Reprinted from Biochemistry, 1991, 30.

Complete Amino Acid Sequence of Rat Liver Cytosolic Alanine Aminotransferase†

Mariko Ishiguro,‡,§ Masami Suzuki,‡ Koji Takio,§ Takeo Matsuzawa,‡ and Koiti Titani*,‡,§

Division of Biomedical Polymer Science, Institute for Comprehensive Medical Science, and Department of Biochemistry, School of Medicine, Fujita Health University, Toyoake, Aichi, Japan 470-11, and Laboratory for the Aging Process Research, Frontier Research Program, Institute of Physical and Chemical Research (RIKEN), Wako, Saitama, Japan 351-01

Received January 29, 1991; Revised Manuscript Received March 20, 1991

ABSTRACT: The complete amino acid sequence of rat liver cytosolic alanine aminotransferase (EC 2.6.1.2) is presented. Two primary sets of overlapping fragments were obtained by cleavage of the pyridylethylated protein at methionyl and lysyl bonds with cyanogen bromide and *Achromobacter* protease I, respectively. The protein was found to be acetylated at the amino terminus and contained 495 amino acid residues. The molecular weight of the subunit was calculated to be 55 018 which was in good agreement with a molecular weight of 55 000 determined by SDS–PAGE and also indicated that the active enzyme with a molecular weight of 114 000 was a homodimer composed of two identical subunits. No highly homologous sequence was found in protein sequence databases except for a 20-residue sequence around the pyridoxal 5′-phosphate binding site of the pig heart enzyme [Tanase, S., Kojima, H., & Morino, Y. (1979) *Biochemistry 18*, 3002–3007], which was almost identical with that of residues 303–322 of the rat liver enzyme. In spite of rather low homology scores, rat alanine aminotransferase is clearly homologous to those of other aminotransferases from the same species, e.g., cytosolic tyrosine aminotransferase (24.7% identity), cytosolic aspartate aminotransferase (17.0%), and mitochondrial aspartate aminotransferase (16.0%). Most of the crucial amino acid residues hydrogen-bonding to pyridoxal 5′-phosphate identified in aspartate aminotransferase by X-ray crystallography are conserved in alanine aminotransferase. This suggests that the topology of secondary structures characteristic in the large domain of other α-aminotransferases with known tertiary structure may also be conserved in alanine aminotransferase.

Alanine aminotransferase (GPT)[1] (EC 2.6.1.2) is a pyridoxal enzyme which catalyzes reversible transamination between alanine and 2-oxoglutarate to form pyruvate and glutamate. This enzyme localizes in both cytosol and mitochondria in various tissues including liver, heart, and skeletal muscle. The cytosolic enzyme is inducible by glucocorticoid administration, but the mitochondrial enzyme is not (Hopper & Segal, 1964). These isoenzymes participate in nitrogen metabolism and liver gluconeogenesis starting with precursors generated by muscular action (DeRosa & Swick, 1975). The enzyme is one of the two important aminotransferases, GOT and GPT, in clinical aspects, because the enzyme in blood is a representative diagnostic marker of human liver diseases. Physicochemical properties and turnover of the cytosolic enzyme have been described in a number of reports (Segal & Kim, 1963; Hopper & Segal, 1962; Bulos & Handler, 1965; Saier & Jenkins, 1967a,b; Gatehouse et al., 1967; Matsuzawa & Segal, 1968; Golichowski et al., 1977). The relationships between protein structure and function have been extensively studied with GOT (Jansonius et al., 1985), but only a little with GPT. So far, only a 20 amino acid residue sequence around the pyridoxal 5′-phosphate (PLP) binding site of the pig heart enzyme has been reported by Tanase et al. (1979). The best preparation purified from rat liver by previous methods (Gatehouse et al., 1967; Matsuzawa & Segal, 1968) still showed two other minor components as judged by polyacrylamide gel electrophoresis. Recently, purification of the cytosolic enzyme from rat liver to absolute homogeneity as judged by SDS–PAGE was accomplished (Matsuzawa et al., unpublished results) by a modification of previous methods (Gatehouse et al., 1967; Matsuzawa & Segal, 1968). We herein report the complete amino acid sequence of the enzyme and discuss its structural resemblance to other aminotransferases.

Materials and Methods

Materials. Rat liver cytosolic GPT was purified to homogeneity as judged by SDS–PAGE with a modification of previous methods (Gatehouse et al., 1967; Matsuzawa & Segal, 1968) by octyl-Sepharose hydrophobic interaction chromatography, Sepharose CL-6B column chromatography, DEAE-Sepharose column chromatography, and crystallization successively, after the second ammonium sulfate precipitation. The highest specific activity of the crystalline preparation was 500 units/mg at 37 °C as assayed by the method of Segal and Matsuzawa (1970). A value of 6.9 for the absorbance at 278 nm (1%) was used for calculating the enzyme concentration.

Achromobacter protease I, which specifically cleaves lysyl bonds (Masaki et al., 1981), and arginylendopeptidase were generous gifts of Dr. T. Masaki of the Department of Agri-

† This work was supported in part by grants-in-aid from the Fujita Health University (to K. Titani and T.M.).

* To whom correspondence should be addressed at the Division of Biomedical Polymer Science, Institute for Comprehensive Medical Science, School of Medicine, Fujita Health University, Toyoake, Aichi, Japan 470-11.

‡ Fujita Health University.

§ Institute of Physical and Chemical Research (RIKEN).

