Expression of Synthetic Thaumatin Genes in Yeast†

Jar-How Lee,* Joachim L. Weickmann, Raju K. Koduri, Pradip Ghosh-Dastidar, Katsuyuki Saito,‡ Lindley C. Blair, Takayasu Date,§ Jiunu S. Lai, Stanley M. Hollenberg,‖ and Rick L. Kendall‡

INGENE, Inc., 1545 17th Street, Santa Monica, California 90404

Received December 30, 1987; Revised Manuscript Received March 15, 1988

ABSTRACT: Thaumatin is a plant protein that contains 8 disulfides and 207 amino acids in the mature form. The protein is of potential commercial interest since microgram quantities elicit an intense sweetness sensation. Two major variants of thaumatin have been identified in our laboratory by using sequence data obtained from thaumatin tryptic peptides. These differ by one amino acid at position 46 (asparagine or lysine), and both proteins differ from previously published sequences. We have synthesized DNA-coding sequences for three of these thaumatin variants using yeast preferred codons. The genes were inserted into an expression vector that contained a yeast 3-phosphoglycerate kinase promoter and terminator, and the vectors were transformed into yeast for expression of the recombinant protein. Upon lysis of the yeast cells, all thaumatin was localized in the insoluble cell fraction. Analysis of the sodium dodecyl sulfate solubilized yeast extracts by gel electrophoresis and Western blotting showed that thaumatin represented about 20% of the insoluble yeast protein. Although expressed at high levels, none of the thaumatins was biologically active (sweet). Preliminary protein folding experiments showed that two of three thaumatin variants could be folded to the sweet conformation.

The African shrub *Thaumatococcus daniellii* Benth produces an extremely sweet-tasting protein in the arils of its fruit. Traditionally, the fruit has been used in West Africa as a sweetener of palm wine, corn bread, and sour fruit. Thaumatin, the sweet protein isolated from this fruit, is about 3000 times sweeter than sucrose on a weight basis (Van der Wel & Loeve, 1972). The nontoxic, low-caloric, noncariogenic nature of thaumatin makes it an excellent candidate for development as a sugar substitute.

Analysis of the purified sweet proteins shows that thaumatin is actually comprised of a mixture of at least five intensely sweet forms. They are thaumatins T_I, T_{II}, T_a, T_b, and T_c, named for their order of elution from a cation-exchange resin[1] (Van der Wel & Bel, 1976; Higginbotham & Hough, 1977). All have molecular masses of approximately 22 kilodaltons (kDa).[2] The entire amino acid sequence has been determined directly for thaumatin I (Iyengar et al., 1979). However, the amino acid sequence deduced from the cDNA sequence differs from that published for the thaumatin I sequence by five amino acids and is believed to be the sequence of thaumatin II (Edens et al., 1982). Each of these two proteins consists of a single unmodified polypeptide chain, 207 amino acid residues in length. The mature proteins contain eight disulfide bonds; a partial three-dimensional structure (backbone only) has been reported (De Vos et al., 1985). No unusual amino acids or posttranslational modifications have been detected for the purified proteins. Sequence analysis of the cDNA indicates that thaumatin II is initially translated as a precursor form, preprothaumatin, with both a 22 amino acid amino-terminal secretion signal and an acidic, 6 amino acid carboxy-terminal extension (Edens et al., 1982).

The biological function of thaumatin in plants remains undetermined. However, proteins having remarkable homology to the thaumatin sequence are present in other plants. Cornelissen et al. (1986) describe a thaumatin-like protein (with 65% sequence homology to thaumatin) induced in tobacco by tobacco mosaic virus. Richardson et al. (1987) showed that a protease inhibitor in maize exhibits a great degree of sequence homology to both thaumatin (52%) and the tobacco protein (57%). They propose that these proteins, as a class, may have roles in response to wounding or infection.

The use of a yeast in vivo expression system for producing thaumatin and closely related derivatives under fermentation conditions may be of value for the commercial development of this sweet protein. In this report, we describe the identification of two new thaumatin variant protein sequences, the design and synthesis of three different thaumatin genes, and their efficient expression in yeast. Sequence analysis of the major components of plant thaumatin has also allowed us to resolve some uncertainties about the amino acid assignments at various positions. The corresponding coding sequences were designed to contain numerous restriction enzyme sites for convenient gene manipulation, and the codons used in highly expressed yeast genes were incorporated into the gene se-

† This work was supported by a contract from the Beatrice Co. to INGENE, Inc.

* Author to whom correspondence should be addressed.

‡ Present address: Department of Biological Chemistry, California College of Medicine, University of California, Irvine, CA 92717.

§ Present address: Department of Biological Chemistry, Kanazawa Medical University, Uchinada-Cho, Ishikawa-Ken 920-02, Japan.

‖ Present address: Department of Biology, University of California, San Diego, La Jolla, CA 92093.

[1] The thaumatin literature is replete with numerous abbreviations used to identify the different forms of this protein. Thaumatins I, II, and III, T_1 and T_2, T_I and T_{II}, and T_a, T_b, and T_c (T_o) have all been used to designate the different homologous proteins. We define thaumatins with published sequences as I and II and thaumatins separated by SP-Sephadex ion-exchange chromatography and sequenced in this laboratory as A and B.

[2] Abbreviations: SP, sulfopropyl; HPLC, high-performance liquid chromatography; FPLC, fast protein liquid chromatography; SDS, sodium dodecyl sulfate; YT, yeast thaumatin; PAGE, polyacrylamide gel electrophoresis; RIA, radioimmunoassay; PGK, 3-phosphoglycerate kinase; kDa, kilodalton(s); BND-cellulose, benzoylated, naphthoylated DEAE-cellulose; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; EDTA, ethylenediaminetetraacetic acid; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; PTH, phenylthiohydantoin; bp, base pair(s); PMSF, phenylmethanesulfonyl fluoride.

