Complete Amino Acid Sequence of Human Liver Cytosolic Alanine Aminotransferase (GPT) Determined by a Combination of Conventional and Mass Spectral Methods†

Mariko Ishiguro,‡,§ Koji Takio,§ Masami Suzuki,‡ Rieko Oyama,‡ 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 Aging Process, Frontier Research Program, Institute of Physical and Chemical Research (RIKEN), Wako, Saitama, Japan 351-01

Received April 29, 1991; Revised Manuscript Received July 24, 1991

ABSTRACT: The complete amino acid sequence of human liver cytosolic alanine aminotransferase (GPT) (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. Isolated peptides were analyzed with a protein sequencer or with a plasma desorption time of flight mass spectrometer and placed in the sequence on the basis of their molecular mass and homology to the sequence of rat GPT. The protein was found to be acetylated at the amino terminus and contained 495 amino acid residues. The M_r of the subunit was calculated to be 54 479, which was in good agreement with a M_r of 55 000 estimated by SDS–PAGE, and also indicated that the active enzyme with a M_r of 114 000 was a homodimer composed of two identical subunits. The amino acid sequence is highly homologous to that of rat GPT (87.9% identity) recently determined [Ishiguro, M., Suzuki, M., Takio, K., Matsuzawa, T., & Titani, K. (1991) *Biochemistry 30*, 6048–6053]. All of the crucial amino acid residues are conserved in human GPT, which seem to be hydrogen bonding to pyridoxal 5′-phosphate in rat GPT by the sequence homology to other α-aminotransferases with known tertiary structures.

Alanine aminotransferase (GPT)[1] (EC 2.6.1.2) is a pyridoxal enzyme which catalizes reversible transamination between alanine and 2-oxoglutarate to form pyruvate and glutamate. In human tissues including liver, kidney, heart, and skeletal muscles, this enzyme localizes in both cytosol and mitochondria as isoenzymes (Katunuma et al., 1962; Noguchi et al., 1977; Kamoda et al., 1980). The cytosolic enzyme is inducible by glucocorticoid, but the mitochondrial enzyme is not (Hopper & Segal, 1964; Noguchi et al., 1977). These isoenzymes participate in cellular nitrogen metabolism and also in liver gluconeogenesis starting with precursors transported from skeletal muscles (DeRosa & Swick, 1975). The cytosolic GPT activity leaked into blood is an important clinical marker of human acute hepatitis, but characteristics of the enzyme itself have not yet been studied in detail. Recently, we have reported the complete amino acid sequence of rat liver cytosolic GPT (Ishiguro et al., 1991). More recently, we have purified human liver cytosolic GPT to a homogeneous state as judged by SDS–PAGE and noted that the human enzyme has some physicochemical properties different from those of the rat enzyme. We herein report the complete amino acid sequence of human liver cytosolic GPT determined by a combination of conventional Edman degradation and newly developed plasma desorption mass spectral analysis and describe its sequence homology to the rat enzyme and to other α-aminotransferases.

Materials and Methods

Materials. Human liver cytosolic GPT was purified to homogeneity as judged by SDS–PAGE by a modification of the previous methods (Gatehouse et al., 1967; Matsuzawa & Segal, 1968): chromatography on tandem DE-52 and CM-52 columns instead of DE-52 column chromatography because the human enzyme is a more cationic protein than the rat enzyme, followed by calcium phosphate gel treatment, ammonium sulfate precipitation, octyl-Sepharose, Sepharose CL-6B, and DEAE-Sepharose column chromatography. About 1 mg of the enzyme was purified from 105 g of fresh remnant liver supplied from autopsy material of a 53-year-old Japanese woman. API, which specifically cleaves lysyl bonds even in the presence of urea (Masaki et al., 1981), was a generous gift of Dr. T. Masaki of the Department of Agricultural Chemistry, Ibaraki University, Ibaraki, Japan. TPCK-treated trypsin was obtained from Cooper Biomedical (Malvern, PA). V8 protease was a product of Miles (Naperville, IL). Cyanogen bromide and succinic anhydride were from Wako Pure Chemical (Osaka, Japan). BNPS-skatole was obtained from Pierce Chemical (Rockford, IL). *N*-Acetyl-L-serine was a generous gift of Dr. S. Tsunazawa of the Institute for Protein Research, Osaka University, Suita, Japan, and *N*-acetyl-L-alanine was obtained from Nacalai Tesque (Kyoto, Japan).

Reduction and Pyridylethylation of GPT. GPT was reduced with tri-*n*-butyl phosphine (Wako Pure Chemical) (Ruegg & Rudinger, 1977) and pyridylethylated with 4-vinylpyridine (Tokyo Kasei Kogyo, Tokyo, Japan) (Hermodson et al., 1973) in 0.1 M Tris-HCl (pH 8.5)/1 mM EDTA

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

* To whom correspondence should be addressed at 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: amu, atomic mass unit; API, *Achromobacter* protease I; BNPS-skatole, 2-[(2′-nitrophenyl)thio]-3-methyl-3-bromo-3*H*-indole; GOT, aspartate aminotransferase; GPT, alanine aminotransferase; MH^+, molecular ion; m/z, mass to charge ratio; 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; TFA, trifluoroacetic acid; TPCK, L-(1-tosylamino)-2-phenylethyl chloromethyl ketone; V8, *Staphylococcus aureus* V8.