[1] Abbreviations: API, *Achromobacter protease* I; BNPS-skatole, 2-[(2-nitrophenyl)sulfenyl]-3-methyl-3-bromoindolenine; GOT, aspartate aminotransferase; GPT, alanine aminotransferase; N^{α}, α-amino; PE, pyridylethyl; PLP, pyridoxal 5′-phosphate; PTH, phenylthiohydantoin; RP-HPLC, reversed-phase high-performance liquid chromatography; SDS–PAGE, polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate; TAT, tyrosine aminotransferase; TFA, trifluoroacetic acid; TPCK, L-(1-tosylamino)-2-phenylethyl chloromethyl ketone; V8, *Staphylococcus aureus* V8.

0006-2960/91/0430-6048$02.50/0

cultural Chemistry, Ibaraki University (Ibaraki, Japan), and of Takara Shuzo Co. (Otsu, Japan), respectively. TPCK-treated trypsin was obtained from Cooper Biomedical (Malvern, PA). *Staphylococcus aureus* V8 protease was a product of Miles (Naperville, IL). Cyanogen bromide and BNPS-skatole were from Wako Pure Chemical (Osaka, Japan) and Pierce Chemical Co. (Rockford, IL), respectively.

Reduction and Pyridylethylation of GPT. GPT was reduced with tri-*n*-butylphosphine (Wako Pure Chemical) (Ruegg & Rudinger, 1977) and pyridylethylated with 4-vinylpyridine (Tokyo Kasei Kogyo, Tokyo, Japan) (Hermodson et al., 1973) in the presence of 7 M guanidine hydrochloride, 0.1 M Tris-HCl (pH 8.5), and 1 mM EDTA in the dark at room temperature overnight. The reaction mixture was separated by gel permeation HPLC on tandem columns of TSK G3000 SW_{XL} (7.8 × 300 mm each) (Tosoh, Tokyo, Japan) in 6 M guanidine hydrochloride containing 10 mM sodium phosphate (pH 6.0), and then desalted by either gel filtration or dialysis and lyophilized.

Enzymatic Digestion and Chemical Cleavage. The PE-protein was digested overnight with API at a substrate to enzyme ratio of 200 (mol/mol) at 37 °C in 50 mM Tris-HCl, pH 9.0, containing 2.0 M urea. The enzyme was fully active in 2.0 M urea as reported by Masaki et al. (1981). Other enzymatic digestions were performed overnight at a substrate to enzyme ratio of 100 (mol/mol) at 37 °C in 0.1 M ammonium bicarbonate, pH 8.0. Peptides were separated by RP-HPLC using a Bakerbond WP-C4 column (4.6 × 50 mm; J. T. Baker, Phillipsburg, NJ). Further purification was carried out by RP-HPLC on a column of Superspher RP Select B (2.1 × 125 mm; Merck, Darmstadt, Germany) or Aquapore PH 300 (2.1 × 30 mm; Applied Biosystems, Foster City, CA) with gradients of acetonitrile into 0.1% aqueous TFA (Mahoney & Hermodson, 1980). The *S*-PE-protein was cleaved with cyanogen bromide in 70% formic acid as described by Gross (1967), and peptides were separated by RP-HPLC on an Aquapore PH 300 column. Cleavage with BNPS-skatole, dilute acid, concentrated HCl, or 70% formic acid followed the procedure of Omenn et al. (1970), Inglis (1983), Titani and Narita (1964), or Landon (1977).

Amino Acid and Sequence Analysis. Amino acid analysis was performed with a Hitachi Model L8500 amino acid analyzer or by the Waters Pico-tag system (Heinrikson & Meredith, 1984; Bidlingmeyer et al., 1984). Amino-terminal sequence analysis was performed with an Applied Biosystems Model 470A or 477A protein sequencer (Hewick et al., 1981) connected on-line to a Model 120A PTH analyzer. Mass spectral analysis was kindly performed by Dr. Carl J. March of Immunex, Seattle, WA, on a BioIon 20 Biopolymer mass analyzer (Applied Biosystems).

Homology Search, Hydropathy Index, and Secondary Structure Prediction. Homologous sequences were searched through protein sequence databases (NBRF-PIR and SWISS-PROT) using DNASIS (Hitachi Softwear Engineering). The hydropathy index was calculated by the method of Kyte and Doolittle (1982), and the secondary structure was predicted by the method of Chou and Fasman (1978). Multiple sequence alignment was performed by CLUSTAL (Higgins & Sharp, 1988).

Results

Sequence Analysis. Sequenator analysis of the intact protein (320 pmol) yielded only minor signals under 8.2 pmol due to impurities in five cycles of degradation, indicating that the amino terminus of the protein is blocked. The blocking group was identified to be an acetyl group by mass spectral analysis of a peptide derived from the amino terminus of the PE-protein as later described.

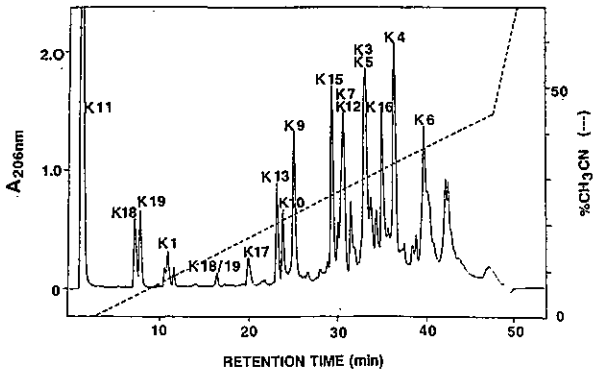

FIGURE 1: Primary separation of peptides generated by digestion with API. The digest (4 nmol) was separated on a Bakerbond WP-C4 column (4.6 × 50 mm) using a TFA/acetonitrile system at a flow rate of 0.5 mL/min. Peptides were monitored at 206 nm. Peptides are identified by the prefix K as in Figure 2.

