

Cucumisin, a Serine Protease from Melon Fruits, Shares Structural Homology with Subtilisin and Is Generated from a Large Precursor*

(Received for publication, August 9, 1994, and in revised form, September 29, 1994)

Hiroshi Yamagata[‡], Takuya Masuzawa, Yuko Nagaoka, Tatsuji Ohnishi, and Teruo Iwasaki

From the Laboratory of Biochemistry, Faculty of Agriculture, Kobe University, Nada, Kobe 657, Japan

Cucumisin is a thermostable alkaline serine protease that is found in the juice of melon fruits (*Cucumis melo* L.). We have determined the complete nucleotide sequence of a cucumisin cDNA (2,552 nucleotides) and deduced the corresponding amino acid sequence. The open reading frame of the cDNA consists of 731 codons and encodes a large precursor (molecular weight, 78,815) relative to the observed size of mature native cucumisin (67 kDa). Sequence comparisons reveal that cucumisin has several features in common with the microbial proteases of the subtilisin family. The highly conserved sequences to the proximal regions of the catalytic triad amino acids Asp, His, and Ser, together with the substrate binding site in subtilisin, can be found within the deduced amino acid sequence of the protease domain of the cucumisin precursor. Cucumisin is the first known plant protease with such characteristics. Examination of the primary structure of cucumisin revealed that it is synthesized as a precursor, consisting of four functional domains: a possible signal peptide (22 amino acid residues), an NH₂-terminal pro-sequence (88 residues), a 54-kDa protease domain (505 residues), which is the active enzyme domain of the 67-kDa native cucumisin, and a 14-kDa COOH-terminal polypeptide (116 residues), which arises by limited autolysis of the 67-kDa native cucumisin. This structure of cucumisin suggests that it is probably synthesized as an inactive precursor.

Proteolytic enzymes are ubiquitous in biological systems. The physiological roles of endopeptidases (proteinases) and exopeptidases range from precise cleavage of protein precursors in specialized cellular compartments to extensive proteolysis that results in the degradation of proteins to amino acids (1). In plants, most endopeptidases appear to be thiol enzymes, and those within the papain superfamily have been the best characterized (2). In addition to these thiol proteases, an increasing number of endopeptidases that do not utilize sulfhydryl groups for catalytic activity are now being reported in plant tissues (3). However, the serine proteases have been largely overlooked in plants. This is surprising, since they are the best characterized

endopeptidases in mammals (e.g. the trypsin-chymotrypsin family) and microorganisms (e.g. the subtilisin family).

The fruits of some *Cucurbitaceae* contain significant amounts of serine protease activity. A serine protease isolated from the fruit of the melon, *Cucumis melo* L. var. *Prince* is called cucumisin (EC 3.4.21.25) and has been well characterized (4). Cucumisin-like enzymes are also found in the fruits of *Benincasa cerifera* (white gourd) (5) and *Trichosanthes cucumeroides* (snake gourd) (6). These enzymes have been characterized as nonspecific alkaline serine proteases with pH optima around 10, an optimal temperature of about 70 °C, and a fairly broad substrate specificity. The amino acid sequence Gly-Thr-Ser-Met around the reactive serine residue of cucumisin is identical to that of subtilisin, a *Bacillus subtilis* serine protease (7). The molecular mass of all cucumisin-like enzymes has been reported to be around 50 kDa. Recently, however, we have isolated a putative native form of cucumisin from the juice of the central region within developing melons, which has a molecular mass of 67 kDa (8). This 67-kDa enzyme shows limited autolysis to produce a 54-kDa protease and a 14-kDa polypeptide. This autolysis does not cause the loss of caseinolytic activity (8). The 54-kDa protease is resistant to further processing and is therefore thought to be identical to the cucumisin that has been previously reported. In this report, we refer, therefore, to the 67-kDa native enzyme as "67-kDa cucumisin" and the autolyzed form of the native enzyme as "54-kDa cucumisin."

Although the differing linear arrangement of the catalytic triad residues of subtilisin and trypsin-chymotrypsin has attracted much attention as paradigms for a possible convergent evolution of serine proteases (2), the structure of the plant serine protease and its evolutionary relationship to mammal and microbial serine proteases have remained entirely unknown. We present here the complete nucleotide and deduced amino acid sequence of cucumisin and compare it with those of other serine proteases whose sequences are already known. We have found a high degree of homology of the plant serine protease to the proteases of the subtilisin family. A pathway for the processing of the preproenzyme is also proposed.

EXPERIMENTAL PROCEDURES

Plant Material—Musk melons (*Cucumis melo* L. var. *reticulatus* cv. *Teresa*) were cultivated in the green house at the experimental farm attached to the Faculty of Agriculture, Kobe University from March to August. Fruits were tagged upon self-pollination, and developing fruits were harvested at the 10th day after pollination. The central parts of the fruits were separated from the sarcocarp and testae and immediately frozen in liquid nitrogen and stored at –80 °C until use.

Reagents—Oligo(dT)-cellulose (Type II) was obtained from Collaborative Research Co. Ltd. Restriction enzymes and modification enzymes were obtained from Toyobo Co. Ltd. (Tokyo, Japan). cDNA synthesis and Megaprime DNA labeling kits were obtained from Pharmacia Biotech Inc. and Amersham Corp., respectively. Sequenase 2.0 was purchased from U. S. Biochemical Corp. Gigapack II Gold, Bluescript plasmid, and an Exo-Mung deletion kit were from Stratagene. λ gt11 forward and reverse primers were from Takara Shuzo Co. Kyoto, Japan. α -³⁵S-dCTP

* This work was supported by Grant-in-aid for Scientific Research 01560096 from the Ministry of Education, Science and Culture of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence data reported in this paper will appear in the *GSDB*, *DDBJ*, *EMBL*, and *NCBI* nucleotide sequence data bases under the accession number D32206.

[‡] To whom correspondence should be addressed: Laboratory of Biochemistry, Faculty of Agriculture, Kobe University, Rokkodai-cho 1, Nada-ku, Kobe 657, Japan. Tel.: 81-78-803-0673; Fax: 81-78-803-1297 or 81-78-803-0995.

(1,000 Ci/mmol) and [α - ^{32} P]dCTP (3,000 Ci/mmol) were from Amersham Corp. All other commonly available reagents were of analytical grade.

Protein Sequence Determination—The native 67-kDa cucumisin was purified from the developing melon fruit and inactivated with 1 mM diisopropyl fluorophosphate (DFP)¹ at pH 7.2 as described previously (8). The 14-kDa polypeptide that was an autolytic product of native 67-kDa cucumisin was prepared by incubating 67-kDa cucumisin (10 mg) in 0.1 M sodium phosphate buffer, pH 7.2, at 50 °C for 30 min followed by preparative SDS-polyacrylamide gel electrophoresis and electroelution of the protein band from the gel. Both DFP-treated 67-kDa cucumisin and the 14-kDa polypeptide were dialyzed extensively in water with three changes. The NH₂-terminal sequences of native 67-kDa cucumisin and the 14-kDa polypeptide were determined by Edman degradation using a gas-phase sequencer, Applied Biosystems 477A equipped with phenylthiohydantoin analyzer PTA 120A according to the manufacturer's protocols and by the manual DABITC (4-*N,N*-dimethylaminoazobenzene-4'-isothiocyanate-2'-sulfonic acid) method as described by Chang *et al.* (14).

