BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **229,** 121–127 (1996)
ARTICLE NO. 1767

Expression of a Synthetic Gene Encoding the Sweet-Tasting Protein Thaumatin in *Escherichia coli*

Ignacio Faus,*[,1] Cristina Patiño,† José Luis del Río,* Catalina del Moral,† Heidi Sisniega Barroso,* and Victor Rubio†

**Laboratory of Biotechnology, URQUIMA, S.A. (URIACH GROUP), Arnau de Vilanova, 22-42, 08105 Sant Fost de Campsentelles (Barcelona), Spain; and †Centro Nacional de Biotecnología (CSIC-UAM), Campus Canto Blanco, 28049 Madrid, Spain*

Received October 22, 1996

A synthetic gene encoding the aminoacid sequence of the sweet-tasting protein thaumatin II has been assembled and expressed in *Escherichia coli.* Immunoblotting analysis shows that the expressed recombinant thaumatin has the same molecular weight as the protein from its natural source, the plant *Thaumatoccocus daniellii Benth.* © 1996 Academic Press, Inc.

The thaumatins are proteins with a very sweet taste and the ability to increase the palatability (upgrading or improving other flavours) of food; in industry they are currently extracted from the arils of the fruit of the plant *Thaumatoccocus daniellii Benth* (for a complete review, see Ref.1). Thaumatins can be isolated from these arils in at least five different forms, which can be separated using ion-exchange chromatography. These forms are all single-chain polypeptides with 207 aminoacids and a molecular weight of approximately 22 Kd. Thaumatins I and II, which predominate in the arils, are much sweeter than saccharose (100,000 times sweeter according to one estimate, Ref.1). Besides being natural products, thaumatins I and II are non-toxic, making them a good substitute for common sweeteners in the animal and human food industries.

Despite its advantages, industrial use of thaumatins of plant origin is very limited because of the extreme difficulty involved in obtaining the fruit from which it is extracted. The producing plant, *T. daniellii,* not only requires a tropical climate and pollination by insects, but it must also be cultivated among other trees and yet 75% of its flowers do not bear fruit.

Although attempts have been made to produce thaumatins by genetic engineering in bacteria such as *Escherichia coli* (2), *Bacillus subtilis* (3) and *Streptomyces lividans* (4), in yeasts such as *Saccharomyces cerevisiae* (5) and *Kluveromyces lactis* (1), in the fungus *Aspergillus oryzae* (6), and in transgenic plants such as *Solanum tuberosum* (7), until now the results have been considered disheartening; thus the thaumatin available to industry is very scarce and expensive (1).

As a first step towards the goal of expressing thaumatin in recombinant filamentous fungi, we have assembled, starting with single-stranded oligonucleotides, a synthetic gene encoding thaumatin II. This artificial gene is optimized for its expression in filamentous fungi. We also show that in *Escherichia coli* this synthetic gene directs the synthesis of a protein that is immunologically reactive with an an anti-thaumatin polyclonal antibody. In a second publication we will describe the expression of the synthetic gene encoding thaumatin II in several species of filamentous fungi.

[1] To whom correspondence should be addressed.

0006-291X/96 $18.00

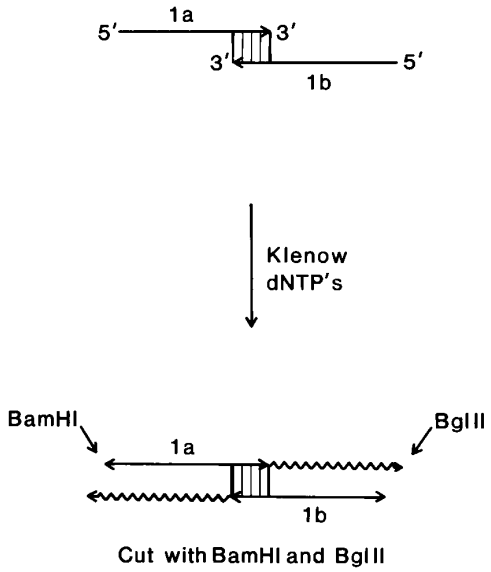

FIG. 1. Strategy used to build FRAGMENT ONE of the synthetic gene encoding thaumatin II. Oligonucloetides *1a* and *1b* have complementary sequences, which are noted. After pairing, and upon addition of dNTPs and the Klenow fragment of DNA polymerase I, elongation takes place. The double-stranded DNA fragment that is generated is then digested with BamHI and BglII prior to its insertion into the pTZ18RN vector. A similar strategy was used to generate fragments 2, 3 and 4.