0006-2960/88/0427-5101$01.50/0

quences. Site-directed mutagenesis (Miyada et al., 1982) has allowed us to produce derivatives of this sweet protein (Huang et al., 1987) and will enable us to study the effect of coding sequence changes on protein folding, stability, and taste properties.

Materials and Methods

Materials. All restriction enzymes were purchased from New England Biolabs, Bethesda Research Laboratories, Pharmacia, or Boehringer Mannheim and used in the buffers recommended by the manufacturers. Benzoylated, naphthoylated DEAE-cellulose (BND-cellulose) was supplied by Boehringer Mannheim while TPCK-trypsin and thaumatin were obtained from Sigma Chemical Co. The latter was further purified by sulfopropyl (SP)-Sephadex ion-exchange chromatography as described below.

Purification of Thaumatins A and B from the Plant Thaumatin Mixture. SP-Sephadex, a strongly acidic cation exchanger, was used to fractionate the naturally occurring plant thaumatin variants. The SP resin was equilibrated in 10 mM sodium phosphate, pH 7.2, and 0.1 mM EDTA (buffer A) and packed into a 5 × 22 cm Amicon column. Two hundred milligram quantities of plant thaumatin were dissolved in 20 mL of the same buffer and loaded onto the column. The bound thaumatin was washed with 500 mL of buffer A followed by elution with 1 L of a linear gradient from 0 to 0.25 M NaCl in the same buffer. Four-milliliter fractions were collected and checked for protein (by absorbance at 280 nm) and conductivity.

Desalting and concentration were done by pressure ultrafiltration using an Amicon stirred cell fitted with an Amicon YM-10 membrane. Three successive washes with distilled water removed greater than 95% of the salt. The concentrated protein was lyophilized to dryness and stored at 4 °C where it is stable indefinitely.

Peptide Isolation and Purification. Plant thaumatin, carboxymethylated according to the procedure of Gurd (1972), was purified as a single peak by fast protein liquid chromatography (FPLC) in 20 mM Tris-HCl, pH 8, using a Mono Q column with a 0 to 0.4 M NaCl linear gradient. The purified protein (1 mg) was desalted, lyophilized, and dissolved in 1 mL of 50 mM ammonium bicarbonate, pH 8.2, containing 25–40 μg of TPCK-trypsin. Samples were incubated at 37 °C for 16 h and lyophilized to dryness in 100-μg aliquots.

The peptide elution profile of TPCK-trypsin-digested alkylated thaumatin was analyzed by reverse-phase high-performance liquid chromatography (HPLC) on an Altex Ultrasphere-EDS (PTH) column. Lyophilized samples (100 μg) were dissolved in 0.1% trifluoroacetic acid and loaded into a C_{18} column. Thaumatin peptides were eluted from the column by using a 5–50% acetonitrile linear gradient over a period of 45 min. The elution profile was monitored at 214 nm, and fractions were collected as individual peaks for sequence and composition analysis.

Amino Acid Analysis and Protein Sequencing. Lyophilized samples of the purified carboxymethylated plant thaumatin or thaumatin peptides were analyzed for amino acid composition after acid hydrolysis for 24 h. Samples were run on a Beckman Model 6300 amino acid analyzer. Data were compiled by using a Hewlet Packard Model 3390A recording integrator.

The same protein and peptide samples were sequenced on the Applied Biosystems Model 470 sequencer (Henrick et al., 1981). A Hewlet Packard Model 1090 instrument equipped with a 220 × 2.1 mm C-18 column was used for PTH analysis (Wittmann-Liebold & Ashman, 1985).

Chemical Synthesis of a Gene Encoding Thaumatin I. The strategy for generating a coding sequence synthetically for thaumatin was similar to that developed by Ito et al. (1982). This method takes advantage of the action of DNA polymerase to produce portions of the complement to the synthesized strand and is illustrated and discussed in Figure 1. Yeast-preferred codons were used in the makeup of the thaumatin I gene so that it could be expressed efficiently in yeast (Bennetzen & Hall, 1982). An ATG initiation codon was included at the 5′ end, and two consecutive termination codons were placed at the 3′ end of the gene. In order to facilitate the cloning process, the synthetic gene was also designed to contain numerous unique restriction enzyme sites. The total length of the thaumatin I coding sequence is 630 bp.

Genetic Manipulations. Genetic manipulations were performed as described in Maniatis et al. (1982) and Sherman et al. (1979). *Escherichia coli* strains HB101 (Bolivar, 1977) and MC1061 (Casadaban & Cohen, 1980) were used in routine transformations and plasmid preparations.

Derivatives of the *E. coli*–yeast shuttle vector pJDB209 (Beggs, 1981; see Figure 4) were introduced into yeast strain AH22 (*MATa leu2 his4 gal4*; Hinnen et al., 1978) by the lithium acetate procedure (Ito et al., 1983) with one minor modification. The AH22 cells were plated onto SD-leu plates in 0.1 mL of YPD containing 100 μg/mL leucine to allow growth while the plasmid was being established.

Preparation of Yeast Cell Extracts. Yeast cells grown to stationary phase in 200 mL of minimal medium were harvested by centrifugation, and 0.3 g wet weight of cells was transferred into a 1.5-mL Eppendorf tube. In order to distinguish between thaumatin expression in the soluble or insoluble cell fractions, yeast cells were resuspended and broken with glass beads in 50 mM Tris-HCl, pH 7.5, containing 2 mM PMSF and 0.1 mM EDTA. After removal of the glass beads, the suspension of lysed cells was centrifuged and the supernatant transferred to a new tube. The pellet in turn was resuspended in 50 mM Tris-HCl, pH 7.5, containing 1% sodium dodecyl sulfate (SDS). The pellet proteins were solubilized by boiling in a water bath for 10 min and separated from the insoluble fraction by centrifugation. The supernatant and solubilized material from the cell pellet were analyzed for released proteins by SDS gel electrophoresis.

