

Expression in *Escherichia coli* of Sin a 1, the major allergen from mustard

Manuel A. GONZÁLEZ DE LA PEÑA, Rafael I. MONSALVE, Eva BATANERO, Mayte VILLALBA and Rosalía RODRÍGUEZ

Departamento de Bioquímica y Biología Molecular, Facultad de Química, Universidad Complutense, Madrid, Spain

(Received 16 November 1995/22 January 1996) – EJB 95 1902/2

Sin a 1, the major yellow mustard allergen, is a seed storage protein that belongs to the 2S albumin family. It is composed of two disulfide-bonded polypeptide chains. The cloning of this allergen has been carried out by means of the polymerase chain reaction using non-degenerate oligonucleotides encoding the N-terminal and C-terminal regions of the mature protein as primers. Five genomic nucleotide sequences have been analyzed, encoding both mature polypeptide chains linked by the internal processed fragment. The sequence data show the existence of microheterogeneities at ten positions, demonstrating the polymorphism exhibited by the natural protein. One of the genomic clones was expressed in *Escherichia coli* by fusion to glutathione *S*-transferase from *Schistosoma japonicum*. The resulting chimeric protein was purified by affinity chromatography on a glutathione–Sephrose 4B matrix, and digested with thrombin to release the recombinant allergen. The recombinant Sin a 1 is recognized by rabbit polyclonal and mouse monoclonal antisera raised against natural Sin a 1, as well as by the IgE of mustard-sensitive human sera. In addition, recombinant Sin a 1 possesses a high resistance to trypsin digestion, like the native mustard allergen.

Keywords: allergen; Sin a 1, recombinant allergen; mustard allergy; food allergy.

IgE-mediated (type I) allergy is the most common immunological disorder, which affects about 20% of the human population in developed countries. The identification and characterization of the antigens involved in IgE-mediated reactions are essential for understanding the mechanisms of type I allergy, as well as for the future development of effective therapies. Although foods are one of the major sources of allergens, molecular studies on food allergens are scarce [1, 2], in comparison to those on allergens from other biological origins, such as pollens, dust mites or venom insects [3]. Food allergens are usually stable proteins, which are capable of crossing the gut barrier and triggering the IgE-mediated reactions. These proteins have been studied mainly in cow's milk [4, 5], hen's eggs [6, 7], cod fish [8, 9], and peanut [10, 11].

Many spices have been described that lead to allergic reactions after their ingestion [12]. Among them, mustard is one of the most reactive when tested *in vivo* [13]. The major allergen from yellow mustard (*Sinapis alba*) seeds, Sin a 1, has been characterized as a basic, polymorphic, low-molecular-mass protein (14.1 kDa), composed of two polypeptide chains, of 39 and 88 amino acids, held together by two disulfide bonds [14]. The amino acid sequence of Sin a 1 allergen has been determined by

Edman degradation [15], and its antigenic structure studied by using monoclonal antibodies [16]. Sin a 1 belongs to the 2S albumin protein family, also called napins. These proteins are synthesized *in vivo* as a precursor polypeptide which contains three additional fragments with respect to the mature protein: a N-terminal leader region, an internal processed fragment (IPF) and an extra amino acid residue [17]. The main biological function of napins seems to be as a nitrogen supply to the germinating seedling. Recently, an antifungal activity has also been described for these proteins [18].

In recent years, the use of recombinant DNA technology has proved to be a convenient tool to obtain homogeneous allergen preparations [19]. The production of recombinant allergens could also allow the identification of residues involved in allergenic determinants, as well as allowing for the isolation of well-defined proteins which are required for diagnosis and treatment of IgE-mediated allergic disorders. In the present study, we report the bacterial expression and purification of an immunologically active recombinant Sin a 1 allergen.

MATERIALS AND METHODS

Isolation of Sin a 1 allergen. Natural Sin a 1 allergen was purified from yellow mustard (*Sinapis alba*) seed flour as reported [14].