MATERIALS AND METHODS

Construction of the Synthetic, Artificial and Completely Optimized Gene Encoding Thaumatin II

Strategy for building the synthetic gene encoding thaumatin II. The method chosen for assembling the synthetic gene of thaumatin II is shown in Figure 1. Eight long DNA oligonucleotides were supplied by Isogen Bioscience, Inc. (Netherlands). The eight oligonucleotides are designed in pairs, and they can be hybridized because of the complementary nature of their sequences. They were named *1a, 1b; 2a, 2b; 3a, 3b;* and *4a, 4b.* After pairing, the single-stranded areas were filled with the Klenow fragment of DNA polymerase (Taq polymerase was also used). This resulted in 4 double-stranded DNA fragments named *1, 2, 3* and *4.* These fragments were disgested with the appropriate restriction enzymes to obtain cohesive or blunt ends, and then ligated to the desired vector (Fig.2). The synthetic gene was assembled in 2 fragments which were then joined to the appropriate expression vector.

Assembly of the first 332 pairs of bases of the synthetic gene encoding thaumatin II. Plasmid pTZ18R (Phamacia, Inc.) was subjected to site-directed mutagenesis in order to generate a unique NcoI site in between the BamHI and PstI sites of its polylinker (see Fig.2). In a first step towards assembling a complete gene, oligonucleotides 1a, 1b, 2a and 2b were elongated, digested with the appropriate restriction enzyme, and joined to the pTZ18RN plasmid, thus generating pTZ18RN(1/2).

One μg of oligonucleotide *1a* and 1 μg of *1b* were mixed in a buffer solution containing 40 mM Tris-HCl, pH 8.0, 10mM $MgCl_2$, 5mM DTT, 50 mM NaCl and 50 μg/ml of bovine serum albumin (BSA). The mixture (17 μl) was heated for 5 minutes at 70°C and then cooled slowly to 65°C for about ten minutes (appropriate temperature for hybridization of the oligonucleotide pairs). Then, 2 μl of a mixture of the four deoxynucleotides (dGATC) (2.5 mM of each dNTP) and 1 μl of modified T7 DNA polymerase enzyme (Sequenase brand from U.S. Biochemical Corp.) were added, giving a final volume of 20 μl. The reactions were allowed to proceed for 30 minutes at 37°C, followed by 10 additional minutes at 70°C (to inactivate the Sequenase). The reaction products were digested with Bam HI and Bgl II at 37°C for 3 hours. The following extractions were performed on the DNAs: once with phenol, once with phenol:chloroform and once with chloroform; they were then precipitated with ethanol, resuspended in TE buffer, and frozen at −20°C until later use. This reaction resulted in the synthesis of **fragment one** (see Figs. 1 and 2).

Oligonucleotides *2a* and *2b* were processed in the same way except that the final products were digested with Bgl II and Nco I, thus generating **fragment two** (see Fig.2).