Isolation of Poly(A)⁺ RNA—The central parts of the fruits (200 g) were ground to a fine powder with a mortar and pestle in liquid N₂. The powder was thawed in 40 ml of homogenizing buffer (0.25 M Tris-HCl, pH 7.5, containing 4.5 M guanidium isothiocyanate, 50 mM EDTA-Na₂, and 5% β -mercaptoethanol) at 65 °C. Subsequently, 85 g of solid guanidium isothiocyanate was added, and the mixture was homogenized with a Physcotron (Nichion Rika, Tokyo, Type NS-50) for 3 min. After filtration through sterilized gauze, the filtrate was centrifuged at 8,400 $\times g$ for 10 min at 12 °C, and 4 ml of 20% Sarkosyl was added to the 200 ml of supernatant and incubated at 65 °C for 2 min. After CsCl (0.1 g/ml) was added to the supernatant, the mixture was layered on the cushion of 15 ml of 5.7 M CsCl containing 0.1 mM EDTA-Na₂, pH 8.0, and 0.2% diethyl pyrocarbonate, and ultracentrifuged (Hitachi 55P-72) at 36,000 rpm for 20 h at 18 °C using an RP50-2 rotor. The pellet was resuspended in a solution containing 5 mM EDTA-Na₂, 0.5% Sarkosyl, 0.2% diethyl pyrocarbonate, and 5% β -mercaptoethanol, pH 8.0, and following the addition of 0.25 volume of 1 M potassium acetate to the suspension, total RNA was precipitated with 2.5 volumes of ethanol at -20 °C. Poly(A)⁺ RNA was isolated by oligo(dT)-cellulose column chromatography (9). Template activities of the poly(A)⁺ RNA were measured in a protein-synthesizing system *in vitro* that was prepared from wheat germ (10).

cDNA Cloning—Most of the methods used for molecular cloning were based on those of Sambrook *et al.* (9). Complementary DNA with an *EcoRI/NotI* linker was synthesized with a cDNA synthesis kit, cloned into λ gt11, and packaged using Gigapack II Gold extracts. The host strain used for screening plaques was *Escherichia coli* Y1090r⁺.

Immunoscreening of the cDNA library was performed using a cucumisin monospecific polyclonal antibody by the method of Mierendorf *et al.* (11). Since all five positive clones obtained by immunoscreening code for only a part of the amino acid sequence of cucumisin, rescreening by plaque hybridization was done using one of the clones, λ MSP30, as a probe. Insert DNA of the λ MSP30 clone was amplified by polymerase chain reaction using λ gt11 forward and reverse primers, isolated by agarose gel electrophoresis, and randomly labeled with [α - ^{32} P]dCTP. This was then used to screen the cDNA library as described by Sambrook *et al.* (9). Positive plaques from the cDNA library were purified, the insert DNA was amplified by polymerase chain reaction, and their sizes were determined by agarose gel electrophoresis. Thirty-three positive clones with inserts ranging from 0.5 to 2.6 kilobases were obtained after three rounds of plaque purification.

Subcloning, DNA Sequencing, and Sequence Analysis—The insert cDNA in λ MSP177 clone was isolated from the purified phage by digestion with *NotI* and subcloned into the *NotI* site of plasmid vector pBlue-script SK⁺. This plasmid subclone (pMSP177) was further digested with restriction enzymes and subcloned into the appropriate sites of pBlue-script SK⁺. The nucleotide sequence of single- or double-stranded cDNA inserts was analyzed from both directions by the dideoxynucleotide chain termination method (12) using a Sequenase Version 2.0 sequencing kit as described in the manufacturer's instruction manual. A series of nested deletions was produced from the 5'- and 3'-ends of pMSP177 clones with exonuclease III and mung bean nuclease using an Exo-Mung deletion kit according to a protocol from the manufacturer's instruction manual. Also, the sequences of some regions were determined with specific sequence primers (P1, CGAAGCTTCATACGACAA; P2, CCCACAAGGCCTTACTATCT; and P3, GAAGTATCGCATTCTCTCCA) which were synthesized on an Applied Biosystems 380A oligonucleotide

synthesizer (Foster City, CA). The amino acid sequence was compared with other serine protease sequences in the Protein Identification Resource compiled by the National Biomedical Research Foundation by using the IDEAS program developed by Kanehisa *et al.* (13) and the Hitachi Prosis program (Hitachi Software Engineering Co., Ltd., Yokohama, Japan).

Antibody Preparation and Purification—The purified native 67-kDa cucumisin was inactivated with 1 mM DFP, mixed with Freund's complete adjuvant, and used to immunize Japanese white rabbits. To purify the anti-cucumisin antibody, DFP-treated 67-kDa cucumisin was first coupled to CNBr-activated Sepharose 4B. Then, crude immunoglobulin was prepared by fractionation with 30–60% saturated (NH₄)₂SO₄, desalted by dialysis against phosphate-buffered saline, and applied to the above column. After the column was washed with 5 volumes of coupling buffer (0.1 M NaHCO₃ containing 0.5 M NaCl), bound antibody was eluted with 1 volume of 0.2 M glycine HCl (pH 2.3). Fractions from the column were neutralized with 0.4 M potassium phosphate buffer (pH 8.0), pooled, dialyzed against phosphate-buffered saline, concentrated by reverse dialysis against a saturated solution of (NH₄)₂SO₄, and then again dialyzed against phosphate-buffered saline. The antibody obtained was tested for specificity by Western blot analysis (15) and used for the immunoscreening of the cDNA library.

Analyses of RNA Gel Blots and Genomic DNA Gel Blots—Poly(A)⁺ RNA (10 μ g) was separated in a 1.0% agarose gel containing 0.66 M formaldehyde (9), transferred onto Hybond-N nylon membranes (Amersham Corp.) and UV cross-linked for 3 min. The λ MSP30 cDNA was labeled with [α - ^{32} P]dCTP by random labeling and used as a probe. After the RNA gel blots were prehybridized in a solution containing 6 $\times$ SSC, 5 $\times$ Denhardt's solution, 50% formamide, 0.1% SDS, and 0.1 mg/ml denatured salmon sperm DNA for 1 h at 50 °C, the blots were hybridized with approximately 10⁷ cpm labeled probe in 10 ml of prehybridization solution overnight at 50 °C. Blots were washed twice with 2 $\times$ SSC containing 0.1% SDS at 25 °C for 5 min, then three times with 1 $\times$ SSC containing 0.1% SDS at 50 °C for 30 min and exposed to New RX 50 film (Fuji) with an intensifying screen at -80 °C for 1 or 2 days.