The digest of the PE-protein (ca. 4 nmol) with API was separated into 16 fractions by RP-HPLC as shown in Figure 1. Three major fractions (K11, K7 and K12, and K3 and K5) were further purified on a Superspher RP Select B column or an Aquapore PH 300 column by RP-HPLC (data not shown). Sixteen major peptides designated as K1 through K19 except for K2, K8, and K14 (Table I), accounting for 99% of the entire protein, were thus isolated and subjected to sequence analysis. A free lysine (K8, residue 290) and two dipeptides, Gly-Lys (K2, residues 17–18) and Ala-Lys (K14, residues 389–390), were not recovered. They may have been overlooked in the breakthrough peak. Cyanogen bromide fragments of the PE-protein (ca. 4 nmol) were separated in a similar manner by RP-HPLC (data not shown). Thirteen peptides isolated, M1 through M17 except for M6, M12, M15, and M16, were analyzed to provide overlaps of 19 K peptides. The PE-protein was cleaved with 70% formic acid at a single Asp–Pro bond (residues 149–150). The mixture was separated by size-exclusion HPLC on tandem columns of TSK G3000 SW_{XL} and G4000 SW_{XL}. Peptide DP1 thus isolated was subjected to sequence analysis to fill gaps in the sequence obtained with peptide K5–E1. To obtain alignment of the six tryptic subpeptides of the peptide K6, a single tryptophanyl bond (between residues 213 and 214) in peptide K6 was cleaved with BNPS-skatole. After extraction of excess reagents with 1-chlorobutane, the mixture was separated by RP-HPLC. The isolated peptide K6–W1 was subjected to sequence analysis.

Analysis of 16 major K peptides (K1–K19) and their subpeptides provided most of the sequence as shown in Figure 2. Of the 16 peptides, only peptide K1 was N^{α}-blocked, as was the intact protein, indicating that this peptide was derived from the amino terminus of the protein. Mass spectral analysis of this peptide (ca. 200 pmol) showed that the peptide had an MH^+ of 1773.9, in accord with the sequence of Ac-Ala-Ser-Arg------.

Thus, analysis of 16 K and 13 M peptides and their subpeptides yielded the complete amino acid sequence of rat cytosolic GPT of 495 residues starting with Ac-Ala, as summarized in Figure 2.

Sequence Homology. Homology search through the two protein sequence databases (NBRF-PIR and SWISS-PROT) found only the 20-residue PLP binding site sequence of the pig GPT and rat TAT as related proteins. Functionally closely related GOT was not found among the 400 proteins which scored highest in the homology search. However, when the

Table I: Amino Acid Compositions of Intact PE-GPT and Its Peptides Derived by Digestion with API[a]

peptide residue no.	K1 1–16	K3 19–56	K4 57–104	K5 105–168	K6 169–263	K7 264–289
Asp/Asn (D/N)	2.7 (3)	1.7 (2)	5.6 (5)	3.3 (6)	7.0 (7)	2.4 (3)
Thr (T)		1.5 (2)	2.3 (2)	1.7 (2)	3.2 (3)	
Ser (S)	2.7 (3)		1.8 (2)	4.6 (6)	2.4 (2)	2.1 (2)
Glu/Gln (E/Q)	2.8 (2)	6.1 (7)	7.3 (5)	6.0 (6)	14.6 (12)	5.2 (5)
Pro (P)		1.7 (2)	4.3 (5)	2.8 (3)	5.6 (6)	
Gly (G)	2.2 (1)	1.6 (2)	2.5 (2)	5.5 (7)	5.5 (5)	1.7 (2)
Ala (A)	2.0 (2)	2.0 (2)	5.0 (5)	5.3 (7)	10.6 (12)	2.0 (2)
Val (V)	1.2 (1)	3.7 (6)	2.1 (3)	2.2 (2)	6.0 (8)	2.0 (2)
Met (M)		0.8 (1)	1.1 (1)	0.8 (1)		1.2 (1)
Ile (I)		0.6 (1)	2.0 (3)	5.4 (8)	4.5 (6)	
Leu (L)	1.5 (1)	3.4 (5)	3.3 (4)	4.0 (4)	9.1 (11)	2.0 (2)
Tyr (Y)		0.8 (1)	0.9 (1)	1.5 (2)	3.2 (4)	1.6 (2)
Phe (F)			3.4 (4)	1.0 (1)	2.5 (2)	2.3 (3)
His (H)				0.8 (1)		0.6 (1)
Lys (K)	1.4 (1)	0.9 (1)	1.3 (2)	1.1 (1)	1.2 (1)	1.5 (1)
Arg (R)	2.1 (2)	4.2 (5)	3.1 (3)	4.7 (6)	10.0 (10)	
PE-Cys (C)		0.5 (1)	0.3 (1)	+ (1)	1.7 (5)	
Trp[b] (W)	ND (0)	ND (0)	ND (0)	ND (0)	ND (1)	ND (0)
total residues	16	38	48	64	95	26
yield (%)	84	55	50	75	23	69