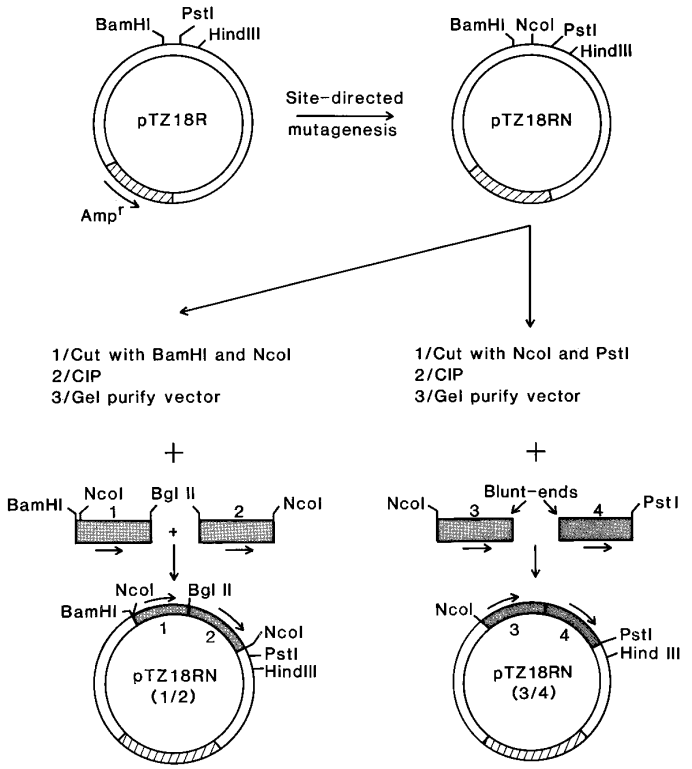

FIG. 2. Diagram of the different steps involved in the asembly of the syntheic gene encoding thaumatin II. Four double-stranded DNA fragments were prepared, and inserted into vector pTZ18RN as shown. The gray areas indicate the sequence of the thaumatin-encoding synthetic gene.

Plasmid pTZ18RN was digested sequentially with BamHI and Nco I and dephosphorylated with calf intestinal phosphatase (CIP). The linearized fragment of 2871 pairs of bases was recovered from a 0.8% agarose gel and purified (see Fig.2).

The products of reactions 1 and 2 (**fragments one and two**) were then joined with the linearized pTZ18RN and the mixture was used to transform *E. coli* strain NM522. To identify the clones with the insert, a white/blue indicator test was used, on dishes with LB/ampicillin medium supplemented with X-gal and IPTG. Various colonies with inserts of the appropriate size contained complete fragments of the 332 pairs of bases of the synthetic gene of thaumatin II. The resulting plasmid was called pTZ18RN(1/2).

Assembly of the second 305 pairs of bases of the synthetic gene encoding thaumatin II. In this case, an alternative approach was used using Taq DNA polymerase and the PCR technique. Before the annealing stage, oligonucleotides *3b* and *4a* were labelled at the 5′ ends with a phosphate group using standard techniques; The oligonucleotides were called *3b** and *4a**.

One ug of *3a* and 1 μg of *3b** were incubated in a reaction mix (18 μl) containing 10 mM Tri-HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$ and 0.1 mg/ml of gelatin. The samples were incubated for 5 minutes at 70°C and for five more minutes at 65°C. At this point, each dNTP was added (G, A, T, C) at a final concentration of 2 mM and 2.5 units of Ampli Taq DNA polymerase (Perkin-Elmer Cetus). PCR cycling was conducted as follows : 1 minute at 94°C; 1 minute at 55°C; and 1 minute at 72°C for 30 cycles, followed by a final extension at 72°C for five minutes. Samples were then extracted with phenol:chloroform, resuspended in 10 μl of restriction enzyme buffer and incubated with Nco I at 37°C for 3 hours. After extraction and precipitation with ethanol, DNAs were dissolved in TE buffer and frozen at −20°C until later use (**fragment three**, Fig.2).

Oligonucleotides *4a** and *4b* were processed as described above, except that the final products were digested with Pst I (**fragment four**, Fig.2).

Ligation of the three fragments was done as per the same procedure mentioned above, except that pTZ18RN was digested with Nco I and Pst I, treated with calf intestinal phosphatase (CIP) and finally purified from an agarose gel