Gel Electrophoresis and Protein Determination. Samples for SDS–polyacrylamide gel electrophoresis (PAGE) were prepared by the method of Laemmli (1970) and run on a discontinuous 15% acrylamide–0.086% bis(acrylamide) gel (Blattler et al., 1972) at constant voltage. An acetate–urea gel system (Reisfeld et al., 1962) was used to separate the basic thaumatins on a discontinuous 15% acrylamide–0.2% bis(acrylamide) gel at constant voltage. Protein concentration was measured by the Bio-Rad colorimetric assay or by the absorbance at 280 nm (the absorbance of a 1 mg/mL solution of thaumatin is 1.00 ± 0.05) or at 214 nm for peptides.

Immunology. Antibodies against purified plant thaumatin A were induced in young adult male New Zealand rabbits by multiple subcutaneous injections with 100 μg of thaumatin A in Freund's complete adjuvant. Booster injections contained 20–40 μg of thaumatin. Sera were initially screened for antibodies cross-reactive with thaumatin by the double immunodiffusion assay in agar (Ouchterlony, 1962) at pH 7.5, 25 °C. Due to the high antibody titer, diluted serum was used directly without removing other serum proteins.

Plant thaumatin A was iodinated with carrier-free ^{125}I as described by Gospodarowicz (1973), except that a 5–10-fold higher level of hydrogen peroxide was required for the iodi-

