

CLONING AND EXPRESSION OF THE MAJOR ALLERGEN FROM YELLOW MUSTARD SEEDS, *Sin a* I

M.A. González de la Peña¹, M. Villalba¹, J.L. García-López² and R. Rodríguez^{1*}

¹Departamento de Bioquímica y Biología Molecular, Facultad de
Ciencias Químicas, Universidad Complutense, 28040 Madrid, Spain

²Centro de Investigaciones Biológicas (CSIC), Velázquez 114,
28006 Madrid, Spain

Received December 15, 1992

The cloning and bacterial expression of the major allergen from yellow mustard (*Sin a* I) is reported. Cloning has been carried out by means of the polymerase chain reaction using non-degenerate oligo primers encoding the N- and C- terminal regions of the mature protein. Two nucleotide sequences (SA2S1 and SA2S2) have been found and analysed. The observed polymorphism suggests the existence of multiple isoforms for *Sin a* I allergen. SA2S2 has been expressed as a fusion protein, linked to the choline-binding domain of the *Streptococcus pneumoniae* murein hydrolase. The resulted fusion protein is recognized by *Sin a*-I-specific antibodies. © 1993 Academic Press, Inc.

The cloning and recombinant expression of allergens is relevant because allows the production of large quantities of pure protein for basic and clinical research. They are also potentially useful for the design and production of recombinant sequences to be used as competitive analogues of the IgE binding in clinical applications. However, so far there is a little number of cloned allergens (1), and the production of recombinant allergens has been limited to proteins from mites and plant pollens.

Food allergies are frequent immunological diseases (2). Among them, table mustard produces atopic responses in sensitive individuals which occasionally lead to anaphylactic shocks (3). The major allergen from yellow mustard (*Sinapis alba*) seeds is the 2S storage protein, *Sin a* I, which has been isolated (4), immunologically characterized (5) and sequenced (6). The 2S storage proteins from Brassicaceae (napins) are

* To whom correspondence should be addressed.

Abbreviations: PCR, polymerase chain reaction; IPTG, isopropyl-thio- β -D-galactoside; C-LYTA, C-terminal domain of murein hydrolase.

encoded by genes without introns, and all of them are synthesized as polypeptide precursors which must be processed by specific maturation proteases to render the two chains of the native protein (7-12). Due to both its small size and its relatively well known antigenic structure, *Sin a* I could serve as a model system to study the basic structural requirements for immune recognition of allergens. We have carried out in this research the cloning and bacterial expression of *Sin a* I, in order to obtain modified forms of this allergen.

MATERIALS AND METHODS

PCR amplification: Genomic DNA was isolated from *Sinapis alba* L. seeds by a method adapted from Murray and Thompson (13). The sequence encoding the precursor of *Sin a* I allergen without its leader sequence was amplified by the polymerase chain reaction (PCR) using non-degenerate oligo primers encoding the N- and C- terminal amino acid sequences of the *Sin a* I allergen (6). The codon usage was selected in basis of the nucleotide sequences of homologous proteins (7-12) and *EcoRI* restriction sites were incorporated at the 5' ends of each primer to facilitate ligation of the amplified DNA into vectors (14). Genomic DNA (50 ng) was amplified in 100 μ l of PCR mixture containing 0.5 μ M deoxynucleotide triphosphates (dNTPs), 0.5 μ M oligonucleotide primers, the manufacturer's buffer and 2 U of *Taq* polymerase (Boehringer). For reamplifying PCR products, the PCR reactions contained 1 ng of template in the same reaction mixture described above. All the reactions were carried out in quadruplicate.

Cloning of PCR products and sequencing: The purified PCR-generated DNA was digested with *EcoRI*, cloned into plasmid vector pUC18 and sequenced by double-stranded dideoxy chain termination method (15,16), using the Sequenase Kit (US Biochemicals).

Expression and immunodetection of the fusion protein: The SA2S2 PCR product was inserted into the pCE17 plasmid (17) and transfected into *E. coli* cells (DH1). The expression was obtained after induction with 1 mM IPTG, and the chimeric protein was detected in Western blotting (18) using rabbit polyclonal sera directed against the carrier protein (C-LYTA) or *Sin a* I.