peptide residue no.	K9 291–313	K10 314–337	K11 338–341	K12 342–372	K13 373–388	K15 391–422
Asp/Asn (D/N)		2.4 (2)		1.0 (1)		2.6 (2)
Thr (T)	1.1 (1)			1.0 (1)		1.2 (1)
Ser (S)	3.8 (4)			3.3 (4)		1.1 (1)
Glu/Gln (E/Q)	5.8 (4)	6.0 (4)	1.2 (1)	2.5 (2)	7.1 (6)	7.8 (6)
Pro (P)	1.9 (2)			5.9 (7)		4.0 (4)
Gly (G)	1.6 (1)	4.5 (5)	0.9 (1)	1.0 (1)		1.9 (2)
Ala (A)	1.9 (1)	1.8 (1)		1.1 (1)	4.0 (4)	2.8 (2)
Val (V)	2.1 (2)	2.9 (4)		2.9 (4)	0.8 (1)	2.5 (3)
Met (M)	1.9 (2)	1.8 (2)	0.8 (1)	2.7 (3)		1.0 (1)
Ile (I)						0.8 (1)
Leu (L)	1.6 (2)			2.7 (3)	1.4 (2)	2.2 (3)
Tyr (Y)	1.2 (1)	1.9 (2)				0.9 (1)
Phe (F)	1.0 (1)	1.0 (1)		1.0 (1)	0.8 (1)	1.7 (2)
His (H)	1.8 (1)					
Lys (K)	0.9 (1)	0.9 (1)	1.0 (1)	1.1 (1)	0.8 (1)	0.7 (1)
Arg (R)		1.0 (1)		1.0 (1)	1.1 (1)	1.1 (1)
PE-Cys (C)		0.4 (1)		0.6 (1)		0.4 (1)
Trp[b] (W)	ND (0)	ND (0)	ND (0)	ND (0)	ND (0)	ND (0)
total residues	23	24	4	31	16	32
yield (%)	50	50	69	77	63	52

peptide residue no.	K16 423–475	K17 476–482	K18 483–489	K19 490–495	intact PE-GPT 1–495
Asp/Asn (D/N)	1.9 (1)				29.3 (32)
Thr (T)	2.6 (3)			1.2 (1)	17.3 (16)
Ser (S)	1.5 (1)		1.3 (1)	1.2 (1)	24.5 (27)
Glu/Gln (E/Q)	10.0 (9)	+ (1)		1.5 (1)	74.9 (71)
Pro (P)	3.3 (4)				34.3 (33)
Gly (G)	4.6 (6)				36.0 (36)
Ala (A)	4.3 (3)		1.2 (1)		45.1 (44)
Val (V)	1.9 (3)				36.4 (39)
Met (M)	0.2 (3)				14.9 (16)
Ile (I)	1.3 (2)				17.8 (21)
Leu (L)	3.4 (6)	4.0 (4)	1.0 (1)		49.1 (48)
Tyr (Y)	0.8 (1)			1.1 (1)	16.2 (16)
Phe (F)	2.5 (4)		1.1 (1)	1.0 (1)	22.0 (22)
His (H)	0.6 (1)		2.1 (2)	1.0 (1)	5.3 (7)
Lys (K)	0.6 (1)	1.2 (1)	0.9 (1)		11.7 (19)
Arg (R)	2.0 (2)	1.7 (1)			33.6 (33)
PE-Cys (C)	0.8 (3)				14.5 (14)
Trp[b] (W)	ND (0)	ND (0)	ND (0)	ND (0)	ND (1)
total residues	53	7	7	6	495
yield (%)	61	86	96	70	

[a] Results are expressed as residues per peptide or protein by amino acid analysis or, in parentheses, from the sequence (Figure 2). Hydrolysis was carried out with 6 N HCl at 110 °C for 20 h. Each composition was calculated on basis of the integral value of the residue underlined. [b] Trp was not determined (ND).

search was repeated against known sequences of aminotransferases, it was noticed that GOT sequences show sequence similarity, although with marginal significance, in wider regions. Common residues were scattered throughout almost the entire sequence but with very limited clustering. To evaluate whether the weak homology between GPT and GOT