0006-2960/91/0430-10451$02.50/0

containing 7 M guanidine hydrochloride 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 dialysis and lyophilized.

Enzymatic Digestion and Chemical Cleavage. The PE-protein was digested overnight with API at a substrate to enzyme ratio of 100 (w/w) at 37 °C in 50 mM Tris-HCl, pH 9.0, containing 2.0 M urea. Peptides were separated by RP-HPLC using a Bakerbond WP-C4 column (4.6 × 50 mm; J. T. Baker, Phillipsburg, NJ). Further purification of peptides was carried out by RP-HPLC on an Aquapore PH 300 column (2.1 × 30 mm; Applied Biosystems, Foster City, CA) with gradients of acetonitrile into 0.1% aqueous TFA (Mahoney & Hermodson, 1980). The PE-protein was cleaved at room temperature with cyanogen bromide in 70% formic acid as described by Gross (1967). Fragments were separated by HPLC on tandem columns of TSK G2000 SW_{XL} (7.8 × 300 mm each) (Tosoh). Further purification of fragments was achieved by RP-HPLC on an Aquapore PH 300 column. The largest cyanogen bromide fragment (M3) was succinylated in the presence of 8 M urea and 0.5 M sodium bicarbonate (pH 8.5) by a modification of Klotz (1967). After succinylation of the fragment, the reaction mixture was diluted with 4 volumes of deionized water and then digested with TPCK-trypsin at 37 °C overnight. 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. 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. Mass spectral analysis was performed with a BioIon 20 biopolymer mass analyzer (Applied Biosystems) on nitrocellulose-coated targets. Data were collected with 50–100 pmol of peptides for 1–30 million fission events at 15-kV acceleration voltage.

Results

Sequence Analysis. The digest of the PE-protein (ca. 4.1 nmol) with API was separated into 13 fractions by RP-HPLC as shown in Figure 1. One of the major fractions (K4) was further purified by RP-HPLC on an Aquapore PH 300 column (data not shown). Thirteen major peptides designated as K1–K15 except for K6 and K11 (Table I), accounting for more than 99% of the entire protein, were thus isolated and subjected to sequence analysis. A free lysine (K6, residue 290) and one dipeptide, Ala-Lys (K11, residues 389–390), were not recovered. They may have been overlooked in the breakthrough peak. From the cyanogen bromide digest of the PE-protein (ca. 1.3 nmol), nine peptides were isolated, M1–M11/12 (resulted from an uncleaved methionylthreonine bond between residues 467 and 468), and analyzed to provide overlaps of 15 K peptides. The PE-protein was cleaved with 70% formic acid at two Asp-Pro bonds (residues 149–150 and 368–369). The mixture was separated by size-exclusion HPLC on tandem columns of TSK G3000 SW_{XL}. Peptides DP1 and DP2 thus isolated were subjected to sequence analysis to provide overlaps

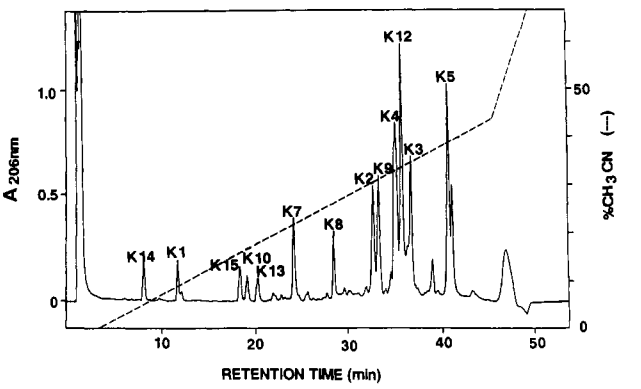

FIGURE 1: Primary separation of peptides generated by digestion of the PE-protein with API. The digest (4.1 nmol) was separated on a Bakerbond WP-C4 column (4.6 × 50 mm) using a TFA/acetonitrile system at a flow rate of 0.5 mL/min. Peptides were monitored at 206 nm. Peptides are identified by the prefix K as in Figure 2.

of K4 and K5 and K9–K12, respectively. The intact PE-protein was cleaved with BNPS-skatole at a single tryptophanyl bond between residues 213 and 214. After extraction of excess reagents with 1-chlorobutane, the mixture was subjected to sequence analysis.