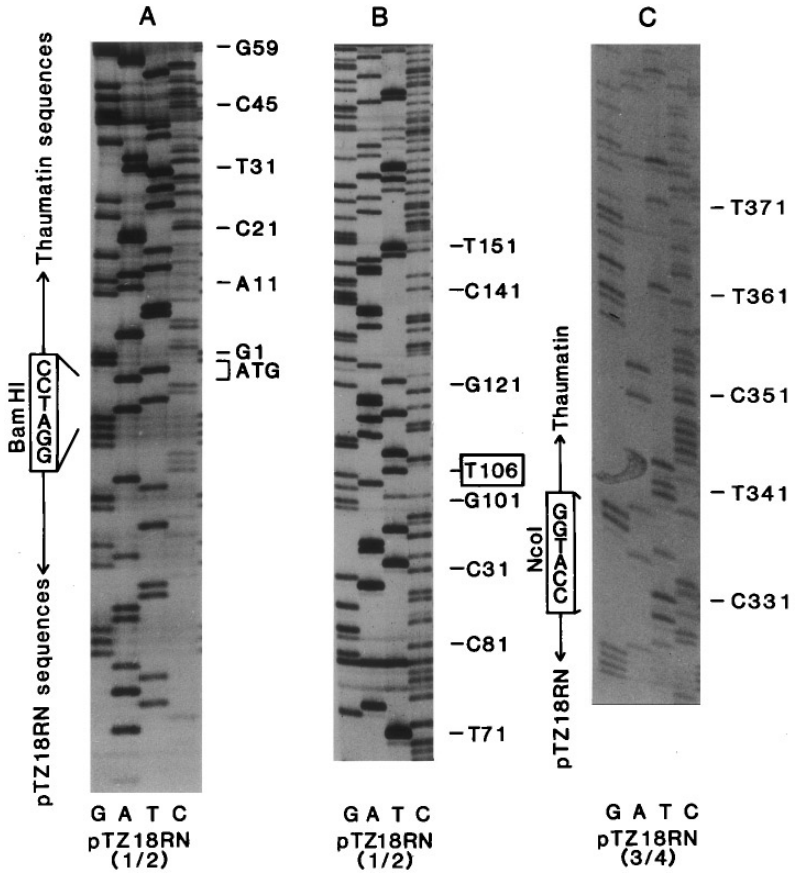

FIG. 3. Representative autoradiographs of the sequence of the synthetic gene. The sequences were generated using the Sanger dideoxy method: (A) nucleotides 1-60 of the synthetic gene; (B) nucleotides 70-170; (C) nucleotides 330-370.

(see Fig.2) Ligation reactions contained 15% polyethylene glycol (PEG), which is known to aid in blunt-end ligations. Reaction mixes were then used to transform *E. coli* strain NM 522. A white/blue selection was made again. After analyzing the transformants, one clone was isolated which contained the 305 bp fragment of the second part of the thaumatin II gene. This plasmid was called pTZ18RN (3/4).

Sequence analysis. The identity of the synthetic gene was verified by analyzing its sequence using the Sanger method. A sequentiation kit was used (version 2.0) from United States Biochemical Corp. The sequence of the synthetic gene was determined without ambiguity by: (1) sequentiation of the two gene strands; and (2) performing parallel sequentiation reactions with dITP to destabilize the potential secondary structures which could form due to GC rich areas in the synthetic gene. Representative autoradiographs are shown in Figure 3.

Subcloning of the Synthetic Thaumatin Gene in the E.coli Expression Vector pGEX-KG

Vector pGEX-KG was disgested with NcoI and HindIII, and ligated to the NcoI-HindIII fragment from pTZ18RN(3/4) containing fragments 3/4 of the thaumatin II gene, generating plasmid pECThI. This plasmid was then disgested with NcoI and ligated to the NcoI-NcoI fragment from pTZ18RN(1/2) containing fragments 1/2 of the thaumatin II gene, generating the expression vector pGEX-KG[Th]. This construct was used to transform *E.coli* strain DH5, using standard transformation techniques.

Subcloning of the Synthetic Thaumatin Gene in the E.coli Expression Vector pET-8c

Vector pET-8c was disgested with BamHI, repaired with the Klenow fragment of DNA polymerase I, and digested with NcoI. This vector was ligated with a 305-bp fragment that was generated by digesting pT18RN(3/