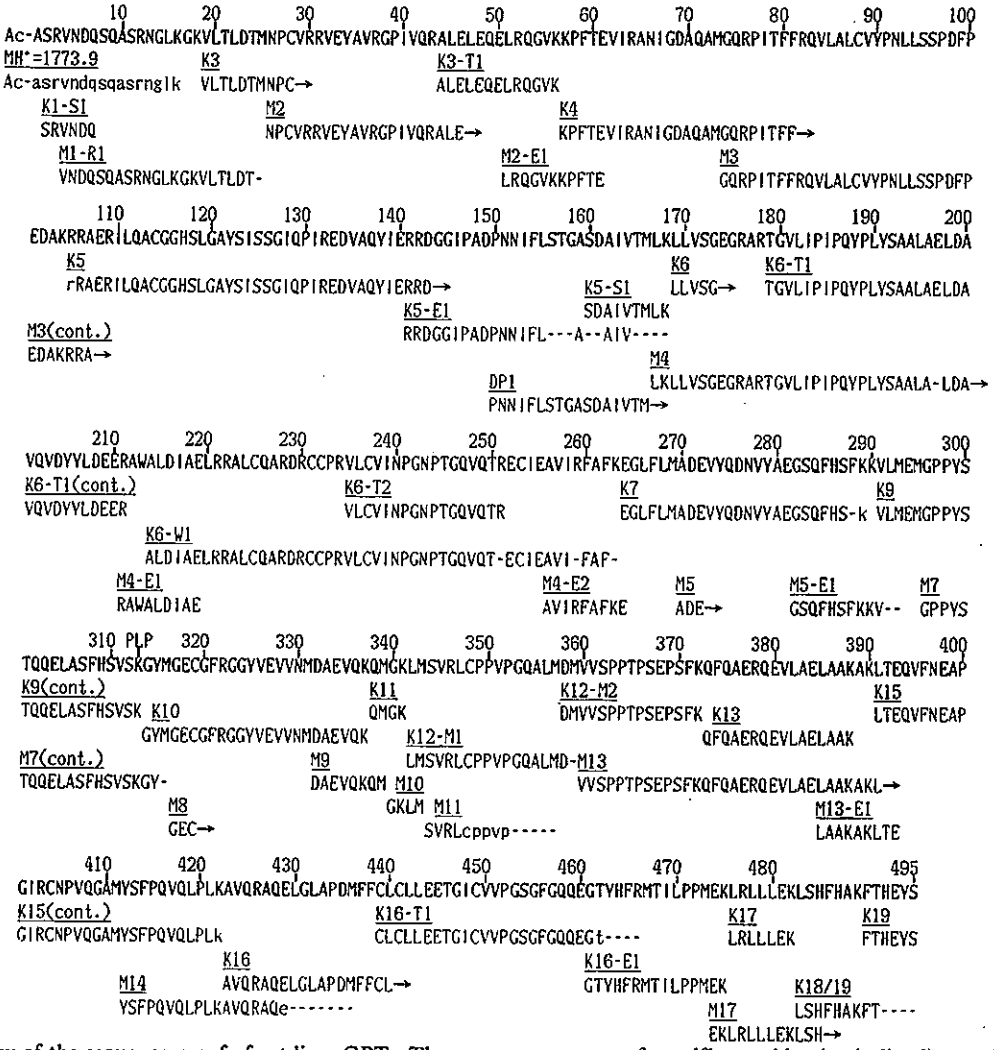

FIGURE 2: Summary of the sequence proof of rat liver GPT. The proven sequences of specific peptides (underlined) are given in one-letter code below the summary sequence (boldface type). Prefixes K and M denote peptides generated by cleavage of the *S*-PE-protein at lysyl bonds (with API) and methionyl bonds (with cyanogen bromide), respectively; DP indicates a peptide generated by cleavage of *S*-PE-GPT with 70% formic acid. The products of lysyl or methionyl cleavage are numbered from the amino terminus toward the carboxyl terminus of the protein. Subpeptides are identified by hyphenated suffixes, where the following code indicates the subdigesting agent: E, V8 protease; K, API; T, trypsin; R, arginylendopeptidase; S, chemical cleavage with 12 N HCl; M, chemical cleavage with cyanogen bromide; W, chemical cleavage with BNPS-skatole. Peptide sequences written in upper case letters are proven by Edman degradation unambiguously; those in lower case letters indicate tentative identifications. Unidentified residues are shown by dashes or horizontal arrows; the latter indicate long unidentified sequences. Ac denotes an acetyl group. The result of the mass spectral analysis of the N-terminal peptide is indicated above the N-terminal sequence in lower case letters. The putative PLP binding site is indicated at residue 313.

really reflects their evolutionary relationship, amino acid sequences of rat cytosolic GPT and aminotransferases of various origins were aligned by using a multiple sequence alignment program, CLUSTAL (Higgins & Sharp, 1988), with standard parameters. Close examination of the sequence alignment obtained revealed that most of the residues involved in binding of PLP and substrate in GOT are conserved in GPT, indicating that these two proteins are products of extensive divergent evolution. From the alignment, sequence identities between GPT and TAT, cytosolic GOT, and mitochondrial GOT were calculated to be 24.7%, 17.0%, and 16.0%, respectively (Figure 3).

Hydropathy Profile and Secondary Structure Prediction. The hydropathy profile of rat cytosolic GPT (Figure 4) indicated a rather even distribution of hydrophobic and hydrophilic regions throughout the molecule. When compared to that of rat cytosolic GOT, the general patterns were similar, especially in the middle of the molecule (the PLP binding large domain, residues ca. 88 through 385 by the numbering system in Figure 3). Prediction of secondary structure by the method of Chou and Fasman (1978) indicated the presence of α-helices (42.2%) and β-sheets (25.9%) in rat cytosolic GPT. Again, some resemblance in the predicted structures of rat cytosolic GPT and GOT was found.