Sequenator analysis of the intact protein (50 pmol) yielded only minor signals under 6.5 pmol due to impurities in five cycles of Edman degradation, indicating that the amino terminus of the protein is blocked. The analysis of 13 major K peptides (K1–K15) and their subpeptides provided most of the sequence as shown in Figure 2. Of the 13 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. Peptide K1 was cleaved with 2% formic acid at Asp (residue 6) and separated on a Superspher RP Select B column (2 × 119 mm; Merck, Darmstadt, Germany). The amino-terminal peptide K1–D1, Ac(ASSTG), thus isolated was subjected to further cleavage with 12 N HCl at Ser and Thr. The mixture was separated on the same column, and a fraction with the retention time for Ac-Ala (retention times of Ac-Ala and Ac-Ser were determined with synthetic Ac-Ala and Ac-Ser) was collected. After acid hydrolysis of this fraction, only Ala was detected by amino acid analysis, indicating that the amino terminus is Ac-Ala. The blocking group was further confirmed to be an acetyl group by mass spectral analysis. Analysis of peptides M1 and K1 derived from the amino terminus of the PE-protein yielded MH^+ values of 2623.0 and 1941.4, respectively (Table II), in accord with the sequences of Ac-Ala-Ser-Ser...Hse (residues 1–25) and Ac-Ala-Ser-Ser...Lys (residues 1–18). Thus, analyses of 13 K and 9 M peptides and their subpeptides yielded the complete amino acid sequence of human cytosolic GPT of 495 residues starting with Ac-Ala, as summarized in Figure 2.

Mass Spectral Analysis. Unambiguous data were easily obtained for peptides of MH^+ values less than 4000. For peptides of MH^+ values above 5000, the signals tended to broaden and it became increasingly difficult to obtain the precise m/z values. In such cases, doubly or triply charged signals were also evident. Most of the major signals were obtained as the protonated free form accompanied by the sodium forms except for fragments M2 and M9 (Table II, Figure 3), which seemed to give the protonated monosodium form as the major signal. With cyanogen bromide fragments, major signals were obtained as the lactone forms. Mass numbers for fragments M8/9 and M9 were obtained from a mixture (65 pmol total), and the intensity of the signals was very low. The rather big discrepancy between the calculated