For genomic DNA gel blot analysis, 15 μ g of musk melon genomic DNA, which was prepared from 10 g of young leaves using cetyltrimethylammonium bromide as described by Rogers and Bendich (16), was digested with each restriction enzyme. The DNA fragments were electrophoresed on 0.8% agarose gels, transferred onto a Hybond-N nylon filter, and UV cross-linked. As a quantitative control, λ MSP30 cDNA of severalfold equivalents to that in 15 μ g of melon genomic DNA was electrophoresed in parallel. Amounts were calculated from an estimated weight of the haploid genome of melon of 1.0 pg (16). Hybridization, using λ MSP30 cDNA as a probe which was labeled by random priming, was carried out overnight at 42 °C in 6 $\times$ SSC containing 50% formamide, 0.5% SDS and 0.1 mg/ml denatured salmon sperm DNA. The filters were washed twice for 5 min with 2 $\times$ SSC, 0.1% SDS at 25 °C, for 30 min with 0.1 $\times$ SSC, 0.5% SDS at 37 °C, and for 30 min at 68 °C with the same solution. The filters were exposed as described above for approximately 1 day.

RESULTS

Amino Acid Sequencing—The sequences of the first 13 amino acid residues from the NH₂ terminus of the native 67-kDa cucumisin and the first 7 amino acid residues from the 14-kDa polypeptide that is an autolytic product of native 67-kDa cucumisin were determined by Edman degradation as Thr-Thr-Arg-Ser-Trp-Asp-Phe-Lys-Gly-Phe-Pro-Leu-Thr and Gly-Asp-Tyr-Ser-Ala-Cys-Thr, respectively.

Molecular Cloning of Cucumisin cDNA—When the original cDNA library (1 $\times$ 10⁶ clones) was screened, 5 clones (λ MSP7, 11, 20, 28, and 30), which reacted with the monospecific antibody against 67-kDa cucumisin (see "Experimental Procedures") were identified. Restriction endonuclease mapping and sequencing of these 5 clones indicated that all these clones are likely to represent the transcripts from the same gene and that each contained the coding region for only a part of the cucumisin precursor. Using λ MSP30 (1,860 nucleotides, the clone containing the most 5' sequence) as a probe, the cDNA library (3 $\times$ 10⁵ clones) was rescreened by plaque hybridization, and a further 33 clones were obtained. One of the longest clones, λ MSP177 was subcloned into the *NotI* site of pBlue-script SK⁺, and this pMSP177 clone was subsequently analyzed.

¹ The abbreviation used is: DFP, diisopropyl fluorophosphate.

The complete nucleotide sequences of both strands of the pMSP177 cDNA clone were determined, and all restriction sites were bridged by separate sequencing reactions with different deletion clones and by using specific synthetic sequence primers. The nucleotide sequence and the sequence of the protein predicted from the only long open reading frame are presented in Fig. 1A. The insert contains 2,552 base pairs including the poly(A) tail. The nucleotide sequence of λ MSP30 cDNA was identical to the sequence from C⁴⁰⁶ to A²²⁶⁵ in pMSP177 cDNA. RNA gel blots using the λ MSP30 cDNA as a probe gave one band of about 2.7 kilobases in size (data not shown) suggesting that the λ MSP177 cDNA is an almost full-length clone, taking into consideration the fact that plant mRNAs can contain poly(A) tails up to 200 nucleotides long (17).

Assuming the most 5'-Met codon to be the correct starting point for translation, the ATG at position 22–24 has been tentatively designated the initiation codon on the basis of the mRNA size (Fig. 1A). In this case, the NH₂-terminal amino acid sequence has features of typical eukaryotic signal sequences in that it contains a long stretch of hydrophobic amino acid residues (Fig. 1A). This is not surprising, since cucumisin is likely to be a secreted protein that accumulates in the juice of the central part of melons (8) and probably passes first through the endoplasmic reticulum. There are four other Met codons further downstream that are in frame with the coding sequence, nucleotides 133–135, 184–186, 304–306, and 337–339 (Fig. 1A). However, if any of them was the initiation codon, the resulting NH₂-terminal sequence would not have the characteristic features of a signal peptide. The hydropathy profile of the amino acid sequence of the precursor, obtained according to Kyte and Doolittle (18) with a span of 8 amino acid residues, showed that the hydropathy of the NH₂ terminus (from Met¹ to Phe¹⁶ in Fig. 1A) was high enough for it to function as a signal sequence translocating the protein across the membrane, and such a hydrophobic segment was not observed in the following sequence (data not shown). We therefore feel that the ATG at nucleotides 22–24 is likely to be the initiation codon. The encoded protein has a predicted molecular weight of 78,815 and a total of 731 amino acids, which is much larger than the isolated native cucumisin (67 kDa) (8). Based on the role and structural considerations described by Von Heijne (19), we predict that the signal peptidase processing occurs after Ser²² as shown in Fig. 1A. The predicted signal peptide is followed by 88 amino acids of pro-sequence since the sequence that is identical to the 13 NH₂-terminal amino acid residues of the purified 67-kDa cucumisin appears from Thr¹¹¹ in the deduced amino acid sequence (Fig. 1, A and B). Therefore, cucumisin is probably synthesized as a prepro-enzyme with a 110-amino acid NH₂-terminal region that is not part of the active enzyme. The mature cucumisin protein contains 621 amino acids, and its predicted molecular weight is 66,267, which is consistent with the relative molecular weight of the purified native cucumisin (67,000) as determined by SDS-polyacrylamide gel electrophoresis, bearing in mind also that cucumisin contains 2% carbohydrate (4). The amino acid composition predicted for the mature cucumisin protein is very consistent with that of the purified 67-kDa native cucumisin (8).

Since the 8 amino acid residues from Asn⁵¹⁹ were identical with the reported amino acid sequence around the reactive serine residue (7), Ser⁵²⁵ could be the reactive serine residue in the prepro-cucumisin. The amino acid sequence from Gly⁶¹⁶ was identical to the NH₂-terminal amino acid sequence of the purified 14-kDa polypeptide that is the autolyzed product derived from 67-kDa native cucumisin (8) as shown in Fig. 1A. The molecular weights of the polypeptides from Thr¹¹¹ to Thr⁶¹⁵ and that from Gly⁶¹⁶ to the COOH-terminal Val⁷³¹ were calculated

as 53,827 and 12,458, respectively, which are similar to those of 54-kDa cucumisin and a 14-kDa polypeptide determined by SDS-polyacrylamide gel electrophoresis (8). Also, the amino acid composition predicted for the former polypeptide is very similar to that of the purified 54-kDa cucumisin (8). These facts indicate that limited autolysis occurs at the COOH terminus of the 67-kDa cucumisin and that the COOH-terminal 116 amino acid residues are not essential for protease activity.

The predicted polypeptide contains four potential acceptor sites for Asn-linked glycosylation, three of which occur in the 54-kDa cucumisin and one in the 14-kDa fragment. Two have the sequence Asn-X-Ser at Asn⁴⁶⁶ and Asn⁵⁹³, and two have the sequence Asn-X-Thr at Asn⁵⁴² and Asn⁶⁵². The sequences of Asn⁴⁶⁶-Ala⁴⁶⁷-Ser⁴⁶⁸ and Asn⁵⁵²-Arg⁵⁵³-Thr⁵⁵⁴ are identical to those of the glycopeptide isolated from the purified enzyme (20).