Discussion

In general, proof of the sequence of rat GPT involved two primary sets of overlapping peptides derived by cleavage at either lysyl or methionyl bonds. A special problem was presented by an N^{α}-blocked amino terminus. Attempts to remove the acylamino acid from the N^{α}-blocked peptide K1 with acylamino acid releasing enzyme (Tsunasawa & Narita, 1976) failed. The N^{α}-blocked amino acid was finally identified to be an acetylated alanine residue by mass spectral analysis. Although most of the residues in the sequence were identified in the two complementary peptides, there are regions of relatively weak analytical proof at the single-residue overlaps aligning residues 210–212, 263–265, 290–292, and 357–359. The first of these occurred in peptides K6 and M4. From the substrate specificity of V8 protease, the sequence of M4–E1 was most likely preceded by Glu or Asp. The single Asp-Arg in M4 was firmly placed at residues 229–230 in the analysis of K6–W1, and the only possible place to put M4–E1 was at resudes 211–219 to overlap K6–T1 to K6–W1. Among the

```
                10        20        30        40        50        60
cGPT  ASRVNDQSQASRNGLKGKVLTLDTMNPCVRRVEYAVRGPIVQRALELEQELRQGVKKPFT
cTAT  MDSYVIQTDVD-DSL-SSVLDVH-VNIGGRN---SVQGRKKGRKARWDVRPSDMSNKTF-
cGOT  ------------------------------------APPSFFAQVPQAPPVLVFK-LI
mGOT  ------------------------------------SSWWTHVEMGGPDPILG-VT

                70        80        90       100       110       120
cGPT  EVIRANIGDAQAMCQRPITPFRQVLA-LCVYPNLLSSPDFPEDAKRRAERILQACGGHSL
cTAT  NPIRAIVDNMKVQPNPNKTVISLSIGDPTVFGNLPTDPEVTQAMK-------DALDSGKY
cGOT  ADFRDDPDPRKV--NLGVGAYRTDDSQPWVLPVVRKVEQKIANDHSLNHEYLPILGLAEF
mGOT  EAFKRDTNSKKM--NLGVGAYRDDNGKPYVLPSVRKAEAQIAGKN-LDKEYLPIGGLADF

               130       140       150       160       170       180
cGPT  GAYSISSGIQPIREDVAQYIERRDGGIPADPNNIFLSTGASDAIVTMLKLLVSGEGRART
cTAT  NGYAPSIGYLSSREEVASYYHCHEA--PLEAKDVILTSGCSQAIELCLAVLANPGQN---
cGOT  RSCASQLVLGDNSP---ALRENRVGGV------QSLGGTGALRIGADFLGRWYNGTDNKNT
mGOT  CKASAELALGENSE---VLKSGRFVTV-----QTISGTGALRVGASFLQRFFKFSRD---

               190       200       210       220       230       240
cGPT  GVLIPIPQYPLYSAALAELDAVQVDYY-LDEERAWALDIAELRRALCQARDRCCPRVLCV
cTAT  -ILIPRPGFSLYRT-LAESMGIEVKLYNLLPEKSWEIDLKQLE-SLIDEKTAC----LVV
cGOT  PVYVSSPTWENHNGVFSAAGFKDIRSYRYWDAEKRGLDLQGFLNDLENAPEF--SIFVLH
mGOT  -VFLPKPSWGNHTEIARDAGM-QLEGYRYYDPKTCGFDFSGALEDISKIPEQ--SVLLLH

               250       260       270       280       290       300
cGPT  INPGNPTGQVQTRECIEAVIRFAFKEGLFLMADEVYQDNVYAEGSQFHSFKKVLMEMGPP
cTAT  NNPSNPCGSVFSKRHLQKILAVAERQCVPILADEIYGDMVFSD-CKYEPLANL-------
cGOT  ACAHNPTGTDPTEEEWKQIAAVMKRRFLFPFFDSAYQGFASGDLEKDAWAIRYFVSEG--
mGOT  ACAHNPTGVDPRPEQWKEMAAYVKKNNLFAFFDMAYQGFASGDGNKDAWAVRHFIEQG--

               310       320       330       340       350       360
cGPT  YSTQQELASFHSVSKGYMGECGFRGGYVEVVNMDAE----VQKQMGKLMSVRL---CPPV
cTAT  -STNVPILSCGGLAKRWLVP-GWRLGWILIHDRRDIFGNEIRDGLVKL-SQRILGPCTIV
cGOT  ----FELFCAQSFSKN-FGLYNERVGNLTVVGKEHDSVLRVLSQMEKI--VRITWSNPPA
mGOT  ----INVCLCQSYAKN-MGLYGERVGAFTVVCKDAEEAKRVESQL-KIL-IRPLYSNPPL

               370       380       390       400       410       420
cGPT  PGQALMDMVVSPPTPSEPSFKQFQAERQEVLAELAAKAKLTEQVFNEAPGIRCNPVQGAM
cTAT  QG-ALKSILQRTP-------QEFY---HDTLSFLKSNADLCYGALAAIPGLQPVRPSGAM
cGOT  QGARIVATTLSNPELFKEWKGNVKTMADRILT-MRSELRARLEALKTPGTWSHITEQIGM
mGOT  NGARIAATILTSPDLRQGWLQEVKGMADRIGS-MRTQLVSNLKKEGSSHNWQHITDQIGM

               430       440       450       460       470       480
cGPT  YSFPQVQLPLKAVQRAQELGLAPDMFFCLLLEETGICVVPGSGFGQQEGTYHFRMTILP
cTAT  YL-------MVGIEMEHFPEFENDVEFTERLIAEQAVHCLPATCFEYPN---FFRVVITV
cGOT  FSFT-------GLNPKQVEYLVNEKHIYLMPSGRINMCGLTTKNLDYVA----------
mGOT  FCFT-------GIKPEQVERLTKEFSIYMTKDGRISVAGVTSGNVGYLA----------

               490       500       510
cGPT  PMEKLRLLLEKLSHFHAKFTHEYS----------
cTAT  PEV---MMLEACSRIQEFCEQHYHCAEGSQEECDK
cGOT  -----TSINEAVTKFQ-------------------
mGOT  -----HAIHQ-VTK--------------------
```

FIGURE 3: Sequence alignment of rat aminotransferases. Asterisks indicate residues conserved among the sequences. The substrate binding residues in GOT (Arg-292 and Arg-386 in cytosolic GOT) are marked with open circles. Filled circles indicate the residues present in the PLP binding pocket. cGPT, cytosolic GPT; cTAT, cytosolic TAT (Grange et al., 1985); cGOT, cytosolic GOT (Hayashi et al., 1987); mGOT, mitochondrial GOT (Huyuh et al., 1980). In Huyuh et al. (1980), glycine was assigned at position 189 but corrected to tryptophan in Kondo et al. (1987).

