhhhhh
E. coli   VNTLQEKGANRDFSFIIKQNGMFSFSGLT----KEQVLRLREEFGVYAV----ASGRVNVAGMTPDNMAPLCEAIV-AVL
          hhhhhh      hhhhh   sSss      hhhhhhhhhhhsss      Ssss     hhhhhhhhhhh h
                 #    #  #    ### #  #          #     ##       # # # ##    #  ##   #   #
          340       350       360       370       380       390       400       409
```

Fig. 3. **Sequence alignment of *B. circulans*, chicken mitochondrial, chicken cytosolic, and *E. coli* AspATs.** α-Helices (h) and β-strands (s) in the corresponding three-dimensional structures are indicated. The numeration below the sequences refers to the sequence of pig cytosolic AspAT (*23, 24*). The two upper lines show six segments of different aminotransferases chosen on the basis of the highest scores of homology to distinct regions of *B. circulans* AspAT among data obtained through the BLAST search (*17*): gp/D50624/−*Streptomyces virginiae* AspAT, gp/L00673/−*Trypanosoma cruzi* tyrosine aminotransferase, gp/M92094/−*Lupinus angustifolius* AspAT, gp/L05064−*Rhizobium meliloti* AspAT, gp/U15033/−*Arabidopsis thaliana* AspAT. The identification codes refer to the cumulative daily updates to the major release of CDS translations from GenBank (R) Release 90, in which the complete sequences of these aminotransferases are available. The bold letters denote identical residues in *B. circulans* and the segments above. The 25 invariant positions (large letters), including 10 amino acids (large and bold letters) important for catalytical activity (*25*), are indicated (see the text). The identical residues in *B. circulans* and one or more of the chicken mitochondrial, chicken cytosolic, and *E. coli* AspATs are underlined by a single line. Conservative hydrophobic positions of these AspATs are also indicated (#). The alignment does not contain the 21 amino acids of the N-terminus of the *B. circulans* enzyme.

the substrate (*25, 28*), respectively, are also conserved among the four AspATs. In addition to these 25 invariant positions, there are 90 identical amino acids (underlined by a single line) between the *B. circulans* enzyme and at least one of the other three AspATs. The 115 conservative positions are uniformly distributed over the whole primary sequence with the exception of region 145-176.

According to the alignment, Lys272 is the residue in *B. circulans* AspAT which binds to PLP since it corresponds to Lys258 in pig cytosolic AspAT. The sequence around Lys272 (GATKEEFVWGFRVG) in *B. circulans* AspAT differs from the consensus pattern [GS], [LIVMFYTC]-[SA]-K-x(2)-[GSALV]-[LIVMF]-x-[GNAR]-x-R-[LIVMA]-G, common to all other AspATs (*29*). However, the changes at residues 256, 257, and 261 are conservative, and do not considerably disturb the pattern.

The primary structure of *B. circulans* AspAT essentially differs from those of the other AspATs in the N-terminus, region 145-176, and the C-terminus. No aminotransferase with homology to these regions was detected with the BLAST network service. Since the three-dimensional structures of three AspATs in the alignment are known, it is possible to predict where differences are located. The N-terminal arm of each AspAT subunit contains generally about 15 residues (*25*). The small domain consists of about thirty N- and eighty C-terminal residues. Thus, the unique amino acid sequences of *B. circulans* AspAT (see above) can be localized in the N-terminal arm (notice that it is elongated by 35 residues), the whole small domain, and in region 145-176 of the large domain.

