Volume 334, number 2, 153–157 FEBS 13232 November 1993

Molecular cloning of cDNA for the 29 kDa proteinase participating in decomposition of the larval fat body during metamorphosis of *Sarcophaga peregrina* (flesh fly)

Noboru Takahashi, Shoichiro Kurata, Shunji Natori*

Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan

Received 30 July 1993

A cDNA clone for the 29 kDa proteinase participating in tissue disintegration during metamorphosis of *Sarcophaga* was isolated. This proteinase, named *Sarcophaga* cathepsin B, consisted of 256 amino acid residues, and contained three putative N-glycosylation sites. By comparison with other cathepsins B, its unique substrate specificity was partly explained by Ala at position 248.

cDNA cloning; Cathepsin B; Metamorphosis; Fat body decomposition

1. INTRODUCTION

In holometabolous insects, most larval tissues disintegrate during the pupal stage and adult structures develop from imaginal discs [1]. Nothing is known about the molecular mechanism of selective decomposition of larval tissues during metamorphosis [2]. However, we recently demonstrated that in *Sarcophaga peregrina* (flesh fly), hemocytes acquire the ability to dissociate the larval fat body during pupation [3]. These hemocytes were found to express a specific 200-kDa membrane protein [3–5]. These hemocytes are not present in the larval hemolymph, but appear when the larvae pupate and rapidly increase in number at the time of fat body disintegration [4].

We demonstrated that disintegration of the larval fat body was caused by a 29 kDa proteinase that is secreted from the pupal hemocytes on their interaction with the fat body [6,7]. We purified this proteinase from pupal hemocytes to near homogeneity and found that it has unique substrate specificity [8]. Namely, it hydrolyzes both Suc-Leu-Leu-Val-Tyr-MCA and Z-Phe-Arg-MCA, which are specific substrates for chymotrypsin and cathepsin B and L, respectively, but not Suc-Ala-Ala-Pro-Phe-MCA and Z-Arg-Arg-MCA, which are specific substrates for chymotrypsin and cathepsin B, respectively. We isolated a cDNA for the precursor of this 29 kDa proteinase and deduced its amino acid sequence of 324 amino acid residues.

*Corresponding author. Fax: (81) (3) 5684 2973.

2. MATERIALS AND METHODS

2.1. *Cloning of 29 kDa proteinase cDNA*

The 29 kDa proteinase was purified to homogeneity from pupal hemocytes of *Sarcophaga* as described before [8]. This proteinase was digested with lysyl-endopeptidase in 50 mM Tris-HCl buffer (pH 9.0) for 19 h at 30°C at an enzyme/protein ratio of 1:100. The resulting peptides were separated by reverse-phase HPLC, and the partial amino acid sequences of 7 peptides were determined. We also determined the sequence of 16 amino acids from the amino terminal of the 29 kDa proteinase. Two DNA probes, each consisting of a mixture of 23-mer corresponding to Asp-Val-Pro-Glu-Glu-Phe-Asp-Ala and Met-Gln-Asn-Gly-Pro-Val-Glu-Gly in these peptides, were synthesized as GA(T/C)GTICCIGA(A/G)GA(A/G)TT(T/C)GA(T/C)GC (probe A) and ATGCA(A/G)AA(T/C)GGICCIGTIGA(A/G)GG (probe B), respectively.

A cDNA library for NIH-Sape-4, an embryonic cell line of *Sarcophaga*, was constructed with 10 μg of poly(A)$^+$ RNA and 3 μg of vector/ primer DNA by the method of Okayama and Berg [9]. About 40,000 colonies of *E. coli* HB101 carrying recombinant plasmid were transferred onto duplicate sets of nylon filters. After amplification of plasmids by putting the filter on a plate containing 200 μg/ml of chloramphenicol, colonies were lysed with 0.5 N NaOH, and denatured DNA was fixed on the filter. Hybridizations with probe A and B were done in 4 × SSC (1 × SSC = 0.15 M NaCl and 0.015 M sodium citrate)/10 × Denhardt's solution (1 × Denhardt's solution = 0.02%(w/v) each of Ficoll-400, bovine serum albumin, and polyvinylpyrrolidone-40) at 54°C and 55°C, respectively [10]. Colonies that hybridized with both probes were selected. For nucleotide sequencing of cDNA, various deletion derivatives of the DNA fragment were prepared using exonuclease III and mung bean nuclease [11]. Then each deletion derivative was sequenced by the dideoxy chain termination method of Sanger et al. [12]. The nucleotide sequences of both strands were determined.