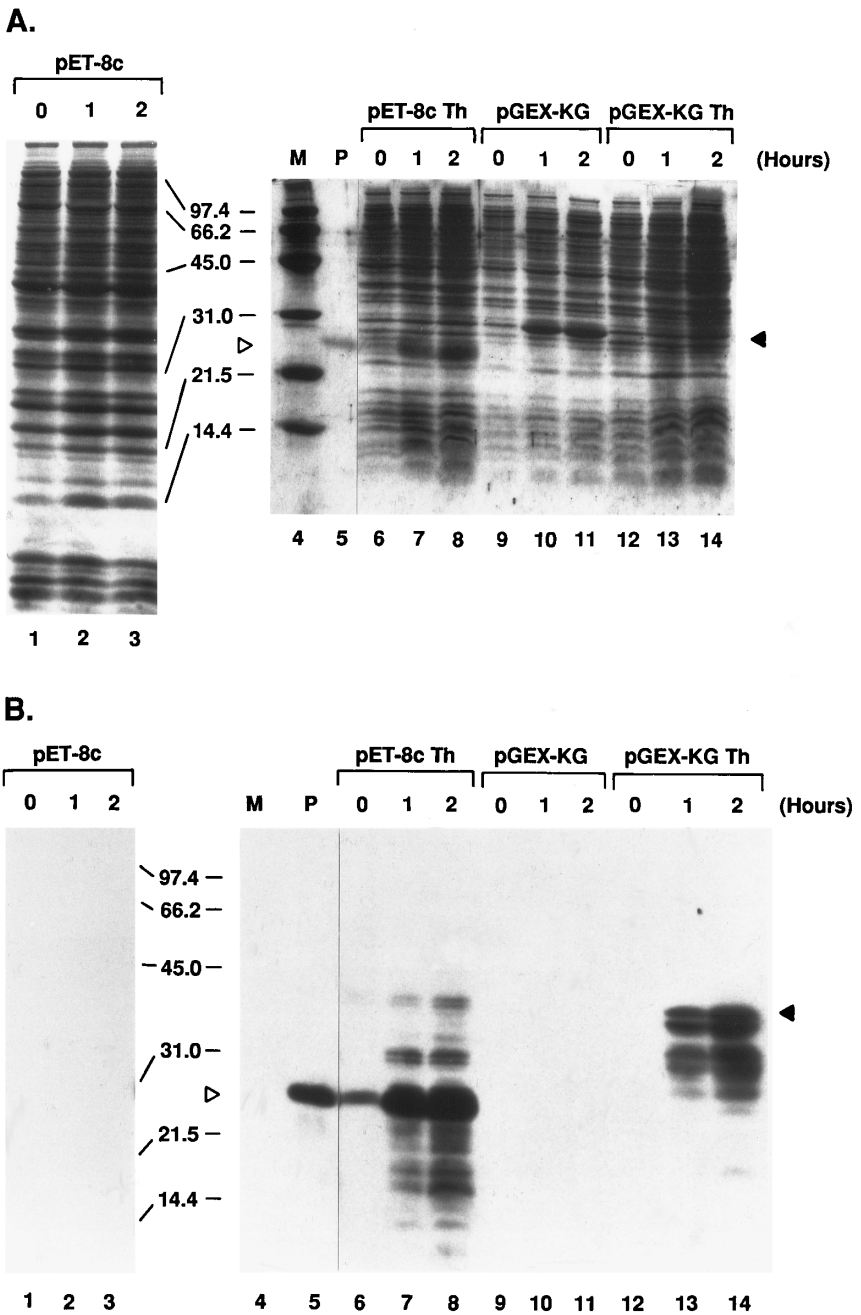

FIG. 4. SDS-PAGE analysis of the expression of thaumatin II in *Escherichia coli.* Cell pellets were re-supended in SDS-loading buffer and resolved in a 14% polyacrylamide gel. In both panels A and B, the numbers in between the two parts of the panel indicate the molecular weight of protein standards (lane 4). Lane 5 contains an aliquot of commercial thaumatin (CTC Organics). The open arrow indicates the position in the gel of thaumatin, while the closed arrow points to the 26 Kd glutathione S-transferase (GST) protein from *Schistosoma japonicum* (panel A) and the 48 Kd GST-thaumatin fusion protein (panel B). In panel A, the gel was stained with Coomassie-blue, while panel B shows the results of an immunoblot analysis.

4) with HindIII, repairing the ends with the Klenow fragment of DNA polymerase I, and further digesting the DNA with NcoI. The resulting plasmid was named pET-8c(201). This plasmid was then disgested with NcoI and ligated to the NcoI-NcoI fragment from pTZ18RN(1/2) containing fragments 1/2 of the thaumatin II gene, generating the expression vector pET-8c(Th). This construct was used to transform *E.coli* strain BL21(DE3), using standard transformation techniques.