The overall similarity of the deduced secondary and tertiary structures of *B. circulans* enzyme with other AspATs enabled us to build a model of the alkalophilic enzyme based on AspAT of *E. coli* (Fig. 4). The differences between the model and the *E. coli* enzyme are in the region of the N-terminal arms. Segment 5-15 of *E. coli* AspAT has an irregular conformation (*25*). However, prediction of the N-terminal arm secondary structure of the *B. circulans* enzyme by means of the nearest-neighbor and segment-oriented algorithms (*31, 32*) revealed two α-helices: "a" (MNPLAGQLNEKLQA) and "b" (HVFDMLSTLGKELY), connected by a short loop. Hence, the "b" α-helix replaces irregular segment 5-15, while the "a" α-helix runs in the opposite direction to "b." The position of "a" shown in Fig. 4 is optimal because of the short length of the "a"-"b" loop (three residues), and because of the maximal number of interhelical hydrophobic contacts which the "a" helix can form with the "b" helix as well as with the α-helix 142-150 of the large domain of the same subunit. In addition, due to this disposition, "a" is in contact with α-helix 107*-122* of the large domain of the other subunit, playing the role of an additional binding "cramp-iron" between subunits. Note, that α-helices 107*-122* and 142-150 interact, and the latter belongs to region 145-176 having the unique amino acid sequence. Thus, the unique regions of *B. circulans* AspAT are all located together forming a continuous network on the surface of the dimeric molecule. Apparently, this results in additional stabilization of the conformation and possibly some limitation of domain movement. The influence of deletions of the N-terminal arm of pig cytosolic AspAT on enzymatic activity has been investigated (*33*). The activity of mutants with 1-7 or 1-9 residues deleted showed greatly increased K_m values for dicarboxylic substrates and a markedly decreased k_{cat}/K_m value (*33*).

Since AspAT of the alkalophilic *Bacillus* is an intracellular enzyme, it is assumed from the results of various experiments that the internal pH is 1-2 pH units higher inside alkalophile than neutrophil cells (*3*). In addition, there is a tendency for intracellular enzymes of alkalophiles to have higher pH optima of related magnitude. Hence, the hydrogen ion concentrations are 10-100 times less around alkalophilic enzymes than their neutral counterparts. This should be reflected in the composition of catalytically

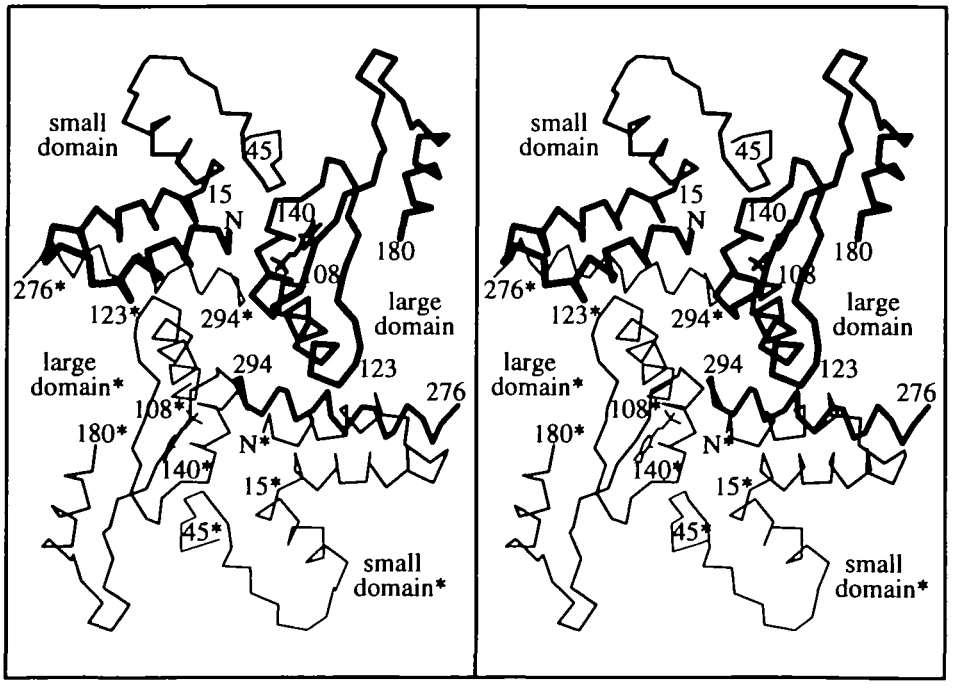

Fig. 4. **Stereo view of a part of a modeled dimeric *B. circulans* AspAT showing the interaction of the two α-helices, "a" and "b" (between N and 15; see the text), with the small and large domains.** The separate subunits are indicated by bold and thin lines. The figure was produced with MOLSCRIPT (*30*).