RESULTS AND DISCUSSION

PCR amplification resulted in two fragments (Fig. 1) with estimated sizes of 440 and 370 bp. The 440 bp PCR-generated fragment had the appropriate size to encode the precursor of *Sin a* I allergen without the leader sequence. This fragment was isolated from agarose gels, PCR-reamplified and cloned into the *EcoRI* site of pUC18.

Two clones (SA2S1 and SA2S2) containing a 435 bp PCR-generated fragment were isolated and sequenced. Both nucleotide sequences (Fig. 2) encode the two polypeptide chains of yellow mustard allergen linked by the connecting peptide. The amino acid sequences deduced from these clones are not identical to the primary



Fig. 1. Agarose gel electrophoresis (2%) of PCR-amplified DNA. The location of the amplified bands is indicated by arrows. 1, Genomic *S. alba* DNA; 2, reamplified 440 pb band. Lane M contains HindIII digested λ DNA.

<i>Sin a</i> I	Pro Ala Gly Pro Phe Arg Ile Pro Lys Cys Arg Lys Glu Phe Gln Gln Ala Gln His Leu	20
SA2S1	<u>C</u> CA <u>G</u> GC <u>G</u> GC <u>C</u> CA <u>T</u> TT AGG ATT CCA AAA TGT AGG AAG CAG TTT CAG CAA GCA CAA CAC CTG	60
SA2S2		
<i>Sin a</i> I	Arg Ala Cys Gln Gln Trp Leu His Lys Gln Ala Met Gln Ser Gly Ser Gly Pro Ser Trp	40
SA2S1	AGA GCT TGC CAA CAA TGG CTC CAC AAG CAG GCA ATG CAG TCC GGT AGT GGT CCA AGC TGG	120
SA2S2		
<i>Sin a</i> I	Thr Leu Asp Asp Gln Phe Asp Phe Glu Asp Asp Met Glu Asn Pro Gln Gly Pro Gln Gln	60
SA2S1	ACC CTC GAC GAT GAG TTT GAT TTT GAA CAC GAT ATG GAG AAC CCA CAG GGA CCA CAG CAG	180
SA2S2		
<i>Sin a</i> I	Arg Pro Pro Leu Leu Gln Gln Cys Cys Asn Glu Leu His Gln Glu Gln Pro Leu Cys Val	80
SA2S1	AGG CCA CCA CTA CTC CAG CAG TGC TGC AAC GAG CTC CAC CAG GAA CAG CCA CTT TGC GTT	240
SA2S2		
<i>Sin a</i> I	Cys Pro Thr Leu Lys Gly Ala Ser Lys Ala Val Lys Gln Gln Val Arg Gln Gln Leu Gly	100
SA2S1	TGC CCA ACC TTG AAA GGA GCA TCC AAA GCC GTT AAA CAG GTT AGA CAA CAG CTG GGG	300
SA2S2		
<i>Sin a</i> I	Gln Gln Gly Gln Gln Gly Pro His Leu Gln His Val Ile Ser Arg Ile Tyr Gln Thr Ala	120
SA2S1	CAG CAG GGG CAG CAG GGA CCG CAG CTG CAG CAT GTA ATT AGC CGT ATC TAC CAG ACT GCT	360
SA2S2		
<i>Sin a</i> I	Thr His Leu Pro Lys Val Cys Asn Ile Arg Gln Val Ser Val Cys Pro Phe Lys Lys Thr	140
SA2S1	ACA CAC TTA CCT AAA GTT TGC AAC ATC AGG CAA GTT AGT GTT TGT CCC TTC AAG AAG ACC	420
SA2S2		
<i>Sin a</i> I	Met Pro Gly Pro Ser	145
SA2S1	<u>A</u> TG <u>G</u> CT <u>G</u> GA <u>C</u> CC <u>T</u> CC	435
SA2S2		

Fig. 2. Nucleotide sequence of PCR products SA2S1 and SA2S2. The deduced amino acid sequence is shown above (SA2S1) or below (SA2S2) the codons. In the first line are represented the differences in respect to the previously determined primary structure of a mature allergen isoform, *Sin a* I (6). The arrows indicate the cleavage sites during the processing of the precursor protein. Line between arrows represents the linker segment, which is absent in the mature allergen. The oligo primers used in the amplification reactions (without the *Eco*RI sites) are underlined.

structure determined for the mature allergen isolated from seeds. This fact confirms the existence of polymorphism in this allergen, which was previously detected during the elucidation of its primary structure by Edman degradation (6).