Isolation of genomic DNA. Total genomic DNA was isolated from yellow mustard seeds (Herborem) by a method adapted from [20]. 0.1 g flour obtained from yellow mustard seeds was suspended in 600 µl extraction buffer (50 mM Tris/HCl, pH 8.0, 10 mM EDTA, 1% cetyltrimethylammonium bromide) containing 0.7 M NaCl and 1% 2-mercaptoethanol. The sample was incubated at 56°C for 20 min, with occasional vigorous shaking. 600 µl chloroform/isoamyl alcohol (24:1, by vol.) were subsequently added and the emulsified mixture was centrifuged at 12000×g for 5 min at room temperature. The

Correspondence to R. Rodríguez, Departamento Bioquímica y Biología Molecular, Facultad de Química, Universidad Complutense, E-28040 Madrid, Spain

Abbreviations. IPF, internal processed fragment; IPTG, isopropylthio- β -D-galactoside; GST, glutathione *S*-transferase; rSin a 1, recombinant Sin a 1.

Enzymes. Trypsin (EC 3.4.21.4); *Taq* DNA polymerase (EC 2.7.7.7); ribonuclease A (EC 3.1.27.5); T4 DNA ligase (EC 6.5.1.1); thrombin (EC 3.4.21.5); Sequenase version 2.0 polymerase (EC 2.7.7.7); restriction endonucleases *Bam*HI and *Eco*RI (EC 3.1.21.4).

Note. The novel nucleotide sequences published here have been submitted to the EMBL sequence data library and are available under accession numbers X91799 (SIN1), X91800 (SIN2), X91801 (SIN3), X91802 (SIN4), and X91798 (SIN5).

aqueous phase was collected and mixed with 600 µl extraction buffer. The mixture was incubated for 20 min at room temperature and centrifuged for 10 min under the same conditions described above. The pellet was redissolved in 600 µl 1 M NaCl and the DNA was extracted with phenol/chloroform and precipitated with ethanol as described in [21]. The new pellet was redissolved in 200 µl 10 mM Tris/HCl, pH 8.0, containing 1 mM EDTA, and digested with 100 µg/ml ribonuclease (Sigma) for 2 h at 37°C. DNA was reprecipitated and dissolved in 50 µl sterile water.

Synthetic oligonucleotides. Non-degenerate oligonucleotide primers used in PCR reactions were synthesized on an Applied Biosystems DNA synthesizer model 381A according to standard procedures [21]. The primers used encode the N-terminal and C-terminal amino acid sequences of the Sin a 1 allergen [15]. The codon usage was selected on the basis of the nucleotide sequences of homologous proteins [17, 22–27]. *Bam*HI and *Eco*RI restriction sites were incorporated at the 5' ends of each primer to facilitate the ligation of the amplified DNA into vectors [28]. The sequences of the primers were: SA-1, sense primer 5'-CCGGGATCCCCAGCCGGCCCATTT-3' and SA-2, antisense primer 5'-CCGGAATTCTCAGGAGGGTCCAGG-CAT-3'.

PCR reactions. 50 ng *S. alba* genomic DNA were amplified in 100 µl reaction mixture containing 1 µM each primer, 200 µM dNTP and 2 U *Taq* DNA polymerase (Boehringer Mannheim GmbH) in 10 mM Tris/HCl, pH 8.3, 1.5 mM MgCl₂, 50 mM KCl, 0.1 mg/ml gelatin. The same reaction mixture, without the template DNA, was used as a negative control for the amplification. After a denaturation step at 95°C for 15 min, the sample was subjected to five cycles of denaturation at 94°C (60 s), annealing at 47°C (60 s), and extension at 72°C (90 s); followed by 20 cycles of denaturation at 94°C (60 s), annealing at 55°C (90 s), and extension at 72°C (90 s). The final extension step was increased to 10 min. Reactions were carried out in quadruplicate, in a PHC-3 Thermal Cycler (Techne Corp.).