important amino acids and those connecting to the catalytical network. It can be speculated that the unique structure of the small domain and the extension of the N-terminus change the chemical milieu at the active site to maintain the apparent pK of the functionally important ionic residues. Detailed characterization of AspAT from alkalophilic *B. circulans* and of a mutant of it lacking the N-terminal extension are now underway.

REFERENCES

1. Volkin, D.V. and Klibanov, A.M. (1989) Minimizing protein inactivation in *Protein Function. A Practical Approach* (Creighton, T.E., ed.) pp. 1-24, IRL Press, Oxford, England
2. Mäkelä, M., Mattsson, P., Schinina, M.E., and Korpela, T. (1988) Purification and properties of cyclomaltodextrin glucanotransferase from an alkalophilic *Bacillus*. *Biotechnol. Appl. Biochem.* **10**, 414-427
3. Horikoshi, K. ed. (1991) *Microorganisms in Alkaline Environments*, Kodansha, Tokyo
4. Krulwich, T.N. and Guffanti, A.A. (1989) Alkalophilic bacteria. *Annu. Rev. Microbiol.* **43**, 435-463
5. Christen, P. and Metzler, D.E., eds. (1985) *Transaminases*, John Wiley and Sons, New York
6. Ford, G.C., Eichele, G., and Jansonius, J.N. (1980) Three-dimensional structure of a pyridoxal-phosphate dependent enzyme, mitochondrial aspartate aminotransferase. *Proc. Natl. Acad. Sci. USA* **77**, 2559-2563
7. Picot, D., Sandmeier, E., Thaller, C., Vincent, M.G., Christen, P., and Jansonius, J.N. (1991) The open/closed conformational equilibrium of aspartate aminotransferase. Studies in the crystalline state and with a fluorescent probe in solution. *Eur. J. Biochem.* **196**, 329-341
8. McPhalen, C.A., Vincent, M.G., and Jansonius, J.N. (1992) X-ray structure refinement and comparison of three forms of mitochondrial aspartate aminotransferase. *J. Mol. Biol.* **225**, 495-517
9. Alexander, F.W., Sandmeier, E., Mehta, P.K., and Christen, P. (1994) Evolutionary relationships among pyridoxal-5′-phosphate-dependent enzymes. Regio-specific α, β and γ families. *Eur. J. Biochem.* **219**, 953-960
10. Battchikova, N., Himanen J.-P., Ahjolahti, M., and Korpela, T. (1996) Phosphoserine aminotransferase from *Bacillus circulans* subsp. *alkalophilus*: Purification, gene cloning and sequencing. *Biochim. Biophys. Acta* (in press, RPP AM011195)
11. Mäkelä, M.J., Paavilainen, S.K., and Korpela, T.K. (1990) Growth dynamics of cyclomaltodextrin glucanotransferase-producing *Bacillus circulans* var. *alkalophilus*. *Can. J. Microbiol.* **36**, 176-182
12. Maniatis, T., Fritsh, E.F., and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
13. Hayashi, H., Inoue, K., Nagata, T., Kuramitsu, S., and Kagamiyama, H. (1993) *Escherichia coli* aromatic amino acid aminotransferase: Characterization and comparison with aspartate aminotransferase. *Biochemistry* **32**, 12229-12239
14. Moos, M., Jr., Nguyen, N.Y., and Liu, T. (1988) Reproducible high yield sequencing of proteins electrophoretically separated and transferred to an inert support. *J. Biol. Chem.* **263**, 6005-6008
15. Fernandez, J., DeMott, M., Atherton, D., and Mische, S.M. (1992) Internal protein sequence analysis: Enzymatic digestion of less than 10 μg of protein bound to polyvinylidene difluoride or nitrocellulose membranes. *Anal. Biochem.* **201**, 255-264