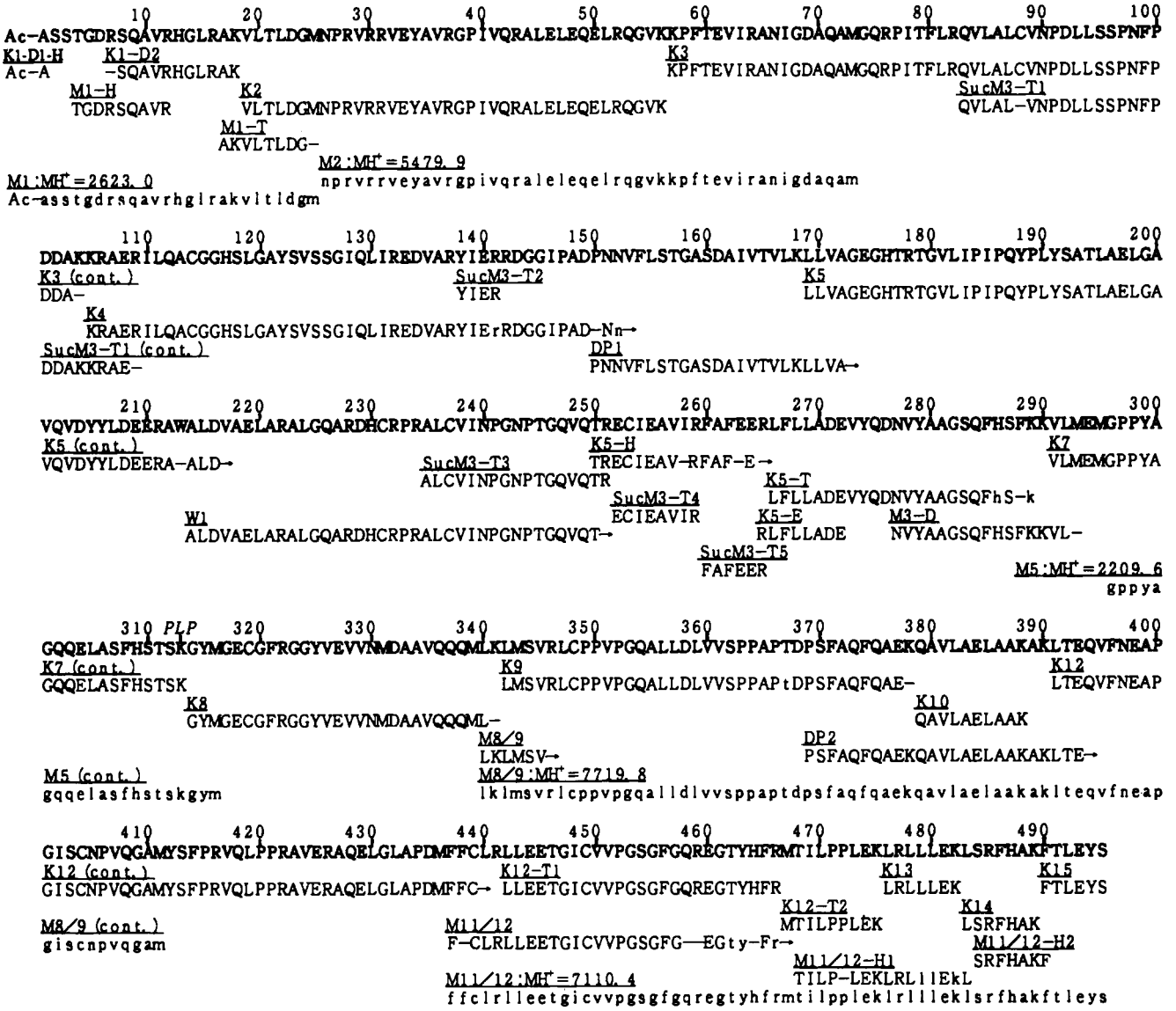

FIGURE 2: Summary of the sequence proof of human liver cytosolic GPT. The proven sequences of specific peptides (underlined) are given in the one-letter code below the summary sequence (bold type). The prefixes K, M, and W denote peptides generated by cleavage of the PE-protein at lysyl (with API), methionyl (with cyanogen bromide), and tryptophanyl bonds (with BNPS-skatole), respectively. The products of lysyl or methionyl cleavage are numbered from the amino terminus toward the carboxyl terminus of the protein. DP indicates peptides generated by cleavage of the PE-protein with 70% formic acid. Suc indicates succinylation of fragment M3 with succinic anhydride. Subpeptides are identified by hyphenated suffixes, where the following codes indicate the subdigesting agents: E, V8 protease; T, trypsin; H, chemical cleavage with 12 N HCl; D, chemical cleavage with 2% formic acid. 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 results of the mass spectral analysis are indicated above the sequences of specific peptides in lower case letters. The putative PLP binding site is indicated at residue 313.

and observed masses for these fragments might have resulted from the low amount. A higher amount of the sample may help ionization and result in signals of better quality. In many cases, differences between the calculated and observed masses were within the channel resolutions which were 0.66, 1.5, and 2.1 amu at m/z 1000, 5000, and 10000, respectively.

Isoelectric Point Prediction. Isoelectric points of human and rat cytosolic GPT, predicted by the methods of Skoog et al. (1986) from the amino acid sequences, were 6.51 and 5.89, respectively.

Discussion

Human liver cytosolic GPT showed a more cationic nature on DE-52 column chromatography at pH 5.7 than rat GPT; i.e., it passed through a DE-52 column but adsorbed to a CM-52 column at pH 5.7. Therefore, we used a tandem column chromatography of DE-52 and CM-52 at pH 5.7 for the first step of purification. The specific activity of the purified human enzyme was ca. 120 units/mg, which was much lower than that of the rat enzyme (ca. 500 units/mg) probably because of the instability. Details of the enzymological

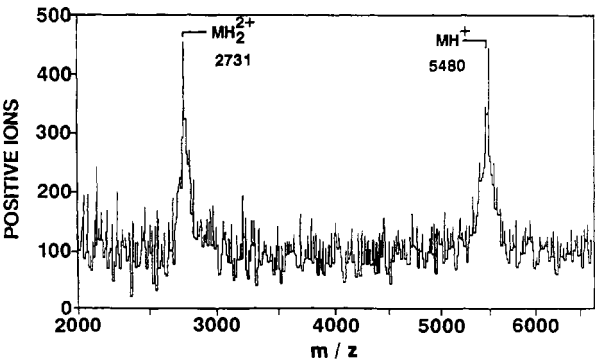

FIGURE 3: Positive-ion californium-252 plasma desorption mass spectrum of fragment M2 (65 pmol). Calculated mass (MH^+) is 5456.

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

peptide residue no.	K1 1–18	K2 19–56	K3 57–104	K4 105–168	K5 169–289
Asp/Asn (D/N)	0.7 (1)	1.9 (2)	5.7 (7)	5.7 (6)	8.3 (9)
Thr (T)	1.2 (1)	1.4 (1)	2.3 (2)	2.1 (2)	5.6 (5)
Ser (S)	2.7 (3)		2.1 (2)	5.4 (6)	3.6 (3)
Glu/Gln (E/Q)	0.9 (1)	6.3 (7)	4.1 (4)	5.4 (5)	17.4 (17)
Pro (P)		3.6 (2)	5.8 (5)	2.3 (2)	7.9 (6)
Gly (G)	2.1 (2)	2.9 (3)	2.5 (2)	6.9 (7)	9.6 (8)
Ala (A)	3.3 (3)	2.7 (2)	5.1 (5)	7.1 (7)	17.0 (16)
Val (V)	0.9 (1)	4.5 (6)	2.7 (3)	4.7 (5)	8.1 (10)
Met (M)		1.0 (1)	1.4 (1)		
Ile (I)		0.5 (1)	2.2 (3)	4.7 (6)	4.0 (5)
Leu (L)	1.0 (1)	5.0 (5)	5.0 (5)	5.0 (5)	14.0 (14)
Tyr (Y)		1.0 (1)		2.0 (2)	5.9 (6)
Phe (F)			2.8 (3)	1.1 (1)	5.1 (5)
His (H)	1.1 (1)			1.0 (1)	4.2 (3)
Lys (K)	0.8 (1)	1.0 (1)	1.6 (2)	0.8 (2)	1.9 (1)
Arg (R)	2.8 (3)	4.8 (6)	3.4 (3)	5.8 (6)	9.3 (9)
PE-Cys (C)			1.7 (1)	0.9 (1)	3.0 (3)
Trp[b] (W)	ND (0)	ND (0)	ND (0)	ND (0)	ND (1)
total res.	18	38	48	64	121
yield (%)	28	56	41	26	19

