Volume 320, number 2, 101–106 FEBS 12271 April 1993

Primary structure determination of mono- and diacylglycerol lipase from *Penicillium camembertii*

Kimiyasu Isobe[a,b] and Kiyoshi Nokihara[a,*]

[a]*Enzyme Technology and Natural Products Chemistry Division, GBF-Gesellschaft für Biotechnologische Forschung mbH, Mascheroder Weg 1, W-3300 Braunschweig, Germany* and [b]*Research and Development Division, Amano Pharmaceutical Co., Kunotsubo, Nishiharu-cho, Nishikasugai-gun, Aichi 481, Japan*

Received 14 December 1992; revised version received 19 February 1993

The complete amino acid sequence of mono- and diacylglycerol lipase from *Penicillium camembertii* was determined. This lipase has a single polypeptide chain consisting of 276 amino acid residues with two disulfide linkages. The primary structure was revealed by sequencing the digests of the intact and *S*-pyridylethylated proteins by trypsin, endoproteinase Lys-C and V8 protease. The two-dimensional electrophoresis was also carried out to confirm the internal sequence. The catalytic triad of this lipase was Ser, Asp and His, and one potential *N*-glycosylation site was also revealed.

Penicillium camembertii lipase; Amino acid sequence; Primary structure; Peptide mapping; Internal sequence analysis; Two-dimensional electrophoresis

1. INTRODUCTION

A unique lipase, which is strictly specific to mono- and diacylglycerols but not triacylglycerols, has been isolated from *Penicillium camembertii* U-150 [1]. This lipase has been purified into four active components by concanavalin A–Sepharose column chromatography [2]. The substrate specificity of all four components were similar to each other, but other enzymatic properties were different between three adsorbed and the unadsorbed components. After enzymatic removal of the carbohydrates from the three adsorbed components, their enzymatic and physico-chemical properties were similar to those of the unadsorbed component, and mapping of tryptic digests, as well as amino acid analysis, indicate no difference among the four components. It was, therefore, concluded that all four active components have the same polypeptide backbone, and multiple forms of this lipase are due to the presence of different carbohydrates, and that their carbohydrates contribute to the stability of this lipase but not the enzyme activity [2].

In view of the unique substrate specificity of this lipase, the analysis of the amino acid sequence and the three-dimensional structure may provide strategies for creating a new unique lipase by protein engineering. The present paper describes the primary structure of the unadsorbed component and gives data on the specificity of this lipase.

Correspondence address: K. Isobe, Research and Development Division, Amano Pharmaceutical Co., Kunotsubo, Nishiharu-cho, Nishikasugai-gun, Aichi 481, Japan. Fax: (81) (568) 210 280.

Present address: Biotechnology Instruments Department, Shimadzu Corporation, Nishinokyo-kuwabara-cho 1, Nakagyo-ku, Kyoto 604, Japan.

2. MATERIALS AND METHODS

2.1. *Chemicals*

Trypsin, *Staphylococcus aureus* V8 protease and endoproteinase Lys-C were purchased from Boehringer-Mannheim GmbH (Mannheim, Germany). Anhydrotrypsin-agarose was from Takara Shuzo Co. (Kyoto, Japan). Acetonitrile, trifluoroacetic acid and ammonium bicarbonate were from J.T. Baker Chemicals (Deventer, Netherlands). Guanidine hydrochloride, dithiothreitol and 4-vinylpyridine were from Pierce (Oud-Beijerland, Netherlands), Sigma Chemie GmbH (Deisenhofen, Germany) and Aldrich Chemie (Steinheim, Germany), respectively. All chemicals used were sequencing grade or the highest grade products available.

2.2. *Purification and alkylation of lipase*

The lipase from *P. camembertii* was purified as described in [2] and the unadsorbed component was used for sequence analyses. The intact protein (1 mg) was dissolved in 80 μl of 0.5 M Tris-HCl, pH 7.5, containing 6 M guanidine-HCl and 2 mM EDTA, and reduced with 2 μl of 1.4 M dithiothreitol for 1 h at room temperature. Pyridylethylation was performed by adding 2 μl of 4-vinylpyridine as described in [3]. After the reaction mixture was allowed to stand for 15 min at room temperature, 10 μl of 1.4 M dithiothreitol was added for quenching. The resulting mixture was dialyzed against water and then lyophilized.