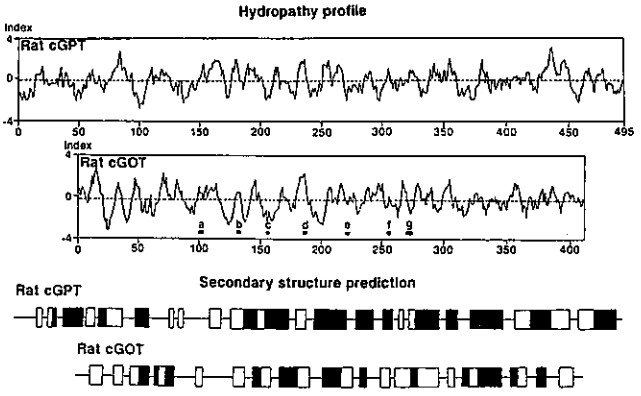

FIGURE 4: Comparison of the hydropathy profile and the predicted secondary structure of rat cytosolic GPT with those of rat cytosolic GOT. The hydropathy profiles were calculated by the method of Kyte and Doolittle (1982) with a window size of 7. The midpoint dashed line corresponds to the grand average of the hydropathy of the protein. Horizontal bars below the hydropathy profile of rat cytosolic GOT indicate the putative locations of β-sheets participating in the seven-stranded pleated-sheet structure which correspond to pig cytosolic GOT of known tertiary structure (Jansonius et al., 1985). Seven β-sheets are labeled a–g. The secondary structures were predicted according to the method of Chou and Fasman (1978). Closed and open boxes indicate α-helix and β-sheet, respectively. cGPT, cytosolic GPT; cGOT, cytosolic GOT.

API peptides, only K7 had the amino-terminal Glu. Therefore, the sequence of M4–E2 clearly indicated the overlap between K6 and K7. The weakest overlap at residues 290–292 was supported by the amino acid composition of M5 (D/N 2.9, E/Q+Hse 5.0, S 2.0, G 1.2, A 2.0, V 2.9, L 1.0, F 1.9, K 1.8, H 1.2, and Y 1.8 mol/mol of peptide). The remaining one occurred at the middle of K12, and the compositional data of K12 (Table I) were in good agreement with the conclusion (Figure 2). From the present study, the rat GPT subunit is composed of 495 residues. The molecular weight of this subunit, calculated from the amino acid sequence, is 55 018, which is in good agreement with a molecular weight of about 55 000 estimated by SDS–PAGE.

Three important results emerged from the determination of the complete amino acid sequence of rat liver cytosolic GPT.

(1) The active enzyme with a molecular weight of 114 000 (Gatehouse et al., 1967) should be a homodimer containing 2 mol of PLP.

(2) The PLP binding site of rat liver GPT is most likely Lys-313. The sequence of residues Gln-303 to Arg-322 of rat liver GPT presented herein coincided with the 20-residue sequence around the PLP binding lysyl residue of pig heart GPT reported by Tanase et al. (1979) except for only a single residue replacement of Tyr-315 to Phe. This also suggests that the sequence around the PLP binding site is highly conserved in GPT beyond tissue and species differences.

(3) The amino acid sequence of rat liver cytosolic GPT presented herein is quite unique among those of pyridoxal enzymes with known structure registered in updated protein databases. The detailed sequence homology search, however, indicated that rat liver GPT is significantly homologous to those of rat liver cytosolic GOT and mitochondrial GOT with 17.0% and 16.0% identity, respectively, and moderately homologous to that of rat liver TAT with 24.7% identity as shown in Figure 3. Mitochondrial GOT was reported to have catalytic activity also toward alanine and aromatic amino acids at 10^5 and 10^3 times slower rates, respectively, than dicarboxylic substrates (Mavrides & Christen, 1978). In this regard, it is interesting that the sequence of GPT is more homologous to that of TAT than that of mitochondrial GOT (Figure 3). The sequence homology between GPT and other α-aminotransferases is much higher in the middle portion comprising the active-site large domain than the amino- and carboxyl-terminal portions comprising the small domain involved in the substrate specificity. In spite of rather low homology to GOT, most of the crucial amino acid residues providing hydrogen bonds to PLP at the PLP binding pocket in the large domain of both cytosolic and mitochondrial GOT with known tertiary structure are also retained in rat liver cytosolic GPT as shown in Figure 5. Furthermore, the tertiary structures of cytosolic and mitochondrial GOTs resemble each other despite rather extensive divergence in their sequences. These suggest that the topology of secondary structures surrounding the PLP binding site in the large domain of GOT, containing a large pleated-sheet structure composed of seven β-strands and sandwiched in eight helices, revealed by X-ray crystallography (Jansonius et al., 1985), may also be conserved in rat liver GPT. Some resemblance in the tertiary structure between GPT and GOT, particularly in the large domain binding to PLP, is also suggested by comparison of the hydropathy index and the secondary structures predicted for GPT with those of GOT as shown in Figure 4.

Two arginyl residues (residues 292 and 386) are involved in substrate binding in pig cytosolic GOT (Figure 5). As seen in the alignment (Figure 3), corresponding residues to Arg-292

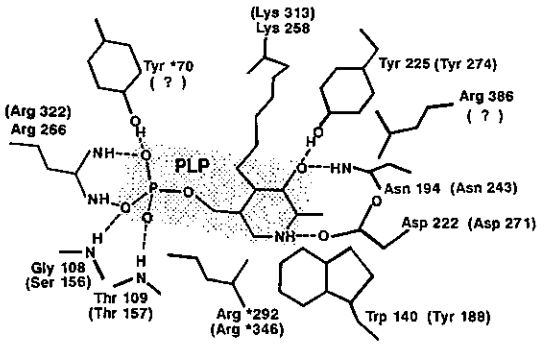

FIGURE 5: Active center of pig cytosolic GOT (rat cytosolic GPT). The shadowed area indicates pyridoxal 5′-phosphate (PLP). All side chains interacting with either the coenzyme PLP or the substrate are included. Residues belonging to the adjacent subunit are labeled with asterisks. The putative residues in the active center of rat cytosolic GPT are indicated in parentheses.

in GOT, which binds the β-carboxyl group of the substrate, are conserved in GPT and TAT, but those of Arg-386 appear to be substituted. The small domain regions containing Arg-386 are so poorly conserved that any change in parameters (i.e., gap penalties) results in different alignment. It is possible that the Arg-386 in GOT, which binds the α-carboxyl group of the substrate, is conserved as Arg-466 in GPT or Arg-417 in TAT as indicated for the latter in the alignment of Mehta et al. (1989). More sequence information is necessary for the real evolutional alignment for regions of the small domain. It is interesting to see how the tertiary structure of GOT is conserved in GPT and TAT.