peptide residue no.	K7 291–313	K8 314–341	K9 342–378	K10 379–388	K12 391–475
Asp/Asn (D/N)		1.6 (2)	1.8 (2)		3.5 (3)
Thr (T)	1.3 (1)		1.4 (1)		4.1 (4)
Ser (S)	2.7 (3)		2.9 (3)		3.3 (3)
Glu/Gln (E/Q)	3.8 (4)	4.2 (5)	3.9 (4)	2.0 (2)	12.8 (13)
Pro (P)	2.1 (2)		7.1 (7)		9.5 (9)
Gly (G)	2.2 (2)	4.4 (5)	1.6 (1)		8.8 (8)
Ala (A)	2.5 (2)	2.0 (2)	4.0 (4)	4.0 (4)	6.6 (5)
Val (V)	1.3 (1)	2.8 (4)	3.6 (4)	1.0 (1)	5.2 (6)
Met (M)	1.6 (2)	2.1 (3)	1.1 (1)		2.5 (3)
Ile (I)					2.8 (3)
Leu (L)	2.1 (2)	1.0 (1)	5.0 (5)	2.3 (2)	9.0 (9)
Tyr (Y)	1.0 (1)	1.7 (2)			2.2 (2)
Phe (F)	1.0 (1)	0.9 (1)	2.0 (2)		5.6 (6)
His (H)	1.0 (1)				1.2 (1)
Lys (K)	0.9 (1)	0.9 (1)	1.0 (1)	1.0 (1)	1.1 (1)
Arg (R)		1.3 (1)	1.5 (1)		6.5 (6)
PE-Cys (C)		0.8 (1)	0.9 (1)		1.1 (3)
Trp[b] (W)	ND (0)	ND (0)	ND (0)	ND (0)	ND (0)
total res.	23	28	37	10	85
yield (%)	39	28	49	53	39

peptide residue no.	K13 476–482	K14 483–489	K15 490–495	intact PE-GPT human 1–495	 rat[c] 1–495
Asp/Asn (D/N)				35.4 (32)	(32)
Thr (T)			1.0 (1)	21.3 (18)	(16)
Ser (S)		0.9 (1)	1.0 (1)	22.3 (25)	(27)
Glu/Gln (E/Q)	1.3 (1)		1.0 (1)	62.3 (64)	(71)
Pro (P)				37.6 (33)	(33)
Gly (G)				37.7 (38)	(36)
Ala (A)		1.3 (1)		46.6 (52)	(44)
Val (V)				38.2 (41)	(39)
Met (M)				12.0 (11)	(16)
Ile (I)				16.5 (18)	(21)
Leu (L)	4.0 (4)	1.0 (1)	1.0 (1)	55.0 (55)	(48)
Tyr (Y)			0.9 (1)	15.0 (15)	(16)
Phe (F)		0.9 (1)	0.9 (1)	20.8 (21)	(22)
His (H)		1.0 (1)		8.7 (8)	(7)
Lys (K)	1.0 (1)	0.9 (1)		18.8 (16)	(19)
Arg (R)	1.0 (1)	1.0 (1)		40.5 (37)	(33)
PE-Cys (C)				14.4 (10)	(14)
Trp[b] (W)	ND (0)	ND (0)	ND (0)	ND (1)	(1)
total res.	7	7	6	495	495
yield (%)	57	52	54		

[a] Results are expressed as residues per peptide or protein by amino acid analysis or, in parentheses, from the sequence (Figure 2). Vapor-phase hydrolysis was carried out with 6 N HCl at 110 °C for 20 h. Each composition was calculated on the basis of the integral value of the residue underlined. [b] Trp was not determined (ND). [c] Taken from Ishiguro et al. (1991).

properties of human liver cytosolic GPT will be published elsewhere. It is well-known that there is a genetic polymorphism in human GPT as reported for the erythrocyte enzyme (Chen et al., 1972) and the liver enzyme (Saha & Bhatta-