Cloning and sequencing of the PCR fragment. The PCR-amplified product was digested with *Bam*HI and *Eco*RI, purified by agarose gel electrophoresis, and ligated into pUC18 plasmid by using standard procedures [21]. The recombinant plasmid was used to transform *E. coli* strain TG1 [29]. DNAs from several pUC18 clones were sequenced using the dideoxynucleotide chain-termination method [30] with the Sequenase kit (US Biochemical Corp.) and [α -³⁵S]-dATP[S]. Sequencing primers for both strands, M13mp18 universal and reverse, were purchased from New England Biolabs. The sequencing was carried out according to the manufacturer's instructions.

Expression of recombinant fusion Sin a 1. After sequencing, the DNA clone SIN2 was subcloned into the *Bam*HI/*Eco*RI sites of the pGEX-2T expression vector [31]. The obtained construction encodes a fusion protein (GST-rSin a 1) consisting of glutathione *S*-transferase (GST) from *Schistosoma japonicum* and a 6-amino-acid peptide (Leu-Val-Pro-Arg-Gly-Ser, a thrombin cleavage site), followed by the Sin a 1 small and large chains linked by the IPF.

Production of the recombinant protein was performed after transforming DH5 α F' cells [32] with the pGEX-2T/SIN2 plasmid. Overnight cultures of transformed cells were diluted tenfold with Luria-Bertani broth medium containing 0.1 mg/ml ampicillin. Cells were grown up to an absorbance value at 600 nm of 0.6. After induction with 1 mM IPTG (isopropyl-thio- β -D-galactoside, Sigma), cultures were harvested by centrifugation at 4000 $\times$ g for 20 min at 4°C. The pellet was resuspended in NaCl/P_i (10 mM sodium phosphate, pH 7.2, 140 mM NaCl, 2.7 mM KCl) containing 1% Triton X-100 and 1 mM phenylmethylsulfonyl fluoride, and sonicated. The obtained cell

lysate was incubated for 30 min at 4°C under continuous stirring. Soluble and insoluble fractions were then separated by centrifugation at 10000 $\times$ g for 10 min at 4°C.

Purification of rSin a 1 allergen. The above soluble fraction (1 l original culture) was incubated with 0.5 ml glutathione-Sepharose 4B (Pharmacia Biotech Inc.) for 1 h at room temperature under continuous shaking. Afterwards, the sample was centrifuged at 500 $\times$ g for 5 min, and the supernatant was discarded. The resin was washed three times with ten volumes of NaCl/P_i and, once the excess buffer was removed, incubated for 1 h with one volume NaCl/P_i containing 10 NIH U/ml thrombin (Pierce Chemical Co.) at room temperature under continuous shaking. After thrombin digestion, the sample was microfuged for 1 min at 4°C, and the supernatant, which contains the released rSin a 1, was collected.

Tryptic digestions. 10 µg natural or recombinant Sin a 1 were dissolved in 10 µl 0.2 M ammonium bicarbonate, pH 8.0, and 0.2 µg trypsin (Worthington), treated with *N*-tosyl-L-phenylalanine chloromethyl ketone, were added. The digestion reaction was carried out at 37°C for 24 h.

Electrophoresis and immunoblotting. Protein samples were analyzed by SDS/PAGE according to Laemmli [33] in 15% polyacrylamide gels. Proteins were visualized by Coomassie brilliant blue staining or electrophoretically transferred onto nitrocellulose membranes (Bio-Rad) for immunostaining as described [34].

For IgE-binding analysis, the membranes were incubated for 2 h at room temperature with a pool of mustard-allergic patient sera diluted 1:10 in blocking buffer (3% skim milk in NaCl/P_i containing 0.1% Tween 20). Afterwards, the membranes were incubated with mouse anti-human IgE monoclonal antibody (kindly donated by Dr Manuel Lombardero, Abelló S. A.) diluted 1:5000 in blocking buffer, and incubated with horseradish-peroxidase-labeled goat anti-mouse IgG (Pierce Chemical Co.) diluted 1:5000 in blocking buffer. The peroxidase reaction was developed with the ECL Western blotting reagent (Amersham Corp.) according to the manufacturer's instructions.