The coding sequence ends with a double stop codon, and the proximal region can form two sequential stem and loop structures (nucleotides 2,205–2,249) as shown in Fig. 1A. In contrast to most mammalian and plant mRNAs, no polyadenylation signals were present within 30 base pairs upstream of the poly(A) tail; however, three AATAAA and one AATAAG sequences were identified further upstream of the poly(A) sequence. It is not certain which sequence facilitates polyadenylation of this mRNA.

The complete sequence of the precursor indicated that autolyzed 54-kDa cucumisin is located in the central part of the primary structure of the precursor. Thus, it was clear that the precursor consists of four portions: the possible NH₂-terminal signal peptide (22 amino acid residues), the pro-sequence (88 residues), the active 54-kDa protease (505 residues), and a COOH-terminal polypeptide with unknown function (116 residues) (Fig. 1B).

Cucumisin Belongs to the Subtilisin Family—Since initial searches of the protein sequence data base identified similarity with several microbial serine proteases of the subtilisin family, the amino acid sequences of these proteins were compared with that of the protein predicted from the pMSP177 cDNA. The similarity is particularly evident within the proposed catalytic domains (Fig. 2). While Ser⁵²⁵ of the catalytic triad of cucumisin could be identified as described above, two catalytic residues, Asp¹⁴⁰ and His²⁰⁴, can be deduced by sequence similarities to subtilisin. The asparagine residue at position 262 in subtilisin is believed to stabilize the tetrahedral transition state (21). In the predicted amino acid sequence of cucumisin, there was a sequence highly homologous to this region (Fig. 2). Cucumisin also shows some similarities to Kex2 (22) and furin (23), which are proprotein and prohormone processing proteases, and tripeptidyl-peptidase II (24), a human subtilisin-like exopeptidase. However, the similarity to the microbial subtilisin family is greater than it is to these Kex2-like proteins and tripeptidyl-peptidase II, *i.e.* the sequence identities between cucumisin and subtilisin BPN' are 58, 60, and 57% in the D, H, and S region, respectively, while those between cucumisin and Kex2 are 25, 33, and 36%, and those between cucumisin and tripeptidyl-peptidase II are 8, 60, and 50%, respectively (Fig. 2). These facts may explain the broad substrate specificity observed for cucumisin (4). The most striking characteristic of the cucumisin sequence, when compared with all the other microbial and mammalian members of the subtilisin family, is the insertion of a long sequence (216 residues) between the stabilizing Asn³⁰⁸ and the reactive Ser⁵²⁵ in the cucumisin relative to all the other proteases, *e.g.* subtilisin BPN' contains only 65 residues in the corresponding region (25). In contrast to its homology with the subtilisin family, cucumisin has no structural relation to the trypsin-chymotrypsin family.

A

CTGATAGCAAAAAGCACTACTATGTTCTTCTCTAATCTTCAAGCTTTCTTTTTAGTCTTTCTTCAGTAATCGACTCGCTTCTAGATTGGATTCTGACGATGATGGAAAAACATT 120
 L I A K S T T M S S S L I F K L F F F S L F F S N R L A S R L D S D D D G K N I
 1
 TATATTGTATACATGGGAGGAAGCTAGAGGATCCTGATTCTGCTCATTTACATCAGGGCAATGTTGGAACAAGTGTGGCAGCACTTTGCTCCAGAATCTGTGCTCCACCTTAC 240
 34 Y I V Y M G R K L E D P D S A H L H H R A M L E Q V V G S T F A P E S V L H T Y
 AAGAGAAGTTTCAACGGATTGCGAGTGAACTTACTGAAGAAGAAGCTGAAAAGATTGCCAGTATGGAGGTGTGGTGTCTGTGTTTTAAATGAAATGAACGAACCTCATACGACAAGA 360
 74 K R S F N G F A V K L T E E E A E K I A S M E G V V S V F L N E M N E L H T T R
 TCATGGGATTTCTGGGTTTTCCACTAAGTCTTCTCGTGAAGTCAAGTGGAAAGCAACATAGTTGTTGGAGTTTGGACCCGGAATCTGGCCGGAATCTCCAGTTTCGACGATGAA 480
 114 S W D F L G F P L T V P R R S Q V E S N I V V G V L D T G I W P E S P S F D D E
 GGGTTCAGTCTCCACCACCAATGGAAGGGCACTGTGAAACCTCAACAACCTTTCGTTGCAACAGAAAAATTATTGGAGCTCGATCATATCAGATGGCCGTCCTTTCACCCGGT 600
 154 G F S P P P P K W K G T C E T S N N F R C N R K I I G A R S Y H I G R P I S P G
 GATGTGAATGGTCCAAGAGACAAATGGACAGGGACGCACTGCATCAACAGCGGTGGTGGTCTAGTTAGCCAGGCAATCTATACGGTCTCGGGCTCGGACGGCAAGAGGAGGA 720
 194 D V N G P R D T N G H G T H T A S T A A G G L V S Q A N L Y G L G L G T A R G G
 GTTCCCTTAGCGGAATCGCTGCATACAAGGTATGCTGAATGATGGTGTCTGATACAGACATTTTCGAGCATATGACGATGCCATTGGCGGATGGAGTCGATATTATCTCTTTCA 840
 234 V P L A R I A A Y K V C W N D G C S D T D I L A A Y D D A I A D G V D I I S L S
 GTGGTGGGGTAAATCCACGACATTATTTCTGTTGATGCCATTGCCATCGGATCTTCCATGCGATAGAGAGGAATATTAACATCAATTTCTGCTGGGAATGGAGGCCATTAATTTCTTC 960
 274 V G G A N P R H Y F V D A I A I G S F H A V E R G I L T S N S A G N G G P N F
 ACCACCGCAAGCTGTCTCCGTGGCTTCTGTCTGTTGCTGCAAGCACCATGGACAGAAAGTTTGTACACAAGTACAGATTGGTAATGGACAGAGCTTTGAGGGAGTTTCAATTAAACACA 1080
 314 T T A S L S P W L L S V A A S T M D R K F V T Q V Q I G N G Q S F Q G V S I N T
 TTTGATAATCAATCTATCCCTTGTAGTGGCGGTGATATCCCAATCTGTTTTCGATAAGTCCACCTCAAGGTTCTGACGGACAAGTCAGTGAATCCCAATTTGTTAAAGGAAAA 1200
 354 F D N Q Y Y P L V S G R D I P N T G D K S T S R F C T D K S V N P L L K G K
 ATTGTTGTTGTGAAGCGAGTTTCGGTCTCATGAATCTTTAAGTCTTGGATGGTGCAGCGGGTGTCTCATGACATCAAAATACGAGGATTATGCAGACTCTATCCCTTGCCTTCT 1320
 394 I V V C E A S F G P H E F F K S L D G A A G V L M T S N T R D Y A D S Y P L P S
 TCCGTTCTCGACCCAAATGATCTCCTTGCCTTTCGTTATATTTAATCAATCGCTCTCTGGTGCACCACTTTCAAGAGTACCACAATCTCAATGCATCTGCACCTGTGTTGTT 1440
 434 S V L D P N D L L A T L R Y I Y S I R S P G A T I F K S T T I L N A S A P V V V
 TCCTTCTCATCCAGGGTCTAATCTGCACTAAAGATGTTATTAAGCCAGACATAAGTGGTCCAGGAGTGAATTTAGCAGCATGGCCTTCTGTTGCACGAGTTGGTGAATCCGT 1560
 474 S F S S R G P N R A T K D V I K P D I S G P G V E I L A A W P S V A P V G G I R
 AGAAACACACTTTTAAATAATCTCAGGAACATCAATGTCTTGTCCACATATCACTGGAATTGCAACCTACGTTAAACATACAATCTACTTGGTCTCTGCTGCCATCAAGTCAGCA 1680
 514 R N T L F N I I S G T S M S C P H I T G I A T Y V K T Y N P T W S P A A I K S A
 CTCATGACAACCGCTTCAACCATGAATGCTAGGTTCAATCCACAGGAGAGTTTGCATATGGTTCAGGCCATGTTAACCCGCTAAAAGCAGTAAGACCTGGGTGGTTATGATGCAAT 1800
 554 L M T T A S P M N A R F N P Q A E F A Y G S G H V N P L K A V R P G L V Y D A N
 GAAAGCGACTACGTTAAATCTTGTGTGGTCAAGGTTACAACCCAGCGGGTTCGAGTATACCCGGAGACTATAGTGCTTGTACTTCTGGTAATCTGGTAGATGAGGATTTAAAC 1920
 594 E S D Y V K F L C G Q G Y N T Q A V R R I T G D Y S A C T S G N T G R V W D L N
 TATCTTCTTTTGGACTTTAGTATCTCTTTCACAGACTTTCAATCAATCTTCAACAGAACTCTCAGAGTGTCCGCTTCAAGCATCAACATATAGAGCTATGATCTCTGCCCCACAA 2040
 634 Y P S F G L S V S P S Q T F N Q Y F N R T L T S V A P Q A S T Y R A M I S A P Q
 GGCTTACTATCTCAGTGAATCTAATGTTCTATCATTTAATGGCTTGGAGATGAGAAATCTTTTACCTTGACAGTTCGAGGATCAATAAAGGATTGTAGTGTGACGCTTTTGGTG 2160
 674 G L T I S V N P N V L S F N G L G D R K S F T L T V R G S I K G F V V S A S L V
 TGGAGCGATGGTGTACACTATGTGAGAAGCCCTATAACCATCAAGTCTAGTTAATGACTTTTGCCTTTTGAATAAAGAAAAAGGAGTTAAATAAAGAAAGTGAAGTGTGG 2280
 714 W S D G V H Y V R S P I T I T S L V
 AAAGAAGTATCGCATTTCTTCCATTACTCTTCTATTTTGTGTTCTTGTCTATAGTAAAGTGGTGGCGACTTTGAATAAGTGTGTCTACTGTCTACTATTAGTAATAAAA 2400
 AGTGTGCTGTATACGTAGCTTTACCTTCGAGAGGTTTTATTAATTTTGATACTTTCCATGGTGTGTTAGCTCTTAGGGGCAATTTTTCTGGTTACACATGATGTGTAATTCGGT 2520
 AAATGAGGCATTTTATCTTCTGTTAAAAAA