3. RESULTS

3.1. *Isolation and characterization of a cDNA clone for the 29 kDa proteinase*

We isolated 9 hybridization positive clones by screen-

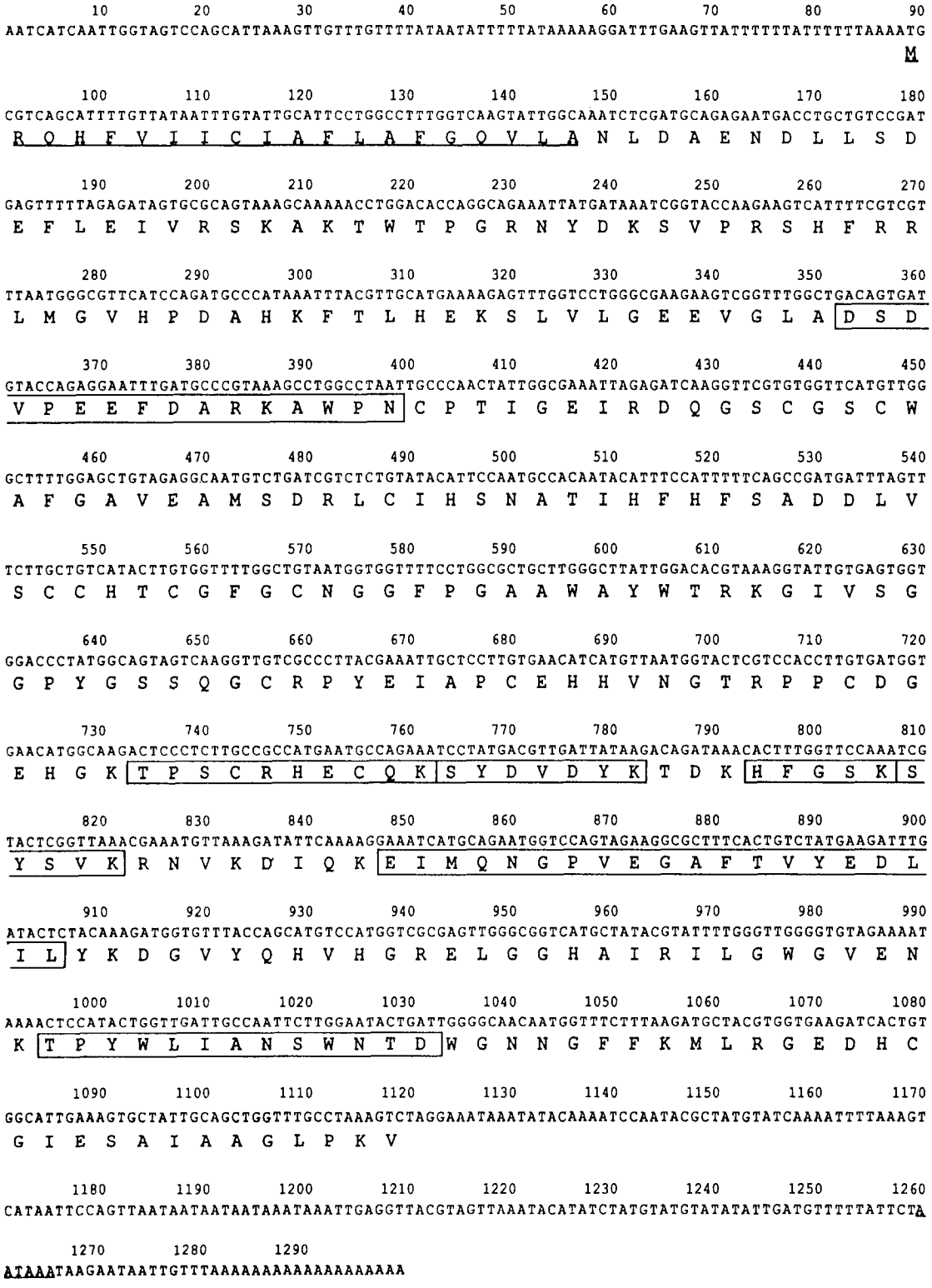

Fig. 1. Nucleotide sequence of cloned cDNA (p29S) encoding the 29 kDa proteinase. The deduced amino acid sequence of the pre-proform of the 29 kDa proteinase is shown below the nucleotide sequence. The putative signal sequence and poly(A) addition signal are underlined, and chemically determined amino acid sequences of the 29 kDa proteinase are boxed. The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank Nucleotide Sequence Databases with the following accession number D16823.

ing 40,000 transformants. These clones contained an insert of about 1.3 kb and gave identical restriction maps, indicating that they were the same. So we determined the nucleotide sequence of one of them, termed p29S. The nucleotide sequence and deduced amino acid sequence of this clone are shown in Fig. 1. This clone contained an open reading frame of 1032 nucleotides, corresponding to 344 amino acid residues. This amino acid sequence contained all the determined sequences of 7 peptides obtained by digesting the 29 kDa proteinase with lysyl-endopeptidase and its amino terminal 16 amino acid residues, indicating that p29S is a cDNA clone of the 29 kDa proteinase.

We assigned the amino terminal amino acid residue of this enzyme as Asp at position 89 and found that this enzyme consists of 256 amino acid residues. The molecular mass of this enzyme was calculated to be 28,267, which was consistent with the value obtained by electrophoretic analysis [8]. Therefore, this proteinase is supposed to