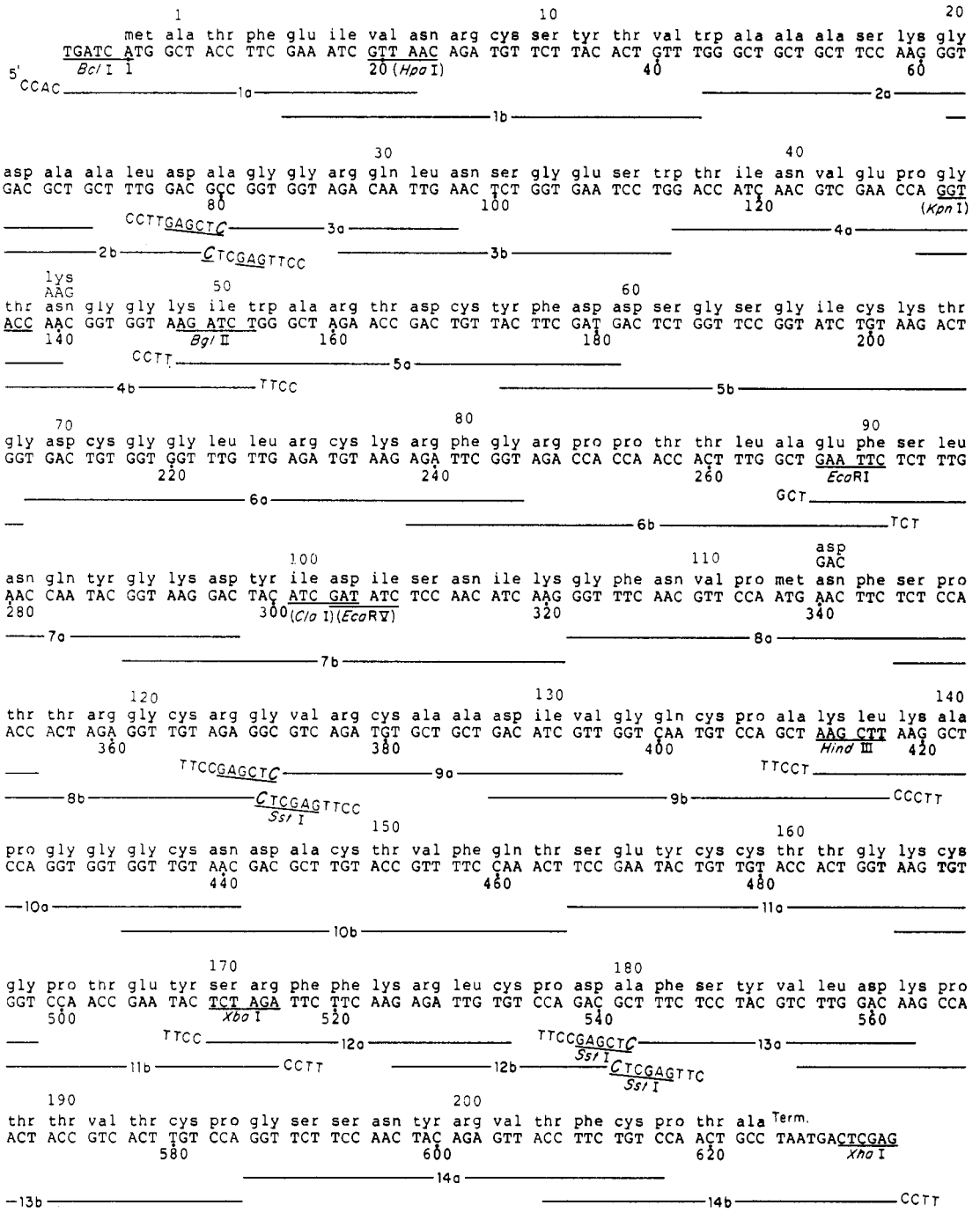

FIGURE 1: DNA sequence and synthesis of the thaumatin genes. The thaumatin gene was assembled from 28 chemically synthesized oligonucleotides ranging in size from 25 to 39 nucleotides. These oligonucleotides are indicated under the gene sequence (1a,b–14a,b). Oligonucleotides were synthesized by the phosphodiester method (Ito et al., 1982). These oligonucleotides formed pairs consisting of a sense and an anti-sense strand. Each had a 9- or 10-nucleotide complementary region that allowed annealing to take place. Enzymatic extension of fragment pairs by the Klenow fragment of DNA polymerase I was used to produce blunt-ended duplexes. Each fragment contained one or two restriction enzyme sites that were used in the sequential insertion of the DNA into the appropriate plasmid vectors. After restriction enzyme digestion of the double-stranded DNA duplexes, one or two duplexes were then inserted into a plasmid vector to generate a cloned gene segment. The sequence of the cloned gene segments was confirmed by DNA sequencing (Sanger et al., 1977). The complete thaumatin gene was pieced together by the sequential joining of contiguous gene segments. An *Sst*I site was introduced in three places to facilitate the cloning of long DNA fragments. When each cloning was completed, the *Sst*I sites were eliminated by "chewing back" with the Klenow fragment of DNA polymerase I and religating. The restriction enzyme sites in parentheses were not used in gene synthesis. Codons above the primary sequence of thaumatin are the change(s) in the thaumatin A and/or thaumatin B sequences.

nation. Reaction conditions were adjusted so that only two to three of the eight tyrosines in the primary sequence were iodinated.

Radioimmunoassay procedures were based on published methods (Weickmann & Glitz, 1982) with several minor modifications. The first and second antibody incubation reactions were reduced from several hours to 60 and 30 min, respectively. The standard curve generated by using a known quantity of diluted plant thaumatin was linear in the range of 20–150 ng of protein per assay. The assay, however, could detect levels of competing protein as low as 1–5 ng. In the absence of inhibitors, 60–70% of the total ^{125}I (ca. 30000 cpm) was precipitated; background precipitation by nonimmune antiserum alone was about 2–5% of the total ^{125}I.

Immunoblot Detection of Thaumatin. The Western Blot procedure used was essentially that described by Towbin et

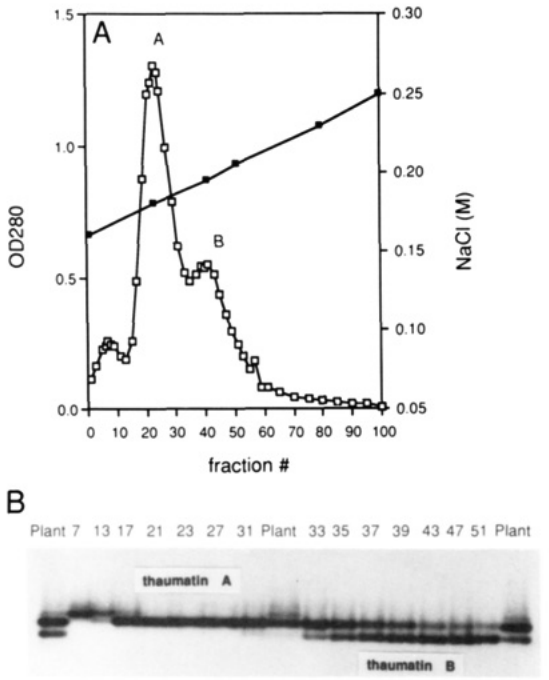

FIGURE 2: Sulfopropyl anion-exchange chromatography of commercially available plant thaumatin. (Panel A) Elution profile at 280 nm (□) showing separation of a plant thaumatin mixture into thaumatins A and B. The salt gradient used to elute the proteins is indicated (■). Panel B shows acid–urea gel analysis of fractions from the SP-Sephadex-separated plant thaumatins. Fractions 21–31 are pure thaumatin A while fractions 39–51, although enriched for thaumatin B, are still contaminated with thaumatin A. A third naturally occurring form of thaumatin can be seen in fraction 7.

al. (1979) using a 2000-fold dilution of rabbit antiserum to plant thaumatin. The protein-bound specific antibodies were labeled with ^{125}I protein A and visualized by autoradiography.

RESULTS

Determination of Thaumatin Protein Sequence. Since there were some disparities between the primary sequences determined for thaumatin by protein sequencing (Iyengar et al., 1979) and the published cDNA sequence (Edens et al., 1982), we decided to reanalyze the sequence of the purified plant proteins. When plant thaumatin was subjected to SP ion-exchange chromatography, two major species were identified and separated (Figure 2A). A convenient way to check for completion of this separation and the homogeneity of the final product is by acid urea gel electrophoresis (Figure 2B). Only the purified proteins were used for sequence determination. Both proteins were carboxymethylated, purified by FPLC chromatography on a Mono-Q column, and digested with TPCK-trypsin. The various fragments were separated by reverse-phase HPLC chromatography on a C_{18} column. Comparison of the HPLC profiles showed an almost identical elution pattern with only slight differences between these two thaumatin variants (Figure 3). The peptides were isolated and purified, and the amino acid composition and sequence of each were determined. In some cases, two peptides copurified and were sequenced simultaneously. Knowledge of the primary sequence of thaumatin from previous studies (Iyengar et al., 1979; Edens et al., 1982) permitted placement (by similarity) of peptides in the sequence. Only one amino acid difference between these two plant forms (asparagine in place of lysine) at position 46 was found when these sequences were compared. However, both sequences differed from the published sequences for thaumatins I and II (see Table I). To avoid confusing these new sequences with those previously published, we designated our sequences thaumatins A and B. The thaumatin A sequence differs from the two published sequences by one and four amino acids and the thaumatin B sequence differs by two and three amino acids, respectively.