ELISA inhibition assays. ELISA inhibition assays were performed as previously described [16] with minor modifications. Microtiter plates were coated with 100 µl/well of natural Sin a 1 allergen (0.1 µg/ml) in NaCl/P_i, and blocked with blocking buffer. For IgE inhibition assays, 100-µl aliquots of a pool of mustard-allergic patient sera (1:10 dilution in blocking buffer), previously incubated with different concentrations of the inhibitor, were subsequently added and incubated for 2 h at 37°C. Binding of human IgE was detected by mouse anti-human IgE monoclonal antibody diluted 1:5000 in blocking buffer, followed by horseradish-peroxidase-labeled goat anti-mouse IgG (Pierce Chemical Co.) diluted 1:5000 in blocking buffer. The peroxidase reaction was performed as described in [16].

For inhibition assays of Sin a 1-specific monoclonal antibodies, blocked wells were incubated with 100-µl aliquots of the corresponding antibody (1:100 in blocking buffer), co-incubated with different concentrations of the inhibitor. Samples in wells were incubated with horseradish-peroxidase-labeled goat anti-mouse IgG, and the peroxidase reaction was developed as in the IgE assays.

RESULTS

Sequence analysis of PCR-amplified clones coding for Sin a 1. Genomic sequences coding for both chains of Sin a 1 allergen and the IPF were amplified by PCR. Non-degenerate primers encoding the N-terminal sequence of the small chain and the C-terminal sequence of the large chain were designed

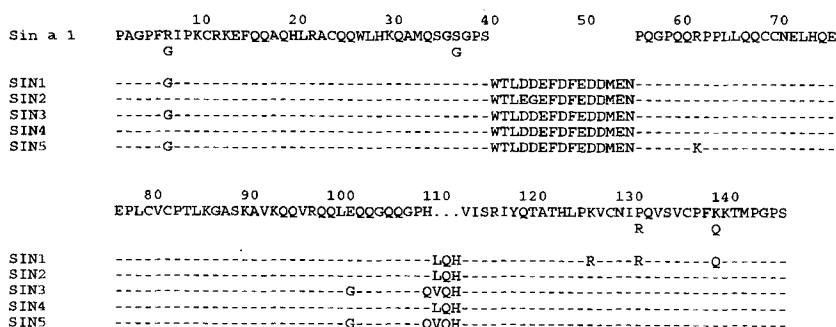


Fig. 1. Deduced amino acid sequences of the obtained PCR clones encoding Sin a 1 allergen. Sequences are shown in comparison with the sequence obtained by Edman degradation of the purified Sin a 1 [15]. Microheterogeneities detected in the natural protein [15] are shown below the main sequence. Residue identity – with respect to the main sequence of Sin a 1 is indicated. Serine and proline at positions 39 and 55 correspond to the C-terminal of the small chain and the N-terminal of the large chain, respectively, in the mature protein. Gaps are indicated by dots (·).

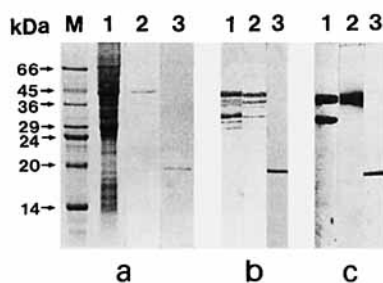


Fig. 2. SDS/PAGE analysis of the recombinant Sin a 1 protein (reducing conditions). (a) Coomassie brilliant blue staining, (b) Western blot immunostained with Sin a 1-specific rabbit polyclonal antiserum, (c) Western blot immunostained with human IgE from mustard-sensitive individuals. Lane M, molecular mass markers (Bio-Rad); lane 1, total cell lysate from *E. coli* strain with pGEX-2T/SIN2 insert; lane 2, GST-rSin a 1 fusion protein eluted from the glutathione-Sepharose 4B matrix; lane 3, rSin a 1 released after thrombin treatment. Before their application on the gel, all the samples were treated with 5% 2-mercaptoethanol at 80°C for 20 min.

based on the primary structure of the natural protein obtained from mustard seeds [15]. Two DNA fragments were produced after amplification (data not shown), one of them approximately 440 bp in size, in good agreement with the size expected from the amino acid sequence of Sin a 1 plus the described IPFs from homologous 2S proteins [17, 22–27]. This fragment was isolated from agarose gels and cloned into pUC18. Five clones were sequenced, all of them encoding the two polypeptide chains of the yellow mustard allergen linked by the IPF. The amino acid sequences deduced from these clones are shown in Fig. 1, in comparison with the sequence obtained from the natural protein by Edman degradation. The SIN2 clone was selected for expression in *E. coli*, due to its similarity with the main sequence obtained for the natural allergen [15].