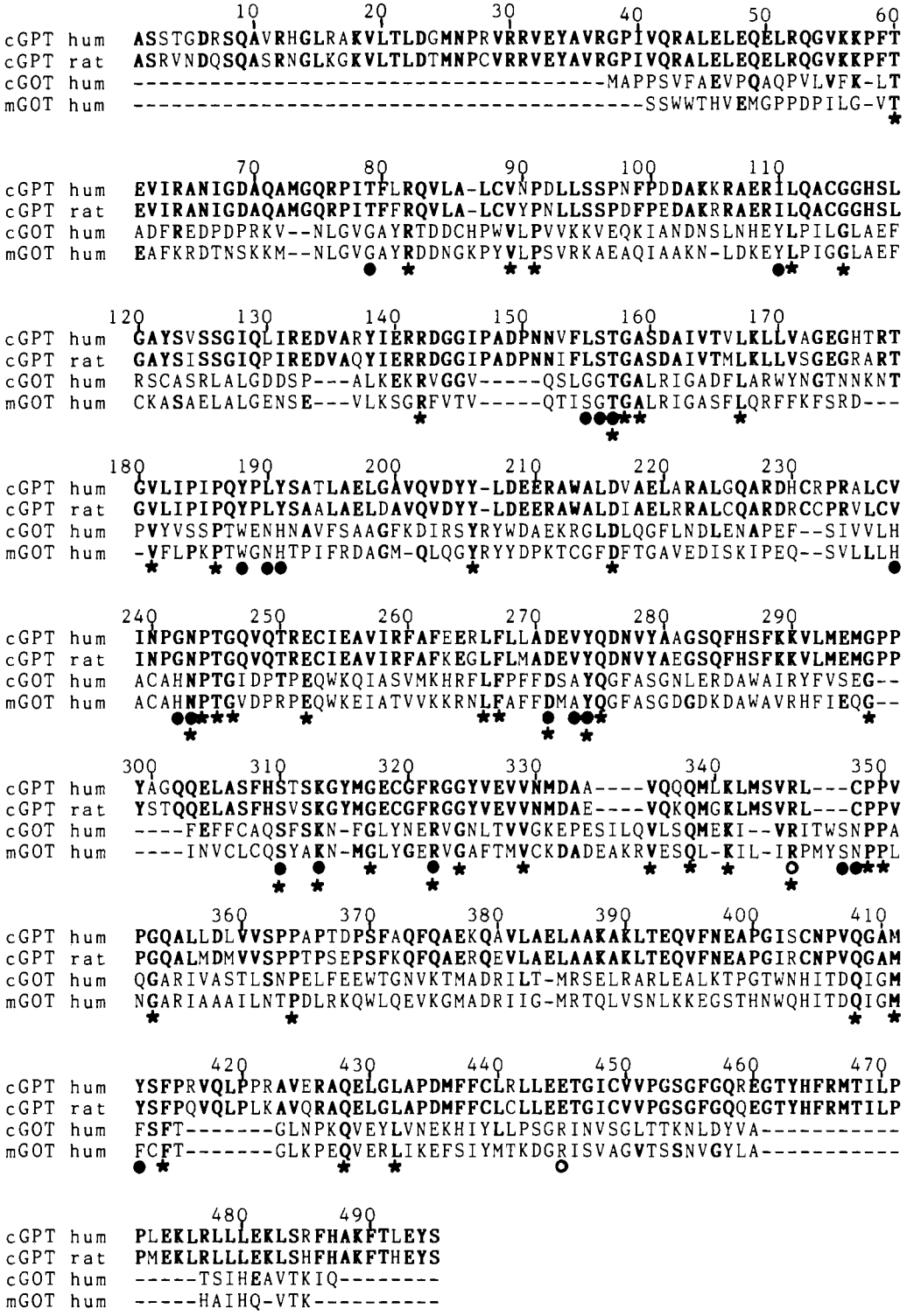

FIGURE 4: Sequence alignment of human GPT, rat GPT, and human GOT. Residue numbers above the alignment represent those of human cytosolic GPT. Bold letters indicate residues identical to human GPT. Asterisks indicate residues conserved among the four sequences. The substrate binding residues in GOT (Arg 292 and Arg 386 in cytosolic pig GOT, Arg 346 and Arg 445 in this figure) are marked with open circles. Filled circles indicate the residues present in the PLP-binding pocket. cGPT hum, cytosolic human GPT; cGPT rat, cytosolic rat GPT (Ishiguro et al., 1991); cGOT hum, cytosolic human GOT (Choudhury et al., 1990); mGOT hum, mitochondrial human GOT (Martini et al., 1985).

charyya, 1989; Kanemitsu et al., 1990) as well. The most popular types among Japanese people seem to be GPT1, GPT2-1, and GPT2 (Ishimoto & Kuwata, 1974; Ueda et al., 1979). The purified enzyme in the present study was probably GPT1 because the predicted isoelectric point was closest to that of the most cationic enzyme. But the cause of the GPT polymorphism remains to be clarified. We are now investigating the other types of human liver cytosolic GPT.

In general, the proof of the sequence of human GPT involved two primary sets of overlapping peptides derived by cleavage at either lysyl or methionyl bonds. As shown in Figure 2, all of the K peptides and their subpeptides and a part of the M peptides were determined by conventional Edman degradation. Mass spectral analysis of the rest of the M peptides and some of the K peptides facilitated the sequence proof by either providing overlaps of K peptides or confirming

Table II: Mass Spectral Analysis of Peptides Derived from GPT[a]

peptide	residue no.	MH^+ calcd	MH^+ obsd	MH_2^{2+} obsd
M1	1–25	2621.9	2623.0	1321
M2	26–73	5456.3	5479.9 (+Na?)	2731
M5	296–317	2209.4	2209.6	1103
M6	317–331	1674.9	1676.5	
M8/9	340–411	7707.1	7719.8	3848
M9	344–411	7221.4	7245.4 (+Na?)	3623
M10	412–436	2794.2	2795.8	1386
M11	437–467	3714.3	3715.6	1860
M11/12	437–495	7101.5	7110.4	3558
K1[b]	1–18	1940.1	1941.4	
K2	19–56	4391.2	4392.8	2197
K7	291–313	2509.9	2509.7	1256
K8	314–341	3187.7	3187.3	1592
K9	342–378	4026.8	4028.3	2025
K1-T1[b]	1–12	1277.4	1277.3	
K1-T2[b]	1–16	1741.0	1741.9	