2.3. *Tryptic digestion*

Each of the intact and *S*-pyridylethylated proteins (1 mg), dissolved in 100 μl of 50 mM Tris-HCl buffer, pH 8.0, containing 2 mM $CaCl_2$, was digested with 10 μg of TPCK-trypsin at 37°C for 18 h, and terminated at 100°C for 5 min.

2.4. *Endoproteinase Lys-C cleavage*

S-Pyridylethylated protein (1 mg) was dissolved in 100 μl of 25 mM

Tris-HCl buffer, pH 8.5, containing 1 mM EDTA, and then digested with 7.5 μg of endoproteinase Lys-C at 37°C for 1 h followed by incubation with additional 7.5 μg of endoproteinase Lys-C for 17 h.

2.5. S. aureus *V8 protease cleavage*

Each of the intact and *S*-pyridylethylated proteins (1 mg) dissolved in 100 μl of 25 mM ammonium bicarbonate buffer, pH 7.8, was digested with 25 μg of *S. aureus* V8 protease at 25°C for 18 h. The reaction was terminated by heating at 100°C for 3 min.

2.6. *Amino acid analysis*

Amino acid analyses were performed with a Biotronik Amino Acid Analyzer LC 5001 by a ninhydrin monitoring. Tryptic fragments purified by a RP-HPLC were hydrolyzed in gas-phase with 6 N HCl containing 3% phenol at 165°C for 1 h [4].

2.7. *Sequence analysis*

The sequence analysis of both intact and *S*-pyridylethylated proteins were carried out using a Protein Sequencer Model PSQ-1 and PPSQ-10 Systems (Shimadzu Corp., Kyoto, Japan) and a Model 470 A Gas-Phase Sequencer in combination with 120 A PTH-analyzer (Applied Biosystems Inc., Weiterstadt, Germany).

2.8. *Internal sequence determination*

Tryptic digests of *S*-pyridylethylated protein (60 μg) were analyzed on a novel automated apparatus for two-dimensional electrophoresis, Shimadzu Model TEP-1 (Shimadzu Corp., Kyoto, Japan), in which the gel of Swank and Munkres [5] was used for the second dimensional separation [6]. The separated digests were electrophoretically transferred onto PVDF membrane and the excised spots were directly sequenced as described in [6].

2.9. *C-terminal sequence analysis*

Anhydrotrypsin-agarose was used for the separation of the C-terminal fragment according to Kumasaki et al. [7]. Tryptic digests of the *S*-pyridylethylated protein (0.1 mg) were applied onto an anhydrotrypsin-agarose column (1 × 1.2 cm) equilibrated with 50 mM sodium acetate buffer, pH 5.0, containing 20 mM $CaCl_2$. The column was washed with the same buffer and each 0.3 ml of fraction was collected. The unadsorbed fraction was re-chromatographed under the same conditions and purified by reverse-phase (RP)-HPLC to provide the C-terminal fragment.

3. RESULTS AND DISCUSSION

Initially it was found that the present lipase was strongly adsorbed in the stationary phase of RP-HPLC. A tryptic digest of the purified lipase was applied to a C_4 RP-HPLC column and almost all the resulting peptides were sequenced. However, the complete sequence could not be determined from the eluates from this column as there were slight differences between the amino acid analysis and sequence results. The whole sequence was therefore determined by sequencing the intact protein using tryptic, V8 protease and endoproteinase Lys-C digests, which were separated on the C_4 RP-HPLC and from the spots of tryptic digests produced by two-dimensional electrophoresis. The primary structure and sequenced fragments of this lipase are summarized in Fig. 1. The lipase has a single polypeptide chain consisting of 276 amino acid residues with two disulfide linkages and has a calculated molecular mass of 29,825.9 Da. To locate the sulfhydryl group and to determine the amino acid sequence, the *S*-pyridylethylated protein was digested with TPCK-trypsin and the resulting peptides were separated (Fig. 2A). Peptides T-1–T-9 and T-11–T-21 in Fig. 2A were single components, while T-10 was found to consist of two components. The sequence of T-10 was identified by sequencing of fragments generated by endoproteinase Lys-C and *S. aureus* V8 protease; L-2, L-3 and G(i)-5, respectively (Fig. 3). Peptides T-2–T-4, T-6, T-7, T-9, T-11, T-15–T-18 and T-20 have Lys residues, and T-1, T-5, T-10, T-14 and T-19 have Arg residues at their C-terminus, respectively. However, the C-terminus of T-8 and T-12 were Gly and His residues, respectively, which seem to be generated by non-specific cleavage. The peptide fragments, L-1, L-2 and L-3, contributed to the alignment of T-1 and T-7, T-10 and T-6, and T-10 and