Expression of rSin a 1. To obtain the expression of Sin a 1 allergen in *E. coli*, the SIN2 clone was inserted into the pGEX-2T plasmid. This construction would produce a fusion protein (GST-rSin a 1) composed of glutathione *S*-transferase and the recombinant allergen, linked through a hexapeptide that contains a thrombin cleavage site [31]. *E. coli* cells transformed with the recombinant plasmid expressed in a high yield a polypeptide of the expected molecular size of about 43 kDa (Fig. 2a). When transferred to nitrocellulose sheets, this protein was recognized by a rabbit polyclonal serum raised against natural Sin a 1 (Fig. 2b), as well as by a pool of sera from mustard-allergic individuals (Fig. 2c).

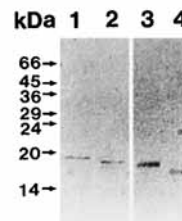


Fig. 3. Trypsin resistance of rSin a 1. SDS/PAGE (under non-reducing conditions) of the following: lane 1, undigested rSin a 1; lane 2, rSin a 1 after 24 h of trypsin digestion; lane 3, undigested natural Sin a 1; lane 4, natural Sin a 1 after 24 h trypsin treatment.

After lytic treatment of the transformed cells, more than 98% of the fusion protein remained insoluble. The formation of inclusion bodies could not be prevented, since it was not affected by culture conditions (temperature, media composition, IPTG concentration or time of induction) or by the *E. coli* strain used. The use of denaturant agents as detergents, organic solvents, urea, or guanidinium chloride, did not result in any improvement of the solubility of the recombinant protein. The expression of other Sin a 1 clones instead of SIN2 did not lead to a higher solubility of the recombinant product.

The soluble fraction was purified by using glutathione-Sepharose resin as described in the Materials and Methods section. The addition of thrombin to the GST-rSin a 1 fusion protein bound to the resin releases a soluble protein of about 18.8 kDa (Fig. 2), which must be composed of the two polypeptide chains present in natural Sin a 1, linked by the IPF. As a consequence of the cloning strategy, rSin a 1 should also contain two extra amino acids, Gly-Ser, at its N-terminus. In Western blot analysis, IgE antibodies from sera of mustard-allergic patients bind to rSin a 1 (Fig. 2), indicating the presence of allergenic epitopes in this recombinant protein.

A remarkable characteristic of the natural Sin a 1 allergen is its resistance to trypsin hydrolysis. After 24 h of trypsin digestion under optimal conditions, rSin a 1 remained essentially undegraded (Fig. 3). SDS/PAGE analysis of the digested products showed the loss of about 1.3 kDa with respect to the untreated rSin a 1. These data agree with the result obtained for the natural protein isolated from mustard seeds (Fig. 3), which indicates that the native conformation was retained by the obtained recombinant allergen.

Inhibition assays. To perform ELISA inhibition assays, microtiter plates were coated with the natural Sin a 1 allergen and incubated with Sin a 1-specific antibodies previously incubated with

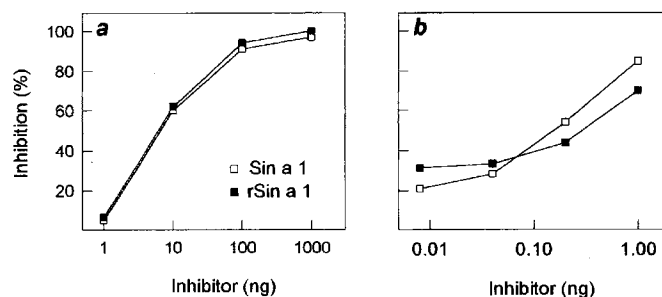


Fig. 4. Competitive inhibition of the binding of Sin a 1-specific polyclonal antibodies to solid-phase Sin a 1 by natural (□) or recombinant (■) Sin a 1. (a) ELISA inhibition of rabbit IgG binding to Sin a 1-coated wells, (b) ELISA inhibition of human IgE binding to Sin a 1-coated wells.