[a] MH^+ for M peptides were calculated as the lactone forms. [b] Peptides observed in a partial tryptic digest of peptide K1.

the sequences obtained by the conventional method, together with the sequence homology to rat GPT (Ishiguro et al., 1991). Mass spectral analysis was also very useful for identification of the amino-terminal blocking group as in analysis of rat GPT. For some reason, acyl-amino acid releasing enzyme (Tsunasawa & Narita, 1976) failed to remove acetylalanine from the N^{α}-blocked peptide K1. Although Ac-Ala itself meets the enzyme specificity, either the length of the peptide (16 residues; the enzyme generally prefers shorter substrate) or the presence of charged residues near the blocked amino terminus may have prevented the cleavage of the Ala–Ser bond. Information on the precise molecular mass of a peptide is of great value for estimation of posttranslational modifications such as the amino-terminal blocking group and phosphorylation. In the present study, we wanted to know its utility as well as limitations. The data presented here are some of our very first trials before establishing routine procedures. Although there are reports concerning many different ways of use in protein chemistry (Roepstorff et al., 1988; Klarskov et al., 1989), we limited our use of the instrument just to molecular mass determination in the present study. We also limited the sample amount to 100 pmol because the amount of the enzyme (starting material) was limited. The sequence shown in Figure 2 contains regions of relatively weak analytical proofs at overlaps aligning residues 56–57, 264–266, 313–314, 339–341, 466–467, 482–484, and 489–491, but these alignments are supported by either the compositional (Table I) or the mass spectral data (Table II) in addition to the sequence homology to rat GPT (Ishiguro et al., 1991). Thus, the only weak point in the proof of sequence is the lack of direct identification of Trp at residue 213. The presence of Trp was monitored by the UV spectrum obtained with a diode array detector during the HPLC separation. Among the isolated peptides, only K5 and M3 showed the characteristic spectrum of Trp-containing peptides. Trp was placed in the sequence at residue 213 by the specificity of the BNPS-skatole and sequence homology with the rat enzyme.

From the present study, the human GPT subunit is composed of 495 residues. The M_r of this subunit, calculated from the amino acid sequence, is 54479, which is in good agreement with a M_r of about 55000 estimated by SDS–PAGE. As shown in Figure 4, the sequence of human GPT can be aligned with that of rat GPT (Ishiguro et al., 1991) without any gap and is highly homologous to it as expected from the species difference. Out of 495 residues, 435 are identical to each other (87.9% identity). Differences between human and rat GPT are scattered throughout the sequence, but there are several regions with relatively many differences such as residues 3–17, 217–235, 263–269, 334–340, 357–380, and 416–425. Although the tertiary structure of human GPT is not yet known, these regions are likely to be not involved in the active site. As discussed in a previous paper (Ishiguro et al., 1991), the amino acid sequence of cytosolic GPT is quite unique among those of pyridoxal enzymes with known structures registered in an updated protein data base. However, the detailed sequence homology search indicated that the sequence of human cytosolic GPT is significantly homologous to those of human cytosolic GOT and mitochondrial GOT with 14.3% and 12.5% identity, respectively. As shown in Figure 4, the homology is much higher in the middle portion of the molecules than the amino- and carboxyl-terminal regions. From the X-ray crystallographic studies on GOT, the middle portion comprises the active site large domain and the amino- and carboxyl-terminal regions comprise the small domain involved in the substrate specificity. In spite of rather low overall homology between GPT and GOT, most of the crucial amino acid residues providing hydrogen bonds to PLP at the binding pocket in the large domain of both cytosolic and mitochondrial GOT are also retained in both rat and human GPT as shown in Figure 4. This suggests that the tertiary structure of GPT should resemble that of GOT in the middle portion of the molecule comprising the active site large domain, but it may be quite different from GOT in the small domain composed of the amino- and carboxyl-terminal regions and involved in the substrate specificity.

Acknowledgments

We thank Dr. M. Naito, Department of Forensic Medicine, School of Medicine, Fujita Health University for the generous gift of the autopsy liver. We also thank Applied Biosystems for their generous offer of the use of a BioIon 20 biopolymer mass analyzer.

References

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

Chen, S., Giblett, E. R., Anderson, J. E., & Fossum, B. L. G. (1972) *Ann. Hum. Genet. (London) 35*, 401–409.

Choudhury, B. K., Setoyama, C., & Shimada, K. (1990) *Biochem. Int. 22*, 583–591.

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.

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

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.

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

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

Ishiguro, M., Suzuki, M., Takio, K., Matsuzawa, T., & Titani, K. (1991) *Biochemistry 30*, 6048–6053.

Ishimoto, G., & Kuwata, M. (1974) *Jpn. J. Hum. Genet. 18*, 373–377.

Kamoda, N., Minatogawa, Y., Nakamura, M., Nakanishi, J., Okuno, E., & Kido, R. (1980) *Biochem. Med. 23*, 25–34.

Kanemitsu, F., Kawanishi, I., Mizushima, J., & Okigaki, T. (1990) *Clin. Biochem. 23*, 121–125.