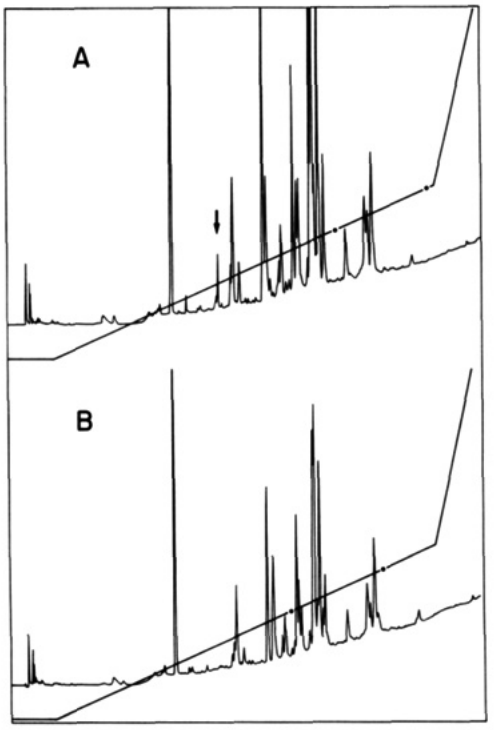

FIGURE 3: HPLC fractionation of carboxymethylated tryptic peptides of thaumatin from (A) pING406 and (B) pING407. The HPLC-purified peptides were collected and sequenced as described under Materials and Methods. Comparison of the two sequences enabled us to detect differences in the amino acid sequences of the two thaumatins.

Table I: Comparison of Amino Acid Sequence Differences between Four Thaumatin Derivatives

	amino acid position				
thaumatin	46	63	67	76	113
A/YT407	Asn	Ser	Lys	Arg	Asp
B/YT406	Lys	Ser	Lys	Arg	Asp
I/YT59	Asn	Ser	Lys	Arg	Asn
II	Lys	Arg	Arg	Gln	Asp

Synthesis of Three Thaumatin Variant Genes. A synthetic gene based on the amino acid sequence of thaumatin I (Iyengar et al., 1979) was constructed as described under Materials and Methods and outlined in Figure 1. Yeast-preferred codons (Bennetzen & Hall, 1982) were used in the thaumatin I gene so that it could be expressed efficiently in yeast. An ATG initiation codon at the 5′ end (start of the protein coding region) and two consecutive termination codons at the 3′ end of the gene were included. The gene was designed to contain multiple restriction enzyme sites to facilitate the cloning process. The completed thaumatin I gene consists of 630 bp bracketed by a *BclI* site at the 5′ end and an *XhoI* site at the 3′ end.

Site-directed mutagenesis (Miyada et al., 1982) and DNA fragment replacement in the thaumatin I gene were used to obtain the thaumatin A and B genes. The thaumatin A protein sequence differs from the thaumatin I sequence only at residue 113 and involves an asparagine to aspartic acid change. A synthetic oligonucleotide (21-mer) was used as a mutagenesis primer to change the codon AAC (asparagine) at position 113 in the thaumatin I gene to the codon GAC (aspartic acid) in the thaumatin A gene.

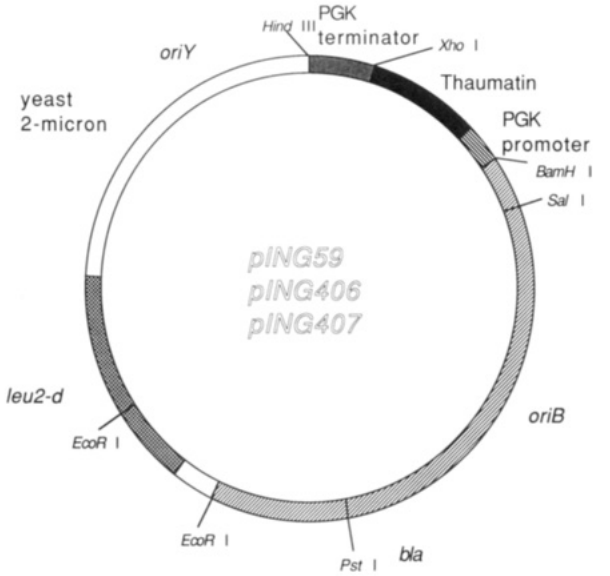

FIGURE 4: Plasmid maps of pING59, pING406, and pING407. The thaumatin expression cassette was inserted between the *Bam*HI and *Hin*dIII site of pJDB209 (Beggs, 1981), an *E. coli*–yeast shuttle vector. The *Fok*I site, which is cleaved after the A of the initiator ATG of the yeast PGK promoter (Hitzeman et al., 1982), was joined to the T4 DNA polymerase filled-in *Bcl*I site of the thaumatin gene. The *Bgl*II site near the 3′ end of the PGK gene was converted to an *Xho*I site attaching an *Xho*I linker to the T4 DNA polymerase filled-in *Bgl*II site. The *Xho*I site was used to join the 3′ end of the thaumatin gene. A restriction map of thaumatin is shown in Figure 1. The differences among pING59, pING406, and pING407 are in the coding sequences shown for thaumatins I, A, and B, respectively.