Expression of Recombiant Thaumatin in Escherichia coli

1 ml. of strains DH5(pGEX-KG), DH5(pGEX-KG[Th]), BL21(DE3)(pET-8c) and BL21(DE3)(pET-8c[Th]) were grown overnight in liquid LB medium supplemented with 100 μg/ml. of ampicillin. The following morning, 50 μl of each of these ''starter'' cultures was added to 5 ml. of LB media (ampicillin also at 100 μg/ml.). The cultures were allowed to grow for 2 hours at 37°C, at which point an aliquot (time 0) was removed. IPTG to 1 mM was added to induce expression of recombinant thaumatin. After one and two hours, aliquots were removed (time points 1 and 2).

Immediately after their removal, samples were processed as follows: the cells were centrifuged in a table-top centrifuge, and after removal of the liquid supernantant were re-suspended in 50 μl of SDS-loading buffer. Samples were then resolved in a 14% SDS-polyacrylamide Laemli gel. Resolved proteins were transferred to nitrocellulose, and the sheets probed with a rabbit polyclonal anti-thaumatin antibody (1/15,000 dilution). The secondary antibody used was a goat anti-rabbit IgG (whole molecule) alkaline phospahatase conjugate. Some gels were also stained directly with a Coomassie Blue solution (0,05% Coomassie Brilliant Blue R-250 (BioRad) in 50% methanol, 10% acetic acid) , and de-stained with a 5% methanol/7% acetic acid de-staining solution.

RESULTS AND DISCUSSION

As a first step towards our goal of obtaining high-levels of recombinant thaumatin, we have assembled a synthetic gene encoding the DNA sequence of the 207 aminoacid protein thaumatin II and demonstrated that it can be expressed in *Escherichia coli.*

Figure 4 clearly shows that upon induction by IPTG, a protein of a molecular weight identical to natural thaumatin II is expressed in *Escherichia coli* strains harboring plasmid pET-8c(Th), but not in strains that harbour the native plasmid pET-8c (compare lanes 1-3 with lanes 5-8 0f Fig.4, panel B). The expression of this recombinant form of thaumatin is certainly strong, since it can also be detected in Coomassie Blue stained gels (lanes 6-8 of Fig.4, panel A).

We have also succeeded in expressing thaumatin a a fusion protein with the 26-Kd glutathione S-transferase (GST) protein, which is encoded by the parasitic helminth *Schistosoma japonicum.* Vector pGEX-KG, a derivative of pGEX-2T (Pharmacia, Inc.) contians the GST protein under the control of a strong *tac* promoter. The vector also contains the *lac* I^q gene, so it expression of the recombinant protein can be induced by addition of IPTG. The insertion of the synthetic thaumatin gene in the multiple cloning site (MCS) of this vector results in the synthesis of a GST-thaumatin fusion protein.

As shown in Fig. 4A, upon induction by IPTG, an *E.coli* strain harboring pGEX-KG synthesizes a 26-Kd protein that is not immunologically reactive with our anti-thaumatin antibody (compare panel A, lanes 9-11 with the corresponding lanes of panel B). On the other hand, an *E.coli* strain that harbours pGEX-KG(Th) can synthesize a high-molecular weight protein that is immunologically reactive with the ant-thaumatin antibody (Fig., 4B, lanes 12-14). The molecular weight of this protein corresponds to the approximate size of the expected GST-thaumatin fusion protein (48 Kd).

Once established the functionality of our synthetic thaumatin II gene, we will attempt to express this protein in different species of filamentous fungi. The results of this work will be reported in a second publication.

REFERENCES

1. Witty, M., and Higginbotham, J. D. Thaumatin, CRC Press, 1994.
2. Edens, L., Heslinga, L., Klok, R., Ledeboer, A. M., Maat, J., Toonen, M., Visser, C., and Verrips, C. T. (1982) *Gene* **18,** 1–12.

3. Illingworth, C., Larson, G., and Hellenkant, G. (1988) *Biotechnology Lett.* **10-8,** 587–592.
4. Illingworth, C., Larson, G., and Hellenkant, G. (1989) *Journal Ind. Microb.* **4,** 37–42.
5. Edens, L., Bom, I., Ledeboer, A. M., Maat, J., Toonen, M. J., Visser, C., and Verrips, C. T. (1984) *Cell* **37,** 629–633.
6. Hahm, Y. T., and Batt, C. A. (1990) *Agric. Biol. Chem.* **54,** 2513–2520.
7. Witty, M. (1990) *Biotechnology Letters* **12-2,** 131–136.