Katunuma, N., Mikumo, K., Matsuda, M., & Okada, M. (1962) *J. Vitaminol. 8*, 68–73.
Klarskov, K., Breddam, K., & Roepstorff, P. (1989) *Anal. Biochem. 180*, 28–37.
Klotz, I. M. (1967) *Methods Enzymol. 11*, 576–580.
Landon, M. (1977) *Methods Enzymol. 47*, 145–149.
Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199–11203.
Martini, F., Angelaccio, S., Barra, D., Pascarella, S., Maras, B., Doonan, S., & Bossa, F. (1985) *Biochim. Biophys. Acta 832*, 46–51.
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.
Noguchi, T., Takada, Y. & Kido, R. (1977) *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1533–1542.
Omenn, G. S., Fontana, A., & Anfinsen, C. B. (1970) *J. Biol. Chem. 245*, 1895–1902.
Roepstorff, P., Nielsen, P. F., Klarskov, K., & Hejrup, P. (1988) *Biomed. Environ. Mass Spectrom. 16*, 9–18.
Ruegg, U., & Rudinger, J. (1977) *Methods Enzymol. 47*, 111–116.
Saha, N., & Bhattacharyya, S. P. (1989) *Hum. Hered. 39*, 110–112.
Skoog, B., & Wichman, A. (1986) *Trends Anal. Chem. 5*, 82–83.
Titani, K., & Narita, K. (1964) *J. Biochem. (Tokyo) 56*, 241–265.
Tsunasawa, S., & Narita, K. (1976) *Methods Enzymol. 45*, 552–561.
Ueda, S., Omoto, K., Park, K. S., & Kudo, T. (1979) *Hum. Hered. 29*, 208–212.

Assignment of the Natural Abundance ^{13}C Spectrum of Proteins Using ^{13}C ^{1}H-Detected Heteronuclear Multiple-Bond Correlation NMR Spectroscopy: Structural Information and Stereospecific Assignments from Two- and Three-Bond Carbon–Hydrogen Coupling Constants†

Poul Erik Hansen
Institute for Life Sciences and Chemistry, Roskilde University, P. O. Box 260, DK-4000 Roskilde, Denmark
Received May 23, 1991; Revised Manuscript Received August 7, 1991

ABSTRACT: Proton-detected heteronuclear multiple-bond ^{1}H–^{13}C correlations (HMBC) previously have been used for assignment purposes in a variety of isotopically enriched proteins. In the present study it is demonstrated that the technique yields an almost complete assignment of the natural abundance ^{13}C spectrum of the protein basic pancreatic trypsin inhibitor (BPTI). In addition, the technique permits additional ^{1}H assignments to be made for this well-studied protein. The intensities of observed correlations permit rough estimates to be made of $^{2}J(C,H)$ and $^{3}J(C,H)$ coupling constants. These couplings can be used for conformational studies of both the side chains and the backbone. Intra- and interresidue coupling between CαH and the carbonyl carbon provides information about the backbone angles ψ and ϕ. Side-chain conformations can be determined from both two- and three-bond carbon–hydrogen coupling constants. The present study of BPTI together with its known high-precision solution structure yields an experimental correlation between resonance intensities and secondary structure. The spectra show the potential of the method in analyzing ^{13}C NMR spectra of nonenriched proteins. The method yields ^{13}C NMR chemical shifts, which are versatile parameters to be used to monitor structural changes, titrations, etc.

Determination of solution structures of proteins by ^{1}H NMR[1] has reached a high degree of refinement (Wüthrich, 1986, 1989; Clore & Gronenborn, 1987, 1989; Bax, 1989). At present, interproton distance constraints derived from NOE spectra and dihedral angles derived from ^{1}H–^{1}H couplings are the only observables used in the structure determination process. In the present paper we explore the use of ^{1}H-detected heteronuclear multiple-bond connectivity (HMBC) to obtain qualitative estimates for a third type of parameter, ^{1}H–^{13}C long-range coupling $^{n}J(C,H)$. These couplings provide a rich source of additional conformational information, not yet thoroughly explored for protein structure determination.

For obtaining the highest quality structure, it is important to obtain stereospecific assignments (Guntert et al., 1989; Driscoll et al., 1989). Stereospecific assignments of a significant fraction of the Cβ methylene protons can often be obtained by using a combination of NOE and $J(H,H)$ coupling information, possibly combined with the search of a data bank. As is well known from model peptide studies, $J(C,H)$ couplings combined with $J(H,H)$ information often also provide sufficient unambiguous information to make such stereospecific assignments. In particular, three-bond $^{3}J(C',H\beta)$ couplings in conjunction with $^{3}J(H\alpha,H\beta)$ couplings can be used to make stereospecific assignments of Cβ methylene protons (Hansen et al., 1975; Cowburn et al., 1983). In addition, the interre-

†Support from the Danish Natural Science Research Council is greatly appreciated.

[1] Abbreviations: BPTI, basic pancreatic trypsin inhibitor; γ, gyromagnetic ratio; HOHAHA, Homonuclear Hartmann–Hahn; NOE, nuclear Overhauser enhancement; NMR, nuclear magnetic resonance; T_1, spin–lattice relaxation time; T_2, spin–spin relation time; HMBC, ^{1}H-detected heteronuclear multiple-bond correlation.

0006-2960/91/0430-10457$02.50/0