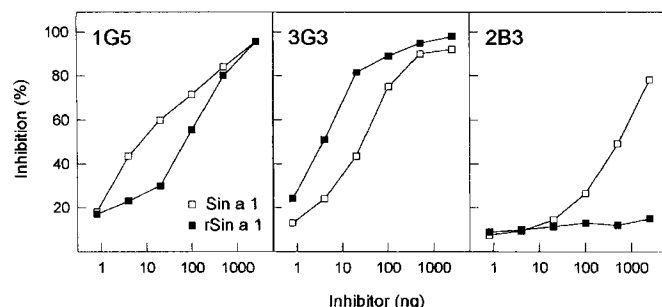


Fig. 5. mAb binding inhibition by natural (□) or recombinant (■) Sin a 1. Polystyrene plates were coated with natural Sin a 1. The binding of Sin a 1-specific mouse mAb 1G5, 3G3, and 2B3 was determined after addition of horseradish-peroxidase-labeled goat anti-mouse IgG.

several dilutions of natural or recombinant Sin a 1. Both antigens were able to inhibit at a similar level rabbit IgG binding to immobilized Sin a 1 (Fig. 4a), which indicates that they share most of their B-cell epitopes.

To compare in more detail the antigenic structure of recombinant and natural Sin a 1, inhibition experiments with three monoclonal antibodies (1G5, 3G3, and 2B3) were performed (Fig. 5). 1G5 and 3G3 antibodies bind to conformational antigenic determinants and belong to different complementation groups [16], while 2B3 recognizes a continuous epitope in the large chain of natural Sin a 1 [35]. Total inhibition of the binding of 1G5 and 3G3 antibodies to solid-phase linked Sin a 1 was obtained by using the recombinant antigen as inhibitor. These monoclonal antibodies display a capacity to bind both rSin a 1 and the natural mustard allergen, although they showed different affinities. However, rSin a 1 was not able to inhibit the binding of 2B3 antibody to natural Sin a 1 adsorbed to polystyrene wells (Fig. 5). This result indicates that the antigenic determinant recognized by this monoclonal antibody is absent in the obtained rSin a 1.

Finally, IgE inhibition assays were performed by using a pool of five sera from mustard-allergic patients. The obtained inhibition curves for the natural and the recombinant allergen (Fig. 4b) show that most of the allergenic determinants are present in rSin a 1.

DISCUSSION

The expression of a soluble, immunologically active rSin a 1 mustard allergen involves remarkable difficulties, since natural Sin a 1 is a cysteine-enriched protein composed of two polypep-

tide chains linked together by disulfide bonds. Here, we describe a simple and efficient method to produce a rSin a 1 allergen in *E. coli*, consisting of the expression of the continuous nucleotide sequence that encodes both polypeptide chains linked by the IPF. It was expected that the product of such a construction should be conformationally very similar to the natural protein, because the folding of this molecule *in vivo* must occur before the proteolytic processing of the IPF [36]. Given that napins are encoded by genes without introns [17, 22–27], it was possible to use a genomic sequence for the expression of Sin a 1 allergen, which, in conjunction with the use of PCR, avoids the tedious screening of cDNA libraries.

By using PCR, five genomic clones have been synthesized and sequenced. Their deduced amino acid sequences show small differences among them and with respect to the primary structure determined for the mature allergen isolated from seeds. The five clones also differ from two previously described genomic Sin a 1 sequences [37]. These results confirm the existence of polymorphism in this allergen, which was previously detected during the elucidation of its primary structure by Edman degradation. Microheterogeneities are mainly found at positions 100–114 of the deduced sequences (Fig. 1). This is located in the hypervariable region, a highly heterogeneous domain present in the large chain of the 2S albumins [38]. The main difference between the deduced amino acid sequences from the five clones analyzed and that from the natural allergen is the insertion of three additional residues at positions 109–111 of rSin a 1. This discrepancy is probably due to geographical variations, since PCR was performed on genomic DNA extracted from Spanish seeds, whereas the natural allergen was isolated from North American yellow mustard seed flour [14].