Both sequences show small differences among them which are mainly located in the linker peptide removed during maturation. It is well established that evolutionary changes are faster in sequences encoding either nonfunctional polypeptide segments (19), or regions removed during the postranslational maturation, e.g. leader peptides or connecting fragments (20). Therefore, a higher heterogeneity should be expected in the nucleotide region encoding the connecting peptide of the *Sin a* I precursor than in that corresponding to the mature protein.

It is interesting to point out that heterogeneities observed near the processing sites can play an important role in the proteolytic maturation of the 2S storage proteins, since they can affect the structural motifs recognized by the endoprotease activities involved in that process. These microheterogeneities could lead to different lengths and sequences in the polypeptide chains of the mature protein as it has been observed for napins (21). On the other hand, Asn at position -1 of the large chain originate a sequence motif that could be recognized by the proteolytic activity recently proposed as a functional maturation protease in seeds (22).

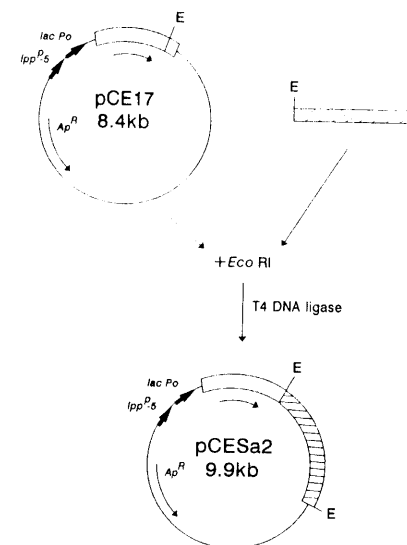


Fig. 3. Construction of pCESa2 encoding a C-LYTA-SA2S2 fusion protein. The white box indicates the sequence encoding the C-LYTA protein. The SA2S2 sequence is indicated by a dashed box. Other symbols are: E, *Eco*RI restriction site; *Ap^R*, ampicillin resistance; *lpp^P-5*, lipoprotein promoter; *lac^o*, lactose promoter/operator.

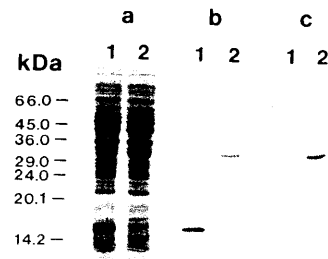


Fig. 4. Immunodetection of the C-LYTA-SA2S2 fusion protein. Lanes 1: crude extract of *E. coli* DH1 [pCE17] used as control; lanes 2: crude extract of *E. coli* DH1 [pCESa2]. (a) SDS/PAGE gel stained with Coomassie blue. (b) Western blot analysis developed with anti-(C-LYTA) serum. (c) Western blot analysis developed with anti-(*Sin a I*) serum. Positions of molecular mass markers are indicated on the left.

In order to obtain the bacterial expression of *Sin a I* allergen, the SA2S2 PCR product was inserted into the pCE17 plasmid. This plasmid contains the 3' moiety of the *lytA* gene from *Streptococcus pneumoniae* (C-LYTA) downstream of a modified *E. coli* lipoprotein promoter (*lpp⁺-5*) and the lactose promoter/operator region (*lac^o*) (Fig. 3). C-LYTA is a choline binding domain that should allow the purification of the recombinant fusion protein by using an affinity chromatography in DEAE-cellulose (23). After induction with IPTG, a cytoplasmic chimeric protein of an apparent size of 32 kDa is detected in cell lysates using rabbit polyclonal sera obtained against C-LYTA or *Sin a I* (Fig. 4). This recombinant protein is produced in an estimated concentration of 8-10 mg per litre of cell culture, and it is accumulated into cells in an insoluble form.