be synthesized as a prepro-enzyme consisting of 344 amino acid residues, and then after cleavage of the signal sequence and the pro-segment to form a mature 29 kDa enzyme. We assigned the signal sequence of this enzyme as 20 amino acid residues from the first Met to Ala at position 20, because this region is hydrophobic, as shown in Fig. 2, and the carboxyl terminal amino acid residues of the signal sequences of all secretory proteins of *Sarcophaga* so far examined are Ala [13,14]. Thus the pro-segment of this enzyme may be the 67 amino acid residues from position 21 to 88.

3.2. *Similarity between the 29 kDa proteinase and mammalian cathepsin B*

As shown in Fig. 3, the amino acid sequence of the 29 kDa proteinase showed significant similarity to that of human cathepsin B [15], the percentage similarity being about 68% when gaps were introduced to optimize the alignment. In human cathepsin B, Cys at position 29, His at position 199 and Asn at position 219 are known to be essential for activity. These 3 amino acids are also conserved in the 29 kDa proteinase at positions 32, 202 and 222, respectively. Human cathepsin B contains 14 Cys residues, which, except for those at positions 29 and 240, participate in formation of 6 disulfide bridges [16]. This may also be the case in the 29 kDa proteinase, as it contains 14 Cys residues at similar positions to those of human cathepsin B. These results strongly suggest that the 29 kDa proteinase of *Sarcophaga* is a cathepsin B. Human cathepsin B contains one putative N-glycosylation site [15]. This site is also present in the 29 kDa proteinase, but in addition, this enzyme has two more putative N-glycosylation sites.