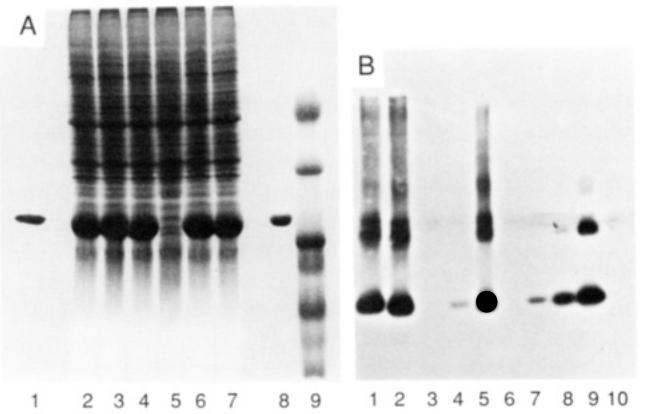

FIGURE 5: (A) SDS gel electrophoresis analysis of different yeast cell extracts showing the quantity of thaumatin expression for several derivatives. These were overloaded for Coomassie Blue staining and Western blot analysis. No thaumatin is detected in the soluble fraction of the cell extract. Lanes 1 and 8, plant thaumatin; lane 2. pING59; lane 3, pING406; lane 4, pING407; lane 5, pING58; lanes 6 and 7, pING406, yeast colonies a and b; lane 9, Bio-Rad standards. (B) Western blot of yeast extract proteins and plant thaumatin standards fractionated by SDS–PAGE. Lane designations are as follows: (1) pING406 and (2) pING407, SDS-solubilized proteins; (3) pING407 and (4) pING406, soluble proteins; (5) pING59, SDS-solubilized proteins; (6) pING59 soluble cell proteins; (7–9) plant thaumatin, 20, 100, and 1000 ng, respectively; (10) pING58, SDS-solubilized proteins.

The thaumatin A gene so generated was used to make a coding sequence for the thaumatin B gene. The amino acid sequence of thaumatin B differs from that of thaumatin A only at position 46. Conversion of the codon at this position was facilitated by the presence of two flanking unique restriction enzyme sites, *Kpn*I and *Bgl*II. Two complementary oligonucleotides, a 12-mer and a 20-mer, were synthesized to form a short double-stranded DNA fragment that contains the codon AAG for lysine at position 46. This replacement oligonucleotide fragment contains *Kpn*I and *Bgl*II cohesive ends which allow the simple substitution of the corresponding *Kpn*I–*Bgl*II fragment containing the codon AAC for asparagine at position 46. Colonies containing the plasmid with the correct DNA insert were identified by colony hybridization (Grunstein & Hogness, 1975) using the shorter synthetic oligonucleotide (12-mer) inserted above as a probe.

Construction of Vectors for Thaumatin Expression. The yeast 3-phosphoglycerate kinase (PGK) promoter, a very strong promoter, was used to direct the expression of the three thaumatin variant genes (I, A, and B) in yeast. The yeast PGK terminator, which provides efficient transcription termination and polyadenylation signals, was used as a terminator for thaumatin gene transcription. A thaumatin expression cassette containing the PGK promoter–thaumatin gene–PGK terminator was cloned into a high copy number *E. coli*–yeast shuttle vector, pJDB209 (Beggs, 1981). These thaumatin expression vectors, pING59, pING407, and pING406 (Figure 4), contain the synthetic genes for the thaumatin variants I, A, and B, respectively. The protein products of these genes are designated YT (yeast thaumatin) 59 (YT59), YT407, and YT406.

Thaumatin Expression in Yeast Cells. Plasmids pING59, pING406, and pING407, together with pING58 (which contains the PGK terminator but no PGK promoter and thaumatin coding sequence), were transformed into yeast strain AH22. The Leu$^+$ yeast transformants were selected on SD-leu minimal medium plates. Leu$^+$ colonies were grown up in shaker flasks and the yeast cells harvested by centrifugation. The cells were lysed, and the extract was checked for protein. Yeast thaumatin was initially detected in cell extracts by the presence of a 22-kDa protein visualized on an SDS–PAGE gel following Coomassie Brilliant Blue R staining. The presence of thaumatin in the yeast extract was confirmed by protein immunoblotting (Western) using thaumatin-specific antibodies.

Thaumatin in yeast transformed with pING59, pING407, and pING406 is found in the insoluble fraction when the yeast extract is centrifuged at greater than 2500*g*. Extraction of the lysed cell pellet with 1% SDS at 100 °C released thaumatin into the soluble phase and allowed visualization of the protein by Coomassie Brilliant Blue R staining after SDS gel electrophoresis. Nonionic detergents, Triton X-100 and NP-40, did not extract significant quantities of thaumatin from the lysed cell pellet (unpublished results). No such protein of similar size was found in the protein extract of the pING58 control culture. Figure 5A shows a comparison of the quantities of thaumatin (and other yeast proteins) solubilized by SDS extraction of the soluble and insoluble fractions of yeast extract for thaumatins YT59, YT406, and YT407. The levels of thaumatin expression appear to be similar for all three derivatives.

The identity of yeast thaumatin was confirmed by Western blot analysis of SDS-solubilized proteins (Figure 5B). A 2000-fold dilution of rabbit antiserum to thaumatin A was used to probe SDS–PAGE-separated proteins that had been transferred to nitrocellulose paper. At this low antiserum dilution, the diversity of low-titer antibodies becomes evident in that denatured, fragmented, and aggregated thaumatins are recognized in the Western blot (Figure 5B; lane 9 even shows antibody recognition of an aggregated form of plant thaumatin that is not dissociated by SDS).

The major protein recognized by the antiserum was localized in the insoluble portion of the yeast extract; it comigrated with the plant thaumatin standard. Proteolytic degradation prod-

ucts of yeast thaumatin were not observed. No cross-reactive protein was observed in the soluble protein fraction nor in the insoluble fraction of the control culture. Since reduction of the thaumatin disulfides results in protein precipitation, the reducing intracellular mileiu of yeast may provide one reason why soluble thaumatin is not found inside these cells.