16. Sanger, F., Nicklen, S., and Coulson, A. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467
17. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410
18. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881-10890
19. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.J., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542
20. Tanaka, T., Tokuda, T., Tachibana, A., Taniguchi, M., and Oi, S. (1990) Purification and some properties of aspartate aminotransferase of *Methanobacterium thermoformicicum* SF-4. *Agric. Biol. Chem.* **54**, 625-631
21. Powell, J.T. and Morrison, J.F. (1978) The purification and properties of the aspartate aminotransferase and aromatic-amino-acid aminotransferase from *Escherichia coli*. *Eur. J. Biochem.* **87**, 391-400
22. Tinoco, I., Borer, P.N., Dengler, B.D., Levine, M.D., Uhlenbeck, O.C., Crothers, D.M., and Gralla, J. (1973) Improved estimation of the secondary structure of ribonucleic acids. *Nature New Biol.* **246**, 40-41
23. Ovchinnikov, Y.A., Egorov, T.A., Aldanova, N.A., Feigina, M.Y., Lipkin, V.M., Abdulaev, N.G., Grishin E.V., Kiselev, A.P., Modyanov, N.N., Braunstein, A.E., Polyanovsky, O.L., and Nosikov, V.V. (1973) The complete amino acid sequence of cytoplasmic aspartate aminotransferase from pig heart. *FEBS Lett.* **29**, 31-34
24. Doonan, S., Doonan, H.J., Hanford, R., Vernon, C.A., Walker, J.M., Airoldi, L.P. daS., Bossa, F., Barra, D., Carloni, M., Fasella, P., and Riva, F. (1975) The primary structure of aspartate aminotransferase from pig heart muscle. Digestion with a proteinase having specifity for lysine residues. *Biochem. J.* **149**, 497-506
25. Jäger, J., Moser, M., Sauder, U., and Jansonius, J.N. (1994) Crystal structures of *Escherichia coli* aspartate aminotransferase in two conformations. Comparison of an unliganded open and two liganded closed forms. *J. Mol. Biol.* **239**, 285-305
26. Mehta, P.K., Hale, T.I., and Christen, P. (1993) Aminotransferases: Demonstration of homology and division into evolutionary subgroups. *Eur. J. Biochem.* **214**, 549-561
27. John, R.A. (1995) Pyridoxal phosphate-dependent enzymes. *Biochim. Biophys. Acta* **1248**, 81-96
28. Okamoto, A., Higushi, T., Hirotsu, K., Kuramitsu, S., and Kagamiyama, H. (1994) X-ray crystallographic study of pyridoxal 5′-phosphate-type aspartate aminotransferases from *Escherichia coli* in open and closed forms. *J. Biochem.* **116**, 95-107
29. Bairoch, A. (1992) The SWISS-PROT protein sequence data bank. *Nucleic Acids Res.* **20**, supplement, 2013-2018
30. Kraulis, P.J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallogr.* **24**, 946-950
31. Solovyev, V.V. and Salamov, A.A. (1994) Predicting α-helix and β-strand segments of globular proteins. *Comput. Appl. Biosci.* **10**, 661-669
32. Salamov, A.A. and Solovyev, V.V. (1995) Prediction of protein secondary structures by combining nearest-neighbor algorithms and multiply sequence alignments. *J. Mol. Biol.* **247**, 11-15
33. Tanase, S., Fukumoto, Y., Pan, Q.-W., Nagashima, F., and Morino, Y. (1991) in *Enzymes Dependent on Pyridoxal Phosphate and Other Carbonyl Compounds as Cofactors* (Fukui, T., Kagamiyama, H., Soda, K., and Wada, H., eds.), Pergamon Press, New York