It is interesting to point out that the IPF shows very little polymorphism. This is important, because heterogeneities located near the processing sites could affect the structural motifs recognized by the endoprotease activities involved in that process. All the clones possess the residue pairs Phe-Asp (positions -9 and -8 of the large chain) and Asp-Asp (-5 and -4), which have been identified as cleavage sites for an acidic endoprotease described in *Brassica napus* seeds as one of the enzymes involved in the maturation process of napins [39].

The fusion protein GST-rSin a 1 was produced in high yield (about 50 mg/l culture), but it was mainly accumulated into cells as an insoluble form (more than 98%). The expression of eukaryotic proteins in *E. coli* leads frequently to insoluble, inactive polypeptides, which accumulate intracellularly in aggregates called inclusion bodies [40]. These structures are favored on proteins rich in disulfide bonds, which is the case of Sin a 1 allergen. Moreover, proline-rich proteins (rSin a 1 has a proline content of 11%) are inefficiently folded in *E. coli* [41]. Therefore, not only rSin a 1 but recombinant 2S albumins in general, are not easily produced in *E. coli*, and, so far, they have not been obtained in a soluble form using this expression system. Only very recently, an *Arabidopsis thaliana* 2S albumin gene has been expressed in yeast [42].

The expression of SIN2 clone using pGEX-2T vector resulted in the expected GST-rSin a 1 fusion protein (60–100 µg/l culture), as well as several other fragments which were also recognized by Sin a 1-specific polyclonal antibodies. These minor polypeptides are probably the products of proteolytic cleavage of the fusion protein, favored by an incorrect folding. The soluble fraction of GST-rSin a 1 fusion protein was isolated by using a glutathione-Sepharose 4B matrix. The subsequent thrombin digestion of the matrix-bound fusion protein resulted in a pure rSin a 1 in soluble form, with only two additional residues in its N-terminal end. The apparent molecular mass of rSin a 1 (18.8 kDa) is greater than the expected

(16.8 kDa), but agrees well with the electrophoretic mobility of the natural allergen, which shows in SDS/PAGE an apparent size of 16.5 kDa, about 2 kDa greater than the real one (14.1 kDa, determined by Edman degradation).

All the immunochemical assays carried out on rSin a 1 showed that its antigenicity and allergenicity are very similar to those of the natural mustard allergen. The antigenic properties of recombinant allergens are usually analyzed by means of Western blotting techniques, in which proteins are previously denatured. Therefore, those experiments can only give us information about sequential epitopes. The ELISA techniques herein employed allow also the study of conformational antigenic determinants. The main antigenic difference between recombinant and natural Sin a 1 lies in the binding to the 2B3 monoclonal antibody. This recognizes a continuous epitope located in the large chain of natural Sin a 1 [16], but it does not bind to the recombinant allergen. This apparent discrepancy might be related to the presence of three additional residues (109–111) in rSin a 1, since the recognition by 2B3 has been proved to be very sensitive to sequence variations at this region of the allergen (positions 50–60 in the large chain of the natural Sin a 1, which correspond to residues 104–117 in Fig. 1) [35].

Although rSin a 1 displays the two immunogenic regions recognized by 1G5 and 3G3 monoclonal antibodies, some differences between natural and recombinant Sin a 1 are also deduced from the inhibition plots. This may be related to slight conformational changes arising from the presence of the IPF and the two additional N-terminal residues on rSin a 1. Such differences may also be a consequence of the polymorphic character of Sin a 1. These monoclonal antibodies, which were obtained against the natural allergen, could exhibit different affinities for its component isoforms, and thus for the single molecular species of rSin a 1.

A remarkable resistance to trypsin digestion is