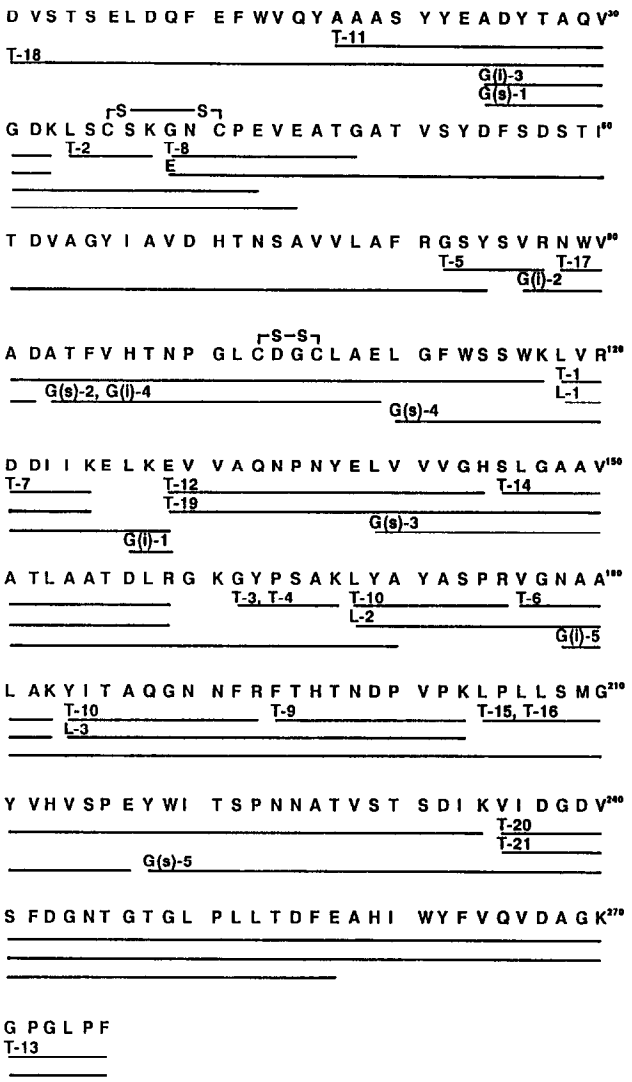

Fig. 1. Primary structure of lipase of *P. camembertii*. The prefixes T, L, G(s) and G(i) denote peptides derived by cleavage of *S*-pyridylethylated proteins with trypsin, endoproteinase Lys-C, *S. aureus* V8 protease, and of intact protein with *S. aureus* V8 protease, respectively. E, sequence obtained by two-dimensional electrophoresis followed by blotting (Fig. 4). The numbers indicate peaks in Figs. 2 and 3.

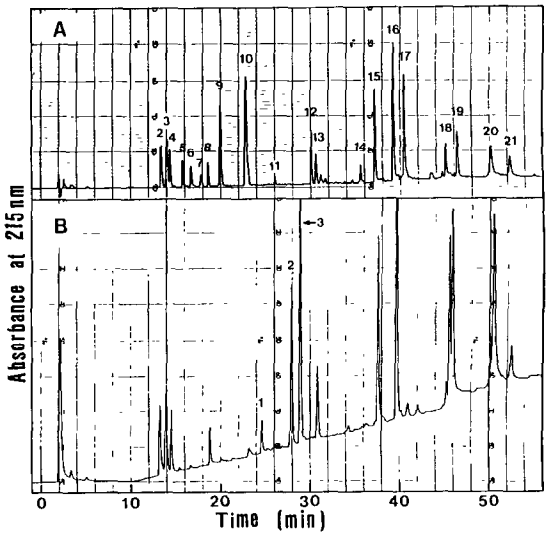

Fig. 2. Separation of peptides obtained by trypsin and endoproteinase Lys-C digestions of lipase from *P. camembertii*. Digests by trypsin (A) and endoproteinase Lys-C (B) were separated using a Vydac C_4 column 214TP5415 with an eluant of 0.1% trifluoroacetic acid for 5 min, followed by a linear gradient (0–60%) with 0.1% trifluoroacetic acid and 0.1% trifluoroacetic acid in acetonitrile at a flow rate of 1 ml/min for 60 min. All peaks were collected manually by monitoring at 215 nm.