B

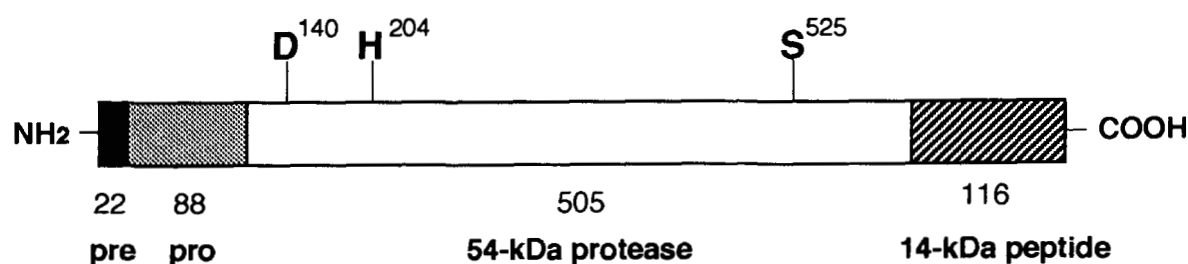


FIG. 1. Nucleotide and deduced amino acid sequence of the pMSP177 cDNA clone and structure of prepro-cucumisin. A, the possible initiation Met of the deduced amino acid sequence is indicated by the number 1. Arrows indicate the possible cleavage site of the signal peptide

D region

1) Cucumisin	132	S N I V V G V L D T G I	143
2) Subtilisin BPN'	131	S N V K V A V I D S G I	142
3) Bacillopeptidase F	219	T G T V V A S I D T G V	230
4) Aqualysin I	158	R G V N V Y V I D T G I	169
5) Proteinase K	136	Q G S C V Y V I D T G I	147
6) Protease B	317	R G V T S Y V I D T G V	328
7) Kex2	167	A G V V A A I V D D G L	178
8) Furin	145	H G I V V S I L D D G I	156
9) TPPII	150	G E V W R A C I D S N E	161

H region

1) Cucumisin	200	D T N G H G T H T A S T A A G	214
2) Subtilisin BPN'	167	D N N S H G T H V A G T V A A	181
3) Bacillopeptidase F	270	D D L A H G T H V T G T M V G	284
4) Aqualysin I	193	D C N G H G T H V A G T I G G	207
5) Proteinase K	170	D G N G H G T H C A G T V G S	184
6) Protease B	353	D G N G H G T H C A G T I A S	367
7) Kex2	209	S D D Y H G T R C A G E I A A	223
8) Furin	190	N D N R H G T R C A G E V A A	204
9) TPPII	207	S G G A H G T H V A S I A A G	231

S region

1) Cucumisin	522	S G T S M S C P H I T G I A	535
2) Subtilisin BPN'	325	N G T S M A S P H V A G A A	338
3) Bacillopeptidase F	448	D G T S M A G P H V S A V A	461
4) Aqualysin I	346	N G T S M A T P H V A G V A	359
5) Proteinase K	326	S G T S M A T P H V A G L A	339
6) Protease B	516	S G T S M A S P H V A G L L	529
7) Kex2	382	G G T S A A A P L A A G V Y	395
8) Furin	365	T G T S A S A P L A A G I I	378
9) TPPII	393	N G T S M S S P N A C G G I	406

Substrate binding site

1) Cucumisin	295	A V E R G I L T S N S A G N G G	310
2) Subtilisin BPN'	249	A V A S G V V V V A A A G N E G	264
3) Bacillopeptidase F	364	W R A A D I F P E F S A G N T D	379
4) Aqualysin I	271	S I A A G V V Y A V A A G N D N	286
5) Proteinase K	253	L Q S S G V M V A V A A G N N N	268
6) Protease B	441	A V E V G I H F A V A A G N E N	456
7) Kex2	301	R D S K G A I Y V F A S G N G G	316
8) Furin	282	R G G L G S I F V W A S G N G G	297
9) TPPII	296	V W K H N I I Y V S S A G N G G	311

FIG. 2. Alignment of cucumisin with other serine proteases. D, H, and S regions represent sequences around the catalytic Asp, His, and Ser residues for the serine proteases. The catalytic residues are labeled by an asterisk. # indicates the stabilizing Asn. Subtilisin BPN' (25) and Bacillopeptidase F (42) are from *Bacillus subtilis*. Aqualysin I is from *Thermus aquaticus* YT-1 (41). Proteinase K is from *Triticum album* (43). Protease B (44) and Kex2 (22) are from *Saccharomyces cerevisiae*. Furin (23) and TPPII (24) are from human. Amino acids that are identical to cucumisin are shown in **boldface**. Amino acid residues of each protease are numbered from the precursor sequence.