Previously, we showed that this 29 kDa proteinase hydrolyses Z-Phe-Arg-MCA, but does not hydrolyze Z-Arg-Arg-MCA, although both are substrates for cathepsin B. Using rat cathepsin B, Hasmain et al. [17]

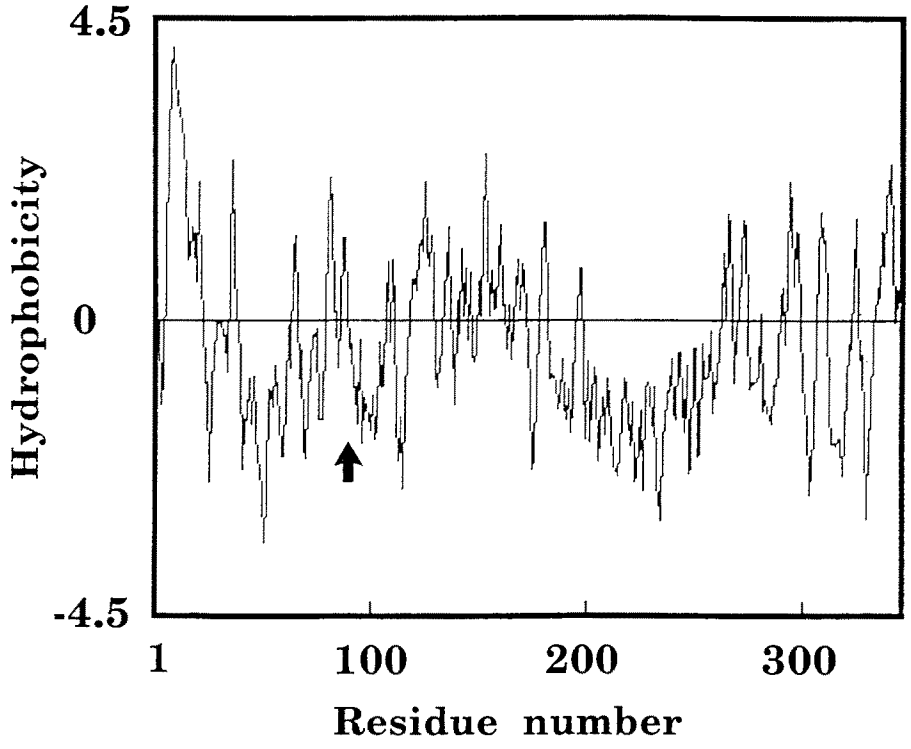

Fig. 2. Hydropathy analysis of the pre-proform of the 29 kDa proteinase. The distribution of hydrophobic and hydrophilic domains was analyzed by the method of Kyte and Doolittle [25]. Hydrophobic and hydrophilic portions are plotted above and below the horizontal line, respectively. The arrow shows the amino terminal of the 29 kDa proteinase.