T-9, respectively. Alignment at the position 144–145 was confirmed by T-19 and G(s)-3. Positions 126–128 and 160–161 were found by sequencing of fragments G(s)-4 and G(s)-3, respectively. However, the fragment position 49–81 was not clearly detected as tryptic peptides by HPLC. The tryptic digest of the *S*-pyridylethylated protein was therefore applied to an automated apparatus for two-dimensional electrophoresis followed by blotting onto PVDF membrane. Spots of the same three runs were excised, combined and sequenced as described by Nokihara et al. [6]. Three spots were identified (Fig. 4) on the membrane with mass between 15.5 and 18.0 kDa. These spots gave an identical sequence for the first 25 residues, position 39–64. One of them gave the sequence of 46 cycles, and the alignment at positions 38–39 and 81–82 were confirmed by G(i)-3, G(s)-1 and T-5, respectively. These results indicate that the portion of 15.5–18 kDa from position 39 was hydrolysis resistant by trypsin, and this portion around 39–81 seems to be tightly adsorbed on the C_4 reverse-phase column, even though this fragment was generated by tryptic digestion. The disulfide linkages were identified by sequencing of the fragments generated by *S. aureus* V8 protease digestion of both intact and *S*-pyridylethylated protein, as described by Nokihara et al. [8]. Intramolecular disulfide bridges, Cys^{36}–Cys^{41} and Cys^{103}–Cys^{106}, were thus identified in G(i)-3 and G(i)-4, respectively.

The C-terminal peptide of this lipase was involved in fragments T-13 and T-21, which had a Phe residue at the C-terminus. Compared with the DNA sequence [9], three amino acids, Lys, Arg and Val, were deleted at the last C-terminal region in the protein. These three amino acids were also not identified by the amino acid analysis of fragment T-13 or by the comparison of amino acids in fragments T-20 and T-21. Mass spectroscopic analysis of T-13 also showed a molecular mass of 587.3 Da as $(M + H)^+$, in accordance with the sequence Gly-Pro-Gly-Leu-Pro-Phe. In addition the molecular mass of the deglycosylated proteins was estimated to be 29,998 ± 19.0, 29,987 ± 1.8 and 30,013 ± 3.0 Da using matrix-assisted ultraviolet laser desorption/ionization mass spectrometry [10], after glycosylated proteins were treated with endoglycosidase H [2]. As the deglycosylated proteins contain one mole of *N*-acetyl-D-glucosamine, these values were in close agreement with that from the amino acid sequence within the limits of variation of the analysis. These results confirm that the C-terminus of the mature lipase is Phe. Three amino acids identified by the DNA sequence are absent in the mature protein and they are not required for enzyme activity or conformational stability.

From the difference in the tryptic maps between the intact and deglycosylated forms of this lipase, it has been assumed that the *N*-glycosylated carbohydrates should be contained in the fragments T-15 and T-16 [2]. The present sequence analyses revealed that fragments T-15 and T-16 have identical amino acid sequences, including the sequence Asn-X-Thr, where X was identified as Ala. Thus Asn^{225} was identified as the potential *N*-glycosylation site in this lipase. The different elution times of T-15 and T-16 seem to be due to the oxidation of the methionine residue in the latter, which cannot be determined by the Edman degradation.