Sequence similarities between the deduced amino acid sequence of cucumisin and other plant proteins currently within sequence data bases (Swiss protein, release 27.0; Genpept, release 79.0; PIR, release 38.0) were also examined. One plant protein from *Lilium longiflorum* with GenBank accession number D21815, showing overall similarity to cucumisin (28% identity at amino acid level) could be found. Also, two short polypeptides encoded by the cDNAs from *Arabidopsis thaliana*, with GenBank T04180 and Z18093, have high homology to the corresponding region of cucumisin. There was also significant similarity between the pro-sequence of the cucumisin precursor and a polypeptide encoded by rice cDNA (GenBank D23159). These plant sequences have not been characterized yet, but, if nothing else, this information seems to confirm that cucumisin-

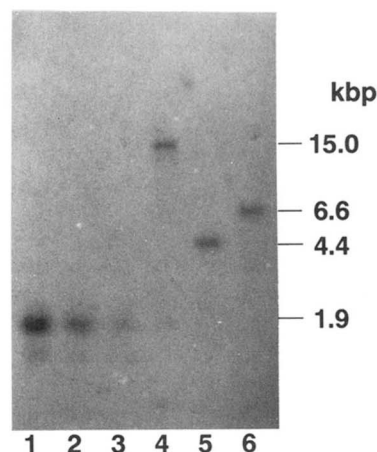


FIG. 3. Hybridization of cucumisin cDNA to restriction endonuclease digested melon leaf DNA. DNA (15 µg) from melon leaves was digested with *Bam*HI (lane 4), *Sac*I (lane 5), or *Xba*I (lane 6), subjected to agarose gel electrophoresis, blotted to Hybond-N nylon membranes and probed with [α - ^{32} P]dCTP-labeled λ MSP30 cDNA. λ MSP30 cDNA was also electrophoresed on the gel. Each amount of loaded λ MSP30 corresponds to twice (lane 1), equal (lane 2), and a half (lane 3) equivalent to the 15 µg of genomic DNA of melon, which was calculated from the weight of the haploid genome to be 1.0 pg (16). The sizes of DNA fragments were determined by DNA standards run in parallel (not shown).

like proteases are widely distributed in several plant species other than *Cucurbitaceae*.

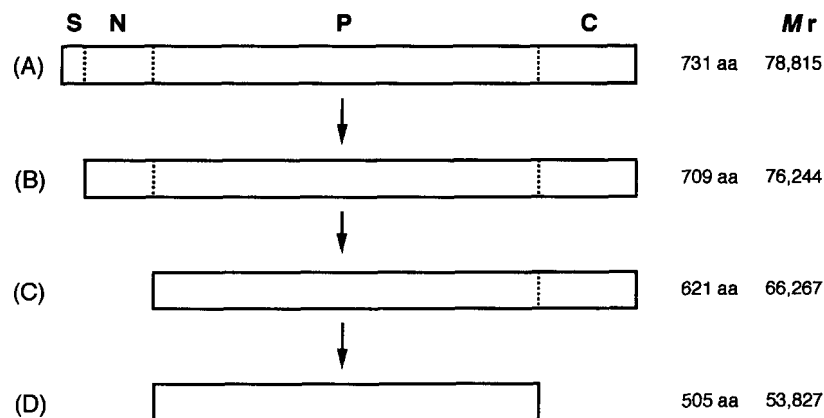
Genomic DNA Gel Blot Analysis—Genomic DNA gel blot analysis, shown in Fig. 3, was performed to determine the complexity of cucumisin-encoding genes. A blot containing musk melon genomic DNA digested with several restriction enzymes was probed with λ MSP30 cDNA. DNA digested with each of the enzymes *Bam*HI, *Sac*I, and *Xba*I gave rise to only a single band at 15.0, 4.4, and 6.6 kilobase pairs, respectively, which hybridized to λ MSP30 cDNA. Although the λ MSP30 cDNA probe has a *Sac*I site near the 5' terminus, we could not detect any other bands in the *Sac*I fragment of the genomic DNA that hybridized to the probe. The most likely explanation for this is that the *Sac*I fragment of the probe is too small (155 base pairs) to detect the hybridization. All of the fragments appeared to hybridize with equal intensity as the single copy reconstructions (Fig. 3). These results indicate that a single gene encoding cucumisin is present per haploid genome in *Cucumis melo* L.

DISCUSSION

Sequence Similarity between Cucumisin and Subtilisins—The COOH-terminal 621 amino acids deduced from the nucleotide sequence of λ MSP177 define the protease domain of cucumisin, which is very closely related to the proteases of the subtilisin family but not to those of the mammalian trypsin-chymotrypsin family. Cucumisin is the first plant protease of the subtilisin family for which the complete primary structure has been clarified. Serine proteases of the subtilisin family have traditionally been thought to be restricted to prokaryotes. Recently, however, subtilisin-like proteases have been found in

(small arrowhead), the processing site of the pro-sequence (vertical arrow) and the autocleavage site of native 67-kDa cucumisin (large arrowhead). The amino acid sequences underlined by a straight line and a double straight line correspond to the NH₂-terminal sequences determined by amino acid sequencing of 67-kDa native cucumisin and the 14-kDa polypeptide that was generated by the autolysis of the 67-kDa native cucumisin, respectively. Amino acid sequences underlined by a dotted line and two double dotted lines correspond to the reported proximal sequences around the reactive serine residue (7) and two glycopeptides (20), respectively. Asterisks identify the Asp, His, and Ser residues corresponding to the catalytic triad in the homologous subtilisins. Underlined nucleotide sequences indicate the putative polyadenylation sequences. Horizontal arrows over the nucleotide sequence indicate possible stem and loop structures. Diamonds indicate the double termination codons. B, diagram of the four-domain structure of prepro-cucumisin, from NH₂ to the COOH terminus; a possible signal peptide (22 amino acid residues), NH₂-terminal pro-sequence (88 residues), mature autolyzed 54-kDa cucumisin (505 residues), and COOH-terminal 14-kDa fragment (116 residues).