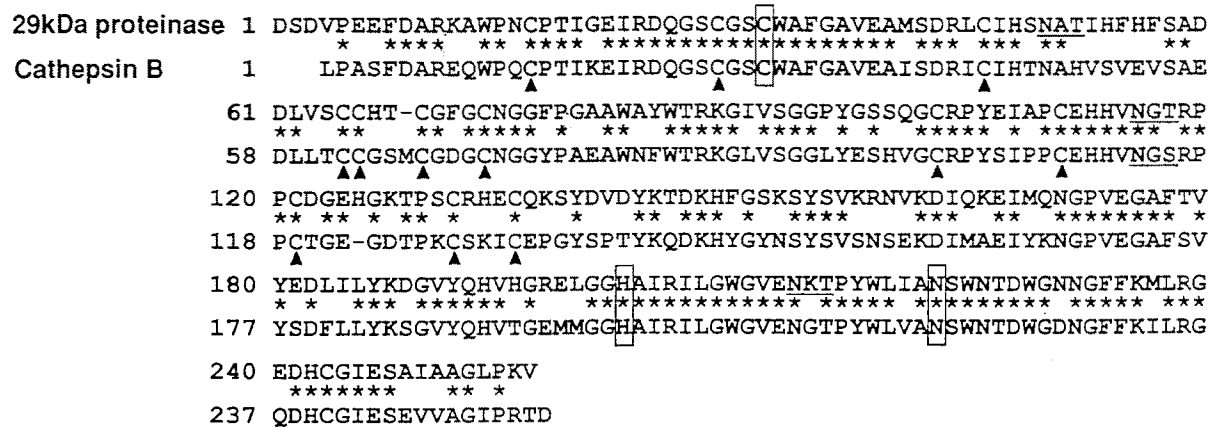

Fig. 3. Comparison of the amino acid sequence of the 29 kDa proteinase with that of human cathepsin B. Amino acid sequences deduced from the corresponding cDNAs are aligned by the FASTA computer program. Amino acid numbers are shown on the left of each line. Asterisks indicate identical amino acid residues. Gaps were introduced to obtain maximal sequence similarity. Conserved amino acid residues forming the active sites are boxed. The arrowheads show Cys residues forming disulfide bridges. Potential N-glycosylation sites are underlined.

recently reported that Glu at position 245 is essential for the hydrolysis of the latter substrate: when this Glu was replaced by Ala, the enzyme lost ability to hydrolyze the latter substrate without loss of activity to hydrolyze the former substrate. As shown in Fig. 4, we found that the corresponding amino acid residue of *Sarcophaga* cathepsin B was Ala, not Glu. Thus, *Sarcophaga* cathepsin B is a naturally occurring mutant with regard to this amino acid residue. This is the first known cathepsin B with an amino acid substitution at this position, and this substitution may explain its unique substrate specificity. However, we cannot explain why this enzyme hydrolyzes one of the two substrates of chymotrypsin that other cathepsins B do not hydrolyze.

4. DISCUSSION

The 29 kDa proteinase purified from pupal hemocytes of *Sarcophaga* has been concluded to be the enzyme that disintegrates the larval fat body during metamorphosis [3,6–8]. In this work, we found by analysis of its cDNA that this proteinase is probably a *Sarcophaga* cathepsin B. Previously, we reported the following three characteristics of this proteinase. (1) It has a narrow pH range and an optimal pH of 6 [8]. (2) It has a unique substrate specificity and hydrolyzes only some of the substrates for chymotrypsin and cathepsin B [8]. (3) It is present in heterogeneous granules in pupal hemocytes and is released from the hemocytes when they interact with the larval fat body [7].

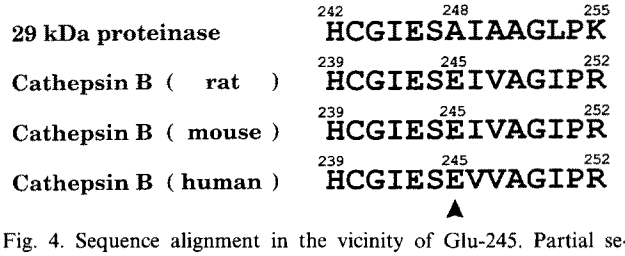

Fig. 4. Sequence alignment in the vicinity of Glu-245. Partial sequences of rat [26], mouse and human cathepsin B [15,27] including Glu-245 are compared with that of the corresponding region of the 29 kDa proteinase. The arrowhead indicates Glu-245, which corresponds to Ala-248 of the 29 kDa proteinase.

These characteristics of the 29 kDa proteinase are not the same as those of mammalian cathepsins B [18–23]. The pH optimum activity of most mammalian cathepsins B is in a much lower pH range. Mammalian cathepsins B hydrolyze both Z-Phe-Arg-MCA and Z-Arg-Arg-MCA, and do not hydrolyze any substrate of chymotrypsin so far tested including Suc-Leu-Leu-Val-Tyr-MCA. Moreover, mammalian cathepsins B are located exclusively in lysozomes and are not normally secreted from the cells.

The difference in substrate specificity of *Sarcophaga* cathepsin B can be explained by the presence of Ala at position 248, but no explanations for the other unique characters of the 29 kDa proteinase could be obtained from its putative amino acid sequence. However, the following two points should be mentioned. (1) Comparison of the amino acid sequences of the 29 kDa proteinase and human cathepsin B showed higher similarities of the amino-terminal 120 and carboxyl-terminal 87 amino acid residues (about 73%) than of those in the intervening region (about 48%). (2) We have no information about the polysaccharide content of this enzyme, but it has two more putative N-glycosylation sites than mammalian cathepsins B. This relatively unique region and the extra polysaccharide chains, if any, might be responsible for the unique characters of the 29 kDa proteinase.