The comparison of the sequence of this lipase with *Rhizomucor miehei* lipase [11] showed that 33% of the

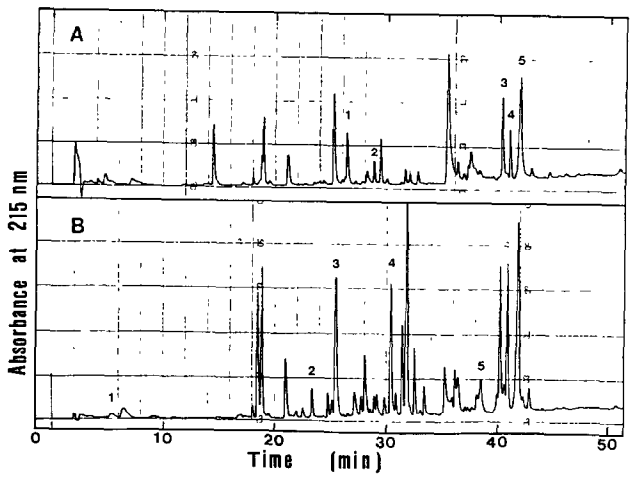

Fig. 3. Separation of peptides obtained by *S. aureus* V8 protease digestion of lipase from *P. camembertii*. Digests of *S*-pyridylethylated (A) and intact (B) proteins were chromatographed on RP-HPLC under the same conditions as Fig. 2.

Table I

Amino acid composition of intact protein and tryptic fragments

Amino acid	Intact protein	Tryptic fragment																			
		T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-10	T-11	T-12	T-13	T-14	T-15	T-16	T-17	T-18	T-19	T-20	T-21
Asx	30.1 (33)						1.0 (1)	1.6 (2)	1.1 (1)	1.9 (2)	2.1 (2)	2.1 (2)	2.2 (2)		1.1 (1)	3.0 (3)	3.0 (3)	3.8 (4)	4.2 (4)	3.4 (3)	6.1 (6)	6.0 (6)
Thr	21.1 (20)								1.0 (1)	1.8 (2)	0.9 (1)	1.2 (1)			1.9 (2)	2.6 (3)	2.6 (3)	2.0 (2)	2.2 (2)	2.0 (2)	3.0 (3)	3.0 (3)
Ser	19.9 (22)		2.3 (2)	0.9 (1)	1.0 (1)	1.6 (2)					0.6 (1)	1.3 (1)			0.9 (1)	4.1 (5)	4.0 (5)	1.9 (2)	2.5 (3)	1.3 (1)	1.3 (1)	1.3 (1)
Glx	18.6 (17)								1.9 (2)		1.0 (1)	2.1 (2)	3.1 (3)			1.0 (1)	1.1 (1)	1.3 (1)	5.8 (6)	3.0 (3)	2.1 (2)	2.3 (2)
Pro	15.1 (13)			1.0 (1)	0.8 (1)				1.0 (1)	1.6 (2)	0.8 (1)		1.3 (1)	2.2 (2)		2.9 (3)	2.8 (3)	1.1 (1)		1.1 (1)	1.1 (1)	2.6 (3)
Gly	23.5 (22)			1.0 (1)	0.8 (1)	1.0 (1)	1.2 (1)		1.5 (2)		1.1 (1)	1.4 (1)	1.2 (1)	2.1 (2)	1.1 (1)	1.1 (1)	1.1 (1)	3.4 (3)	1.2 (1)	2.4 (2)	5.0 (5)	6.2 (7)
Ala	30.0 (30)			1.0 (1)	0.8 (1)		2.7 (3)		1.1 (1)		2.6 (3)	4.6 (5)	1.0 (1)		4.7 (5)	1.0 (1)	1.0 (1)	2.7 (3)	4.6 (5)	5.6 (6)	2.6 (2)	2.7 (2)
Cys/2	4.1 (4)		0.9 (1)						0.7 (1)									(2)				
Val	24.9 (28)	1.0 (1)				1.0 (1)	1.0 (1)		1.0 (1)	1.0 (1)		1.2 (1)	4.1 (5)		1.0 (1)	2.8 (3)	2.8 (3)	2.0 (2)	3.0 (3)	4.7 (6)	3.6 (4)	3.6 (4)
Met	0.5 (1)															0.8 (1)	0.9 (1)					
Ile	7.9 (9)							0.9 (2)			0.8 (1)					1.7 (2)	1.8 (2)				1.2 (2)	1.1 (2)
Leu	20.8 (21)	1.0 (1)	1.1 (1)				1.0 (1)				1.0 (1)		1.0 (1)	1.0 (1)	2.9 (3)	2.9 (3)	2.9 (3)	2.9 (3)	1.1 (1)	3.6 (4)	2.8 (3)	3.5 (4)
Tyr	14.7 (15)			1.0 (1)	0.7 (1)	1.0 (1)					2.9 (3)	3.2 (3)	1.0 (1)			1.9 (2)	1.9 (2)		3.7 (4)	1.1 (1)	1.2 (1)	1.2 (1)
Phe	11.7 (12)									0.9 (1)	1.0 (1)			1.0 (1)				1.8 (2)	1.9 (2)		2.7 (3)	3.5 (4)
His	6.3 (6)									0.9 (1)			1.0 (1)			1.1 (1)	1.0 (1)	2.4 (1)		1.1 (1)	1.1 (1)	1.1 (1)
Trp	7.2 (6)															0.5 (1)	0.7 (1)	2.0 (3)	0.7 (1)		0.7 (1)	0.5 (1)
Lys	11.5 (11)		1.0 (1)	1.1 (1)	2.3 (1)		0.9 (1)	1.0 (1)		1.0 (1)		1.2 (1)				0.9 (1)	0.9 (1)	1.1 (1)	0.7 (1)		0.7 (1)	1.0 (1)
Arg	7.1 (6)	1.0 (1)				1.1 (1)					2.1 (2)				1.3 (1)					1.4 (1)		