FIG. 4. Diagrammatic representation of the proposed processing pathway of the cucumisin precursor. The open bars correspond to the primary sequence of the polypeptide forms. Dotted lines indicate positions of cleavages to follow. S, signal sequence; N, NH₂-terminal pro-sequence; P, protease domain; C, COOH-terminal fragment. A, prepro-cucumisin; B, pro-cucumisin with NH₂-terminal prosequence and COOH-terminal segment; C, 67-kDa native cucumisin; D, 54-kDa autolyzed cucumisin.



lower eukaryotes such as fungi and yeast. Furthermore, it was reported that mammalian endoproteases that have structural identity to yeast Kex2 protease are responsible for cleavage of some biochemically and biomedically important proproteins and prohormones (26). However, cucumisin has a higher similarity with prokaryotic subtilisin type enzymes than any of the mammalian Kex2-like enzymes and tripeptidyl-peptidase II (Fig. 2), possibly reflecting the broader substrate specificity of cucumisin in similarity to the bacterial enzymes. The discovery of cucumisin as a new member of the subtilisin family raises intriguing questions regarding the evolutionary descent of the plant subtilisin-like serine protease. Furthermore the identification of several other proteins in plants that have similarity to cucumisin suggests its widespread occurrence in plants. Since a plant polypeptide hormone has recently been found to be derived from a prohormone precursor in a similar way to animal polypeptide hormones (27), some of the subtilisin-type serine proteases may function in its processing, in a similar fashion as occurs with proproteins and prohormones in mammals (26).

The Structure of Prepro-cucumisin—We found that the precursor of cucumisin comprises four functional domains (Fig. 1): the possible signal peptide (22 amino acid residues), the NH₂-terminal pro-sequence (88 residues), mature autolyzed 54-kDa cucumisin (505 residues), and the COOH-terminal polypeptide (116 residues), which is released from the 67-kDa native cucumisin by autolysis during fruit development.

Since cucumisin is likely to be a secreted protein, we would expect to find a secretory peptide region that is cleaved from the mature protein. The NH₂-terminal amino acid sequence indeed has the general characteristics of signal sequences that direct cotranslational delivery to the lumen of the endoplasmic reticulum. The presumed cleavage site of the signal peptide would be predicted to follow Ser²². Although our knowledge of the molecular mechanisms of compartmentalization and secretion of plant secretory proteins is poor, protein secretion is thought to occur by a default mechanism in plant cells as in mammalian and yeast cells, *i.e.* secretory proteins do not require positive targeting signals to proceed along the secretion pathway (28). We could not detect any homology in prepro-cucumisin with known targeting signals or transit peptides necessary for protein transport to specific organelles such as vacuoles, nuclei, mitochondria, or chloroplasts. This also supports the idea that cucumisin is a secreted protein. Plant thiol proteases also have the signal peptide, for example, the papain precursor of *Carica papaya* has a signal peptide of 26 amino acid residues and is secreted into the latex (29, 30), and endopeptidase B and aleurain of barley seed have 21 and 22 residues of signal peptide, respectively (31).

The possibility of an 88-residue propeptide in the cucumisin

precursor raises questions as to its biological function and the cellular site and mechanisms of its proteolytic processing. In maturing melons the main site of synthesis of protease appears to be in locular tissue just around the seeds,² but most of the protease activity is located in the juice of the central part of the fruits from a very early stage of fruit development onward (8), suggesting that the newly synthesized enzyme is immediately secreted into the juice. The most obvious possibility regarding the pro-peptide function is that it keeps the enzyme inactive and thus protects proteins involved in the secretory pathway from nonspecific proteolytic degradation. If this were the case, one would also expect the final removal of the pro-peptide to occur in the juice following the secretion of the proprotease zymogen from the cells around the seeds. All other precursors of plant proteases so far reported have a pro-sequence; for example, papain precursor has a pro-sequence of 107 amino acid residues (30), and endopeptidase B and aleurain have 111 and 119 residues of pro-sequence, respectively (31).

Another possible function of the propeptide is that it is required for correct folding of the catalytic domain. This has been shown for the pro-sequences of subtilisin (32) and for the α -lytic protease of *Lyso bacter enzymogenes* 495 (33). In the subtilisins secreted by the *Bacillus* species, the inactive precursor of each subtilisin has a signal peptide and an NH₂-terminal pro-sequence (25, 34–36). For example, subtilisin E has a 29-residue signal sequence and a 77-residue prosegment (34), and the mature active protein is formed after self-cleavage of the propeptide (37). The pro-sequence of subtilisin E is essential for guiding the appropriate folding of the active conformation of the enzyme (38), and this function of the pro-sequence can be effective even intermolecularly *in vitro* (32). Random mutagenesis experiments have shown that the pro-peptide consists of a few functional regions that interact with specific regions of mature subtilisin during the folding process (39). Since the pro-peptide is not a part of the final mature subtilisin, it is proposed to function as a built-in intramolecular chaperone for production of an active enzyme (40). Also, the pro-peptide can inhibit the mature enzyme specifically and competitively *in vitro* (40). The NH₂-terminal pro-sequences of several subtilisins contain a common sequence, Tyr-Ile-Val-Gly-Phe-Lys, in the NH₂-terminal region (36). Comparison of the NH₂-terminal pro-sequence of cucumisin and subtilisins reveals only one short homologous sequence, Tyr-Ile-Val (residues 34–36 in cucumisin shown in Fig. 1A). Thus, although the pro-sequence of the cucumisin precursor has no notable structural homology to that of subtilisin, it is probably essential for the production of active cucumisin, and the processing of the pro-peptide may in

² H. Yamagata, K. Hanamori, and T. Iwasaki, unpublished observations.

fact occur by an autocatalytic mechanism as was found for subtilisin.

The four-domain structure of prepro-cucumisin is more similar to the precursors of aqualysin I from *Thermus aquaticus* YT-1 and bacillopeptidase F from *Bacillus subtilis* than it is to the other subtilisins, which do not have a COOH-terminal pro-sequence. Indeed the amino acid sequence of the COOH-terminal 14-kDa polypeptide of the 67-kDa native cucumisin has partial homology to those of aqualysin I and bacillopeptidase F. However, in contrast to the aqualysin I and bacillopeptidase F COOH-terminal sequences, which are cleaved soon after the pro-proteases are secreted to the extra cellular medium (41, 42), the COOH-terminal 14-kDa polypeptide of cucumisin is not released immediately after secretion; while in the fruits at the 10th day after pollination almost all the cucumisin is in the 67-kDa form (8), the 54-kDa form becomes much greater than the 67-kDa form in ripened fruits at the 50th day after pollination.² Since 54-kDa cucumisin is as active as 67-kDa cucumisin, the 14-kDa peptide is not required for the protease activity of cucumisin (8). However, it is not known whether the COOH-terminal sequence is important for secretion of the protein into the extracellular juice from the cells in which the precursor of cucumisin is synthesized. In this regard, it is known that the COOH-terminal 290 amino acids of bacillopeptidase F are not required for either catalytic activity or secretion (42), while the COOH-terminal pro-sequence (105 residues) of Aqualysin I is thought to function in *T. aquaticus* for secretion through the outer membrane of the cell (41).