The amount of intracellular thaumatin produced by yeast did not vary greatly for the three derivatives tested. Yeast thaumatin is easily solubilized by 1% SDS or in 6 M urea but not by washing with high salt (2 M NaCl). In every case, all the immunoreactive protein was found in the insoluble fraction of yeast extracts. Upon subsequent solubilization, electrophoresis showed that each thaumatin represents about 20% of the total insoluble yeast protein as judged by LKB Ultroscan XL laser scanning of the Coomassie Blue stained protein bands (data not shown).

Assays for Sweetness. The sweet-taste threshold of native plant thaumatin is 2 μg/mL (determined experimentally in our laboratory). Larger amounts (25 μg/mL) elicit an intensely sweet sensation. Whenever a thaumatin sample tasted sweet in these tests, the sample was immunoreactive in the radioimmunoassay (RIA), albeit at protein concentrations 1000-fold higher than in the RIA. Since our immunoassay is effective in the nanogram range, it is the most sensitive and quantitative assay developed to test for biological activity in the yeast extract samples. It was determined that at high dilution (1:30000 to 1:70000), low-titer antibodies were lost to dilution effects, and the major antibody in the rabbit serum recognized the folded, native structure of plant thaumatin exclusively. When a yeast extract enriched for thaumatin was assayed for cross-reactive protein by using the thaumatin-specific RIA, no protein recognition was observed (data not shown). Since these high-titer antibodies are directed against biologically active thaumatin, the results suggest that intracellularly produced yeast thaumatin is not biologically active (sweet). The yeast are therefore able to produce considerable intracellular quantities of each thaumatin derivative but lack the ability to process these molecules into functional proteins. When the inactive thaumatin derivatives were purified and subjected to in vitro protein folding (unpublished experiments), YT407 (thaumatin A) and YT406 (thaumatin B) could be folded into an immunoreactive conformation. These proteins were also determined to be sweet by a taste test. We were unable to fold YT59 (thaumatin I) into a biologically active conformation. This protein was not cross-reactive in the RIA at any level and was not perceived to be sweet during subsequent taste tests.

Discussion

Plant thaumatin is comprised of a mixture of related proteins that are easily separated on the basis of their charge differences. The relative ratios of the major species can be quantitated by SP-Sephadex ion-exchange chromatography and acid–urea PAGE (Figure 2). These techniques show that normal fluctuations can and do occur in the ratios of the different forms of thaumatin. Our observations have shown that thaumatin A usually represents >70% of the total thaumatin in commercial samples. However, this amount is variable, and some preparations contain >95% thaumatin A. The relative amounts of these proteins found in the plant may depend upon genetic, climatic, or seasonal variations, or harvest times.

We have sequenced the two major variants of thaumatin present in commercially available samples and found that they differ by one amino acid at position 46 (asparagine versus lysine). The single amino acid difference suggests that thaumatins A and B are derived from two distinct genes. The sequences we determined for thaumatins A and B differ from previously published sequences (Iyengar et al., 1979; Edens et al., 1982) for thaumatins I and II (see Table I). Our sequence data have placed an aspartic acid residue in place of asparagine at position 113 in both thaumatin A and thaumatin B. This assignment agrees with the published cDNA sequence for thaumatin II but contrasts with the published primary sequence data for thaumatin I (Iyengar et al., 1979), which showed that asparagine occupies this position. Aspartic acid at position 113 is important for biological activity. A protein folding protocol developed in this laboratory (unpublished experiments) showed that thaumatin I (Asn-113) was not folded to a sweet form, whereas thaumatins A and B (both Asp-113) were. Other protein renaturation methods (Bradshaw et al., 1967; Light, 1985) were also used, and none was successful in generating even low levels (nanogram quantities) of immunoreactive thaumatin. We therefore suggest that the amino acid assignment at position 113 for thaumatin I is incorrect. Since thaumatin A is sweet, it is unlikely that it represents a natural inactive (nonimmunoreactive) variant.

In order to produce thaumatin in microorganisms, recombinant DNA technology was used to introduce a synthetic functional gene into yeast. The gene was comprised of a yeast PGK promoter and PGK terminator joined to the 5′ and 3′ ends of the thaumatin coding sequence, respectively. This construction ensured efficient transcription and termination of the thaumatin gene and resulted in high levels of protein expression.

The thaumatin gene was chemically synthesized for two reasons. First, multiple restriction sites could be engineered into the thaumatin gene to facilitate gene manipulation. We have taken advantage of these sites to generate alternate genes (namely, thaumatins A and B) from the thaumatin I coding sequence. Second, the gene can be tailored to contain preferred codons for a specific host organism (in this case, yeast). Previously, a codon bias has been shown for highly expressed yeast genes (Bennetzen & Hall, 1982). Another study (Edens et al., 1984) reported the expression of thaumatin cDNA using the yeast glyceraldehyde-3-phosphate dehydrogenase promoter on a similar high copy number vector in yeast. The results showed a very low level of thaumatin expression. Since both the yeast GAPDH and PGK promoters direct high protein expression, the high level of thaumatin expression (ca. 20%) reported herein may result in part from the use of yeast-preferred codons in our synthetic gene.

Polyacrylamide gel analysis of yeast cell extracts containing the thaumatin gene showed that significant quantities of thaumatin are produced and, as expected, remain inside the cell (Figure 5). This result is not surprising since the natural plant signal sequence was not attached to the N-terminus of the protein and the cell therefore lacks the information necessary to destine the protein for export or membrane localization by the signal sequence directed secretory process. Thaumatin is found only in the insoluble cell fraction, but whether it accumulates in the cytoplasm or in an organelle compartment was not determined.

When plant thaumatin is reduced in vitro in the absence of detergents or denaturants, it becomes insoluble. The insoluble nature of yeast thaumatin suggests that the internal reducing environment of these cells apparently prevents disulfide formation. This results in the production of a protein molecule having the amino acid sequence of thaumatin but lacking its native conformation.