Values in parentheses represent theoretical values calculated by their sequence.

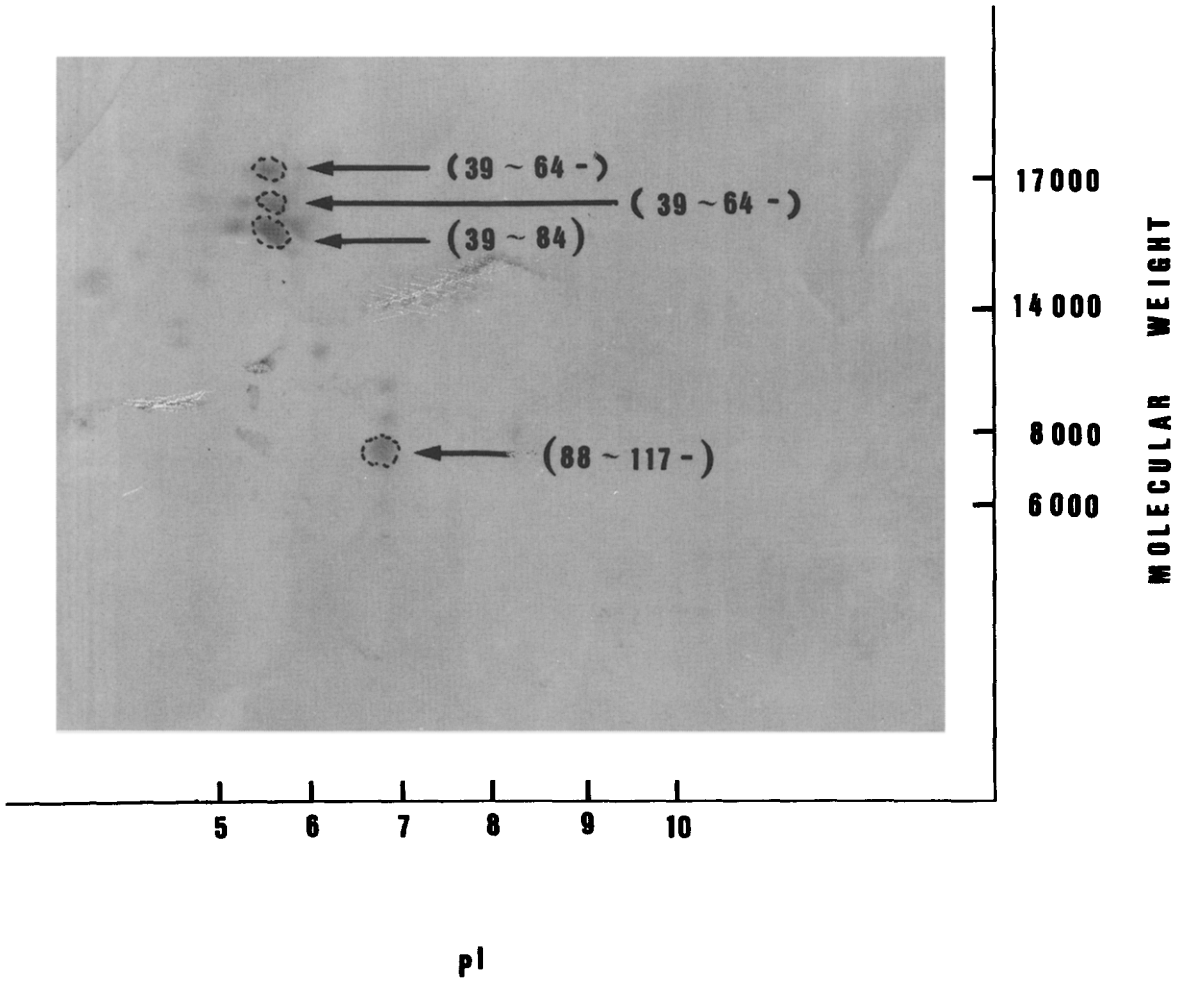

Fig. 4. Separation of tryptic digest of lipase from *P. camembertii* by two-dimensional electrophoresis followed by blotting. Spots indicated by arrow were excised to give the sequence in Fig. 1.

residues were identical and 18% of the exchanges were conserved. The catalytic triad of Asp, His, Ser, which is a well-known conformational feature of serine protease and lipase, was conserved at positions 199, 259 and 145, respectively, and the two disulfide bridges were deleted and one disulfide bridge was newly formed in the present lipase. As the disulfide bridges in the present lipase were very short in length, they may contribute only to local structure stability. Although the sequence homology among the three triacylglycerol lipases from *R. miehei* [12], *Geotrichum candidum* [13] and human pancreas [14] are limited to the region near the active site serine, there is some similarity in their three-dimensional structures. It is therefore presumed that the basic structure of this lipase may be similar with the overall topology of *R. miehei* lipase and that the unique substrate specificity of this lipase may be decided by the difference of the surface structure near the substrate binding region.

Acknowledgements: We thank Prof. R.D. Schmid, Head of Enzyme Technology and Natural Products Chemistry Division of GBF, for his encouragement, R. Getzlaff and R. Christ, GBF, and N. Morita and T. Kuriki, Shimadzu Corp., for their assistance.

REFERENCES

[1] Yamaguchi, S. and Mase, T. (1991) Appl. Microbiol. Biotechnol. 34, 720–725.
[2] Isobe, K., Nokihara, K., Yamaguchi, S., Mase, T. and Schmid, R.D. (1992) Eur. J. Biochem. 203, 233–237.
[3] Thomsen, J. and Bayne, S. (1988) J. Protein Chem. 7, 295–296.
[4] Nokihara, K. and Herbst, F. (1990) in: Peptide Chemistry (N. Yanaihira, ed.) pp. 69–74, Protein Res. Found., Osaka, Japan.
[5] Swank, R.T. and Munkres, K.D. (1971) Anal. Biochem. 39, 462–477.
[6] Nokihara, K., Morita, N. and Kuriki, T. (1992) Electrophoresis 13, 702–707.
[7] Kumazaki, T., Nakako, T., Arisaka, F. and Ishii, S. (1986) Proteins: Struct. Funct. Genet. 1, 100–107.
[8] Nokihara, K., Morita, N., Yamaguchi, M. and Watanabe, T. (1992) Anal. Lett. 25, 513–533.

[9] Yamaguchi, S., Mase, T. and Takeuchi, K. (1991) Gene 103, 61–67.
[10] Hillenkamp, F. and Karas, M. (1990) Methods Enzymol. 193, 280–295.
[11] Boel, E., Huge-Jensen, B., Christensen, M., Thim, L. and Fiil, N.P. (1988) Lipids 23, 701–706.
[12] Brady, L., Brzozowski, A.M., Derewenda, Z.S., Dodson, E., Dodson, G., Tolley, S., Turkenburg, J.P., Christiansen, L., Huge-Jensen, B., Norskov, L., Thim, L. and Menge, U. (1990) Nature 343, 767–770.
[13] Schrag, J.D., Li, Y., Wu, S. and Cygler, M. (1991) Nature 351, 761–764.
[14] Winkler, F.K., D'Arcy, A. and Hunziker, W. (1990) Nature 343, 771–774.