The processing of prepro-cucumisin can be summarized as shown in Fig. 4. First, prepro-cucumisin (the 79-kDa polypeptide) produced in the cells of locular tissue around the seeds is translocated to the lumen of the endoplasmic reticulum with the aid of the signal peptide. Second, the pro-cucumisin (76 kDa) is probably secreted to the juice of the central part of the fruit. Third, the NH₂-terminal pro-sequence is removed, and mature active 67-kDa cucumisin is produced. In the later stages of fruit development, the 14-kDa fragment localized in the COOH terminus of the 67-kDa cucumisin is released by the proteolytic activity of cucumisin itself, and the 54-kDa autolyzed form of cucumisin is produced in the juice.

The accumulation of the large amount of cucumisin protein in melons (more than 10% of total fruit protein (8)) and the single copy character of the corresponding gene (Fig. 3) suggests that the cucumisin gene contains a strong promoter, which also controls the fruit-specific expression of cucumisin. In addition to being of fundamental interest, these facts suggest that the cucumisin promoter may be potentially useful for genetic engineering of fruits in the future.

Acknowledgments—We thank Dr. Bunzo Mikami for protein sequence determination and Drs. Diana R. Horvath and Chris Bowler for the computer analysis and critical reading of this manuscript.

REFERENCES

- Bond, J. S., and Butler, E. (1987) *Ann. Rev. Biochem.* **56**, 333–364
- Glazer, A. N., and Smith, E. L. (1971) in *The Enzymes*, 3rd Ed., (Boyer, P. D.,

- ed) Vol. 3, pp. 501–546, Academic Press, New York
- Ryan, C. A. (1981) in *The Biochemistry of Plants*. Vol. 6, pp. 321–350, Academic Press, New York
- Kaneda, M., and Tominaga, N. (1975) *J. Biochem. (Tokyo)* **78**, 1287–1296
- Kaneda, M., and Tominaga, N. (1977) *Phytochemistry (Oxf.)* **16**, 345–346
- Kaneda, M., Sobue, A., Eida, S., and Tominaga, N. (1986) *J. Biochem. (Tokyo)* **99**, 569–577
- Kaneda, M., Ohmine, H., Yonezawa, H., and Tominaga, N. (1984) *J. Biochem. (Tokyo)* **95**, 825–829
- Yamagata, H., Ueno, S., and Iwasaki, T. (1989) *Agric. Biol. Chem.* **53**, 1009–1017
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Yamagata, H., Sugimoto, T., Tanaka, K., and Kasai, Z. (1982) *Plant Physiol. (Rockv.)* **70**, 1094–1100
- Mierendorf, R. C., Percy, C., and Young, R. A. (1987) *Methods Enzymol.* **152**, 458–469
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467
- Kanehisa, M., Klein, P., Greif, P., and DeLisa, C. (1984) *Nucleic Acids Res.* **12**, 417–428
- Chang, T. Y., Brauer, D., and Wittmann-Liebold, B. (1978) *FEBS Lett.* **93**, 205–214
- Towbin, H., Staehelin, T., and Gordon, J. (1979) *Proc. Natl. Acad. Sci. U. S. A.* **76**, 4350–4354
- Rogers, S. O., and Bendich, A. J. (1986) *Plant Mol. Biol.* **5**, 69–76
- Grierson, D., and Covey, S. (1984) *Plant Molecular Biology*, pp. 22–46, Blackie and Son Limited, Bishopbriggs, Glasgow, Scotland
- Kyte, J., and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Von Heijne, G. (1991) in *Plant Molecular Biology 2* (Herrmann, R. G., and Larkins, B., eds) pp. 583–593, Plenum Press, New York
- Kaneda, M., Kamikubo, Y., and Tominaga, N. (1986) *Agric. Biol. Chem.* **50**, 2413–2414
- Robertus, J. D., Kraut, J., Alden, R. A., and Birktoft, J. J. (1972) *Biochemistry* **11**, 4293–4303
- Mizuno, K., Nakamura, T., Ohshima, T., Tanaka, S., and Matsuo, H. (1988) *Biochem. Biophys. Res. Commun.* **156**, 246–254
- Wise, R. J., Barr, P. J., Wong, P. A., Kiefer, M. C., Brake, A. J., and Kaufman, R. J. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **87**, 9378–9382
- Tomkinson, B., and Jonsson, A.-K. (1991) *Biochemistry* **30**, 168–174
- Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A., and Chen, E. Y. (1983) *Nucleic Acids Res.* **11**, 7911–7925
- Barr, P. J. (1991) *Cell* **66**, 1–3
- McGurl, B., Pearce, G., Orozco-Cardenas, M., and Ryan, C. A. (1992) *Science* **255**, 1570–1573
- Chrispeels, M. J., Dickinson, C. D., and Tague, B. W. (1991) in *Plant Molecular Biology 2* (Herrmann, R. G., and Larkins, B., eds) pp. 575–582, Plenum Press, New York
- Cohen, L. W., Coghlan, V. M., and Dihel, L. C. (1986) *Gene (Amst.)* **48**, 219–227
- Vernet, T., Tessier, D. C., Richardson, C., Laliberte, F., Khouri, H., Bell, A. W., Storer, A. C., and Thomas, D. Y. (1990) *J. Biol. Chem.* **265**, 16661–16666
- Holwerda, B. C., Padgett, H. S., and Rogers, J. C. (1992) *Plant Cell* **4**, 307–318
- Zhu, X., Ohta, Y., Jordan, F., and Inouye, M. (1989) *Nature* **339**, 483–484
- Silen, J. L., and Agard, D. A. (1989) *Nature* **341**, 462–464
- Stahl, M. L., and Ferrari, E. (1984) *J. Bacteriol.* **158**, 411–418
- Vasantha, N., Thompson, L. D., Rhodes, C., Banner, C., Nagle, J., and Filpula, D. (1984) *J. Bacteriol.* **159**, 811–819
- Jacobs, M., Eliasson, M., Uhlen, M., and Flock, J.-I. (1985) *Nucleic Acids Res.* **13**, 8913–8926
- Powers, S. D., Adams, R. M., and Wells, J. A. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3096–3100
- Ikemura, H., Takagi, H., and Inouye, M. (1987) *J. Biol. Chem.* **262**, 7859–7864
- Kobayashi, T., and Inouye, M. (1992) *J. Mol. Biol.* **226**, 931–933
- Ohta, Y., Hojo, H., Aimoto, S., Kobayashi, T., Zhu, X., Jordan, F., and Inouye, M. (1991) *Mol. Microbiol.* **5**, 1507–1510
- Terada, I., Kwon, S.-T., Miyata, Y., Matsuzawa, H., and Ohta, T. (1990) *J. Biol. Chem.* **265**, 6576–6581
- Wu, X.-C., Nathoo, S., Pang, A. S.-H., Carne, T., and Wong, S.-L. (1990) *J. Biol. Chem.* **265**, 6845–6850
- Gunkel, F. A., and Gassen, H. G. (1989) *Eur. J. Biochem.* **179**, 185–194
- Moehle, C. M., Tizard, R., Lemmon, S. K., Smart, J., and Jones, E. W. (1987) *Mol. Cell. Biol.* **7**, 4390–4399