One clear difference between the plant and yeast-produced thaumatins is that the latter has a blocked amino-terminal alanine (Huang et al., 1987). Folding studies, however, have clearly shown that the purified, N-terminal modified yeast thaumatin is indistinguishable from plant thaumatin in biological activity and sweetness threshold. The N-terminal modification apparently has little effect on the folding process or on the interaction of the protein with the sweet receptor. We have utilized our ability to mutagenize specific codons to study the effect of all 20 amino acid replacements at the N-terminus of thaumatin on the processing of this protein by yeast (Huang et al., 1987). Mutagenesis of the thaumatin gene is currently being used to study the effect of amino acid substitutions on thaumatin structure, folding, stability, and taste characteristics.

ACKNOWLEDGMENTS

We thank Brett Einarson for technical assistance on protein sequence analysis, Michael Grunstein for helpful discussions, and Kevin Bryan for help in preparing figures. We also thank Ralph Bradshaw for continued advice and critical reading of the manuscript.

REFERENCES

Beggs, J. D. (1981) *Alfred Benzon Symp. 16*, 383–390.
Bennetzen, J. L., & Hall, B. D. (1982) *J. Biol. Chem. 257*, 3026–3031.
Blattler, D., Garner, E., Van Slyke, K., & Bradley, A. (1972) *J. Chromatogr. 64*, 147–155.
Bolivar, F., Rodriguez, R. L., Bethlach, M. C., & Boyer, H. W. (1977) *Gene 2*, 75–93.
Bradshaw, R. A., Kanarek, L., & Hill, R. L. (1967) *J. Biol. Chem. 242*, 3789–3798.
Casadaban, M. J., & Cohen, S. N. (1980) *J. Mol. Biol. 138*, 179–207.
Cornelissen, B. J. C., Hooft van Huijsduijnen, R. A. M., & Bol, J. F. (1986) *Nature (London) 321*, 531–532.
De Vos, A., Hatada, M., Van der Wel, H., Krabbendam, H., Peerdeman, A., & Kim, S.-H. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 1406–1409.
Edens, L., Bom, I., Ledeboer, A. M., Maat, J., Toonen, M. Y., Visser, C., & Verrips, C. T. (1984) *Cell (Cambridge, Mass.) 37*, 629–633.
Edens, L., Heslinga, L., Klok, R., Ledeboer, A. M., Maat, J., Toonen, M. Y., Visser, C., & Verrips, C. T. (1982) *Gene 18*, 1–12.
Gospodarowicz, D. (1973) *J. Biol. Chem. 248*, 5042–5049.
Grunstein, M., & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 3961–3965.
Gurd, F. (1972) *Methods Enzymol. 25*, 424–438.
Henrick, R. M., Hunkapillar, M. W., Hood, L. E., & Dreyer, W. J. (1981) *J. Biol. Chem. 256*, 7990–7997.
Higginbotham, J. D., & Hough, C. A. M. (1977) in *Sensory Properties of Foods* (Birch, G. C., Brennan, J. G., & Parker, K. J., Eds.) pp 129–149, Applied Sciences, London.
Hinnen, A., Hicks, J. B., & Fink, G. R. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 1929–1933.
Hitzeman, R. A., Hagie, F. E., Hayflick, J. S., Chen, C. Y., Seeburg, P. H., & Derynck, R. (1982) *Nucleic Acids Res. 10*, 7791–7808.
Huang, S., Elliott, R. C., Liu, P.-S., Koduri, R. K., Weickmann, J. L., Lee, J.-H., Blair, L. C., Ghosh-Dastidar, P., Bradshaw, R. A., Bryan, K. M., Einarson, B., Kendall, R. L., Kolacz, K. H., & Saito, K. (1987) *Biochemistry 26*, 8242–8246.
Ito, H., Fukuda, Y., Murata, K., & Kimura, A. (1983) *J. Bacteriol. 153*, 163–168.
Ito, H., Ike, Y., Ikuta, S., & Itakura, K. (1982) *Nucleic Acids Res. 10*, 1755–1769.
Iyengar, R. B., Smits, P. S., van der Ouderaa, F., van der Wel, H., van Brouwershaven, J., Ravestein, P., Richters, G., & van Wasenaar, P. D. (1979) *Eur. J. Biochem. 96*, 193–204.
Laemmli, V. K. (1970) *Nature (London) 227*, 680–685.
Light, A. (1985) *BioTechniques 3*, 298–306.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Miyada, C. G., Soberon, X., Itakura, K., & Wilcox, G. (1982) *Gene 17*, 167–177.
Ouchterlony, O. (1962) *Prog. Allergy 6*, 30–37.
Reisfeld, R. A., Lewis, U. J., & Williams, D. E. (1962) *Nature (London) 195*, 281–283.
Richardson, M., Valdes-Rodriquez, S., & Blanco-Labra, A. (1987) *Nature (London) 327*, 432–434.
Sanger, F., Nickleu, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Sherman, F., Fink G. R., & Hicks, J. B. (1979) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Towbin, H., Staehlin, T., & Gordon, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4350–4354.
Van der Wel, H., & Bel, W. J. (1976) *Chem. Senses Flavour 2*, 211–218.
Van der Wel, H., & Loeve, K. (1972) *Eur. J. Biochem. 31*, 221–225.
Weickmann, J. L., & Glitz, D. G. (1982) *J. Biol. Chem. 257*, 8705–8710.
Wittmann-Liebold, B., & Ashman, K. (1985) in *Modern Methods in Protein Chemistry* (Tschesche, H., Ed.) pp 303–327, Walter de Gruyter, Berlin.