The insoluble nature of this chimeric protein has disabled its purification from the crude bacterial homogenate in a single chromatographic step using DEAE-cellulose. But, insolubility can facilitate the purification of recombinant proteins (24). However, the overall yield of the purification will be dependent on the efficiency of the *Sin a I* allergen renaturation. The structural and immunological characterization of recombinant *Sin a I* should be of unestimable aid to study structure-allergenicity relationships in this protein.

ACKNOWLEDGMENT

This work was supported by grant PB89/087 of the Dirección General de Investigación Científica y Técnica (Spain).

REFERENCES

- 1 Scheiner, O. (1992) Int. Arch. Allergy Immunol. 98, 93-96.
- 2 Johnstone, D.E. (1981) Respiration 42, 34-41.

- 3 Niinimäki, A. and Hannuksela, M. (1981) Allergy 36, 487-493.
- 4 Menéndez-Arias, L., Monsalve, R.I., Gavilanes, J.G., and Rodríguez, R. (1987) Int. J. Biochem. 19, 899-907.
- 5 Menéndez-Arias, L., Domínguez, J., Moneo, I. and Rodríguez, R. (1990) Mol. Immunol. 27, 143-150.
- 6 Menéndez-Arias, L., Moneo, I., Domínguez, J. and Rodríguez, R. (1988) Eur. J. Biochem. 177, 159-166.
- 7 Crouch, M.L., Tenbarger, K.M., Simon, A.E. and Ferl, R. (1983) J. Molec. Appl. Genet. 2, 273-283.
- 8 Ericson, M.L., Rodin, J., Lenman, M., Glimelius, K., Josefsson, L.-G. and Rask, L. (1986) J. Biol. Chem. 261, 14576-14581.
- 9 Josefsson, L.-G., Lenman, M., Ericson, M.L. and Rask, L. (1987) J. Biol. Chem. 262, 12196-12201.
- 10 Scofield, S.R. and Crouch, M.L. (1987) J. Biol. Chem. 262, 12202-12208.
- 11 Baszczynski, C.L. and Fallis, L. (1990) Plant. Mol. Biol. 14, 633-635.
- 12 Raynal, M., Depigny, D., Grellet, F. and Delseny, M. (1991) Gene 99, 77-86.
- 13 Murray, H.G. and Thompson, W.F. (1980) Nucleic Acids Res. 8, 4321-4325.
- 14 Scharf, S., Horn, G.T. and Erlich, H.A. (1986) Science 233, 1076-1078.
- 15 Sambrook, J., Fritsch, E.F. and Maniatis T. (1989) Molecular Cloning: a laboratory manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.
- 16 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- 17 Sánchez-Puelles, J.M., Sanz, J.M., García, J.L. and García, E. (1990) Gene 89, 69-75.
- 18 Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA 76, 4350-4354.
- 19 Poljak, R.J., Amzel, L.M. and Phizackerley, R.P. (1976) Prog. Biophys. Molec. Biol. 31, 67-93.
- 20 Padlan, E.A. and Davis, D.R. (1975) Proc. Nat. Acad. Sci. USA 72, 819-823.
- 21 Monsalve, R.I., Menéndez-Arias, L., López-Otín, C. and Rodríguez R. (1990) FEBS Lett. 263, 209-212.
- 22 Hara-Nishimura, I., Inoue, K. y Nishimura, M. (1991) FEBS Lett. 294, 89-93.
- 23 Sánchez-Puelles, J.M., Sanz, J.M., García, J.L. and García, E. (1992) Eur. J. Biochem. 203, 153-159.
- 24 Marston, F.A.O., Lowe, P.A., Doel, M.T., Schoemaker, J.M., White S. and Angal, S. (1984) BIO/TECHNOL. 2, 800-804.