It seems clear that GPT is a member of the family of α-aminotransferases but it is more closely related to TAT than GOT and others in sequence homology and length of polypeptides. This suggests that GPT and TAT have diverged from a GOT family at an earlier stage of evolution.

ACKNOWLEDGMENTS

We are indebted to Dr. Harold L. Segal of the State University of New York at Buffalo for his interest and encouragement throughout this study and critical reading and editing of the manuscript and to Dr. Carl J. March of Immunex, Seattle, WA, who kindly performed mass spectral analysis of peptides. We are also grateful to Ms. Atsuko Nagata for typing the manuscript.

REFERENCES

Bidlingmeyer, B. A., Cohen, S. A., & Tarvin, T. L. (1984) *J. Chromatogr. 336*, 93–104.

Bulos, B., & Handler, P. (1965) *J. Biol. Chem. 240*, 3283–3294.

Chou, P. Y., & Fasman, G. O. (1978) *Annu. Rev. Biochem. 47*, 251–276.

DeRosa, G., & Swick, R. W. (1975) *J. Biol. Chem. 250*, 7961–7967.

Gatehouse, P. W., Hopper, S., Schatz, L., & Segal, H. L. (1967) *J. Biol. Chem. 242*, 2319–2324.

Golichowski, A., Harruff, R. C., & Jenkins, W. T. (1977) *Arch. Biochem. Biophys. 178*, 459–467.

Grange, T., Guenet, C., Dietrich, J. B., Chasserot, S., Fromont, M., Befort, N., Jami, J., Beck, G., & Pictet, R. (1985) *J. Mol. Biol. 184*, 347–350.

Gross, E. (1967) *Methods Enzymol. 11*, 238–255.

Hayashi, H., Horio, Y., Tanaka, T., Taketoshi, M., & Wada, H. (1987) in *Biochemistry of Vitamin B_6* (Korpela, T., & Christen, P., Eds.) pp 39–42, Birkhäuser, Basel.

Heinrikson, R. L., & Meredith, S. C. (1984) *Anal. Biochem. 136*, 65–74.

Hermodson, M. A., Ericsson, L. H., Neurath, H., & Walsh, K. A. (1973) *Biochemistry 12*, 3146–3153.

Hewick, R. M., Hunkapiller, M. W., Hood, L. E., & Dreyer, W. J. (1981) *J. Biol. Chem. 256*, 7990–7997.

Higgins, D. G., & Sharp, P. M. (1988) *Gene 73*, 237–244.

Hopper, S., & Segal, H. L. (1962) *J. Biol. Chem. 237*, 3189–3195.

Hopper, S., & Segal, H. L. (1964) *Arch. Biochem. Biophys. 105*, 501–505.

Huyuh, Q. K., Sakakibara, R., Watanabe, T., & Wada, H. (1980) *Biochem. Biophys. Res. Commun. 97*, 474–479.

Inglis, A. S. (1983) *Methods Enzymol. 91*, 324–332.

Jansonius, J. N., Eichele, G., Ford, G. C., Picot, D., Thaller, C., & Vincent, M. G. (1985) in *Transaminase* (Christen, P., & Metzler, D. E., Eds.) pp 110–138, Wiley, New York.

Kondo, K., Wakabayashi, S., & Kagamiyama, H. (1987) *J. Biol. Chem. 262*, 8648–8659.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Landon, M. (1977) *Methods Enzymol. 47*, 145–149.

Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199–11203.

Masaki, T., Tanabe, M., Nakamura, K., & Soejima, M. (1981) *Biochim. Biophys. Acta 660*, 44–50.

Matsuzawa, T., & Segal, H. L. (1968) *J. Biol. Chem. 243*, 5929–5934.

Mavrides, C., & Christen, P. (1978) *Biochem. Biophys. Res. Commun. 85*, 769–773.

Mehta, P. K., Hale, T. I., & Christen, P. (1989) *Eur. J. Biochem. 186*, 249–253.

Omenn, G. S., Fontana, A., & Anfinsen, C. B. (1970) *J. Biol. Chem. 245*, 1895–1902.

Ruegg, U., & Rudinger, J. (1977) *Methods Enzymol. 47*, 111–116.

Saier, M. H., & Jenkins, W. T. (1967a) *J. Biol. Chem. 242*, 91–100.

Saier, M. H., & Jenkins, W. T. (1967b) *J. Biol. Chem. 242*, 101–108.

Segal, H. L., & Kim, Y. S. (1963) *Proc. Natl. Acad. Sci. U.S.A. 50*, 912–918.

Segal, H. L., & Matsuzawa, T. (1970) *Methods Enzymol. 17A*, 153–159.

Tanase, S., Kojima, H., & Morino, Y. (1979) *Biochemistry 18*, 3002–3007.

Titani, K., & Narita, K. (1964) *J. Biochem. (Tokyo) 56*, 241–265.

Tsunasawa, S., & Narita, K. (1976) *Methods Enzymol. 45*, 552–561.