Acknowledgement: This work was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Science and Culture of Japan.

REFERENCES

[1] Bodenstein, D. (1950) in: Biology of Drosophila (Demerce, M., Ed.) pp. 275–367, Hafner, New York.
[2] Lockshin, R.A. (1985) in: Comprehensive Insect Physiology, Biochemistry and Pharmacology (Kerkut, G.A. and Gilbert, L.I., Eds.) Vol. 2, pp. 301–317, Pergamon Press, Oxford.
[3] Kurata, S., Komano, H. and Natori, S. (1989) J. Insect Physiol. 35, 559–565.
[4] Kurata, S., Kobayashi, H. and Natori, S. (1991) Dev. Biol. 146, 179–185.
[5] Kobayashi, H., Kurata, S. and Natori, S. (1991) Insect Biochem. 21, 517–522.
[6] Kurata, S., Saito, H. and Natori, S. (1990) Insect Biochem. 20, 461–465.
[7] Kurata, S., Saito, H. and Natori, S. (1992) Dev. Biol. 153, 115–121.
[8] Kurata, S., Saito, H. and Natori, S. (1992) Eur. J. Biochem. 204, 911–914.
[9] Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161–170.
[10] Hanahan, D. and Messelson, M. (1980) Gene 10, 63–67.
[11] Henikoff, S. (1987) Methods Enzymol. 155, 156–165.
[12] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[13] Takahashi, H., Komano, H., Kawaguchi, N., Kitamura, N., Nakanishi, S. and Natori, S. (1985) J. Biol. Chem. 260, 12228–12233.
[14] Matsumoto, N., Okada, M., Takahashi, H., Ming, Q. X., Nakajima, Y., Nakanishi, Y., Komano, H. and Natori, S. (1986) Biochem. J. 239, 717–722.
[15] Ritonja, A., Popovic, T., Turk, V., Wiedenmann, K. and Machleidt, W. (1985) FEBS Lett. 181, 169–172.
[16] Musil, D., Zucic, D., Turk, D., Engh, R.A., Mayer, I., Huber, R., Popovic, T., Turk, V., Towatari, T., Katunuma N. and Bode, W. (1991) EMBO J. 10, 2321–2330.
[17] Hasnain, S., Hirama, T., Huber, C.P., Mason, P. and Mort, J.S. (1993) J. Biol. Chem. 268, 235–240.
[18] Barrett, A.J. (1977) in: Proteinases in Mammalian Cell and Tissues (Barrett, A.J., Ed.) pp. 181–208, Elsevier, Amsterdam.
[19] Willenbrock, F. and Brocklehurst, K. (1985) Biochem. J. 227, 511–519.
[20] Willenbrock, F. and Brocklehurst, K. (1986) Biochem. J. 238, 103–107.
[21] Koga, H., Yamada, H., Nishimura, Y., Kato, K. and Imoto, T. (1991) J. Biochem. 110, 179–188.
[22] Khouri, H.E., Plouffe, C., Hasnain, S., Hirama, T., Storer, A. and Menard, R. (1991) Biochem. J. 275, 751–757.
[23] Ii, K., Hizawa, K., Kominami, E., Bando, Y. and Katunuma, N. (1985) J. Histochem. Cytochem. 33, 1173–1175.
[24] Kominami, E., Tsukahara, T., Hara, K. and Katunuma, N. (1988) FEBS Lett. 231, 225–228.
[25] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[26] Takio, K., Towatari, T. Katunuma, N., Teller, D.C. and Titani, K. (1983) Proc. Natl. Acad. Sci. USA 80, 3666–3670.
[27] Chan, S.J., Segundo, B.S., McCormick, M.B. and Steiner, D.F. (1986) Proc. Natl. Acad. Sci. USA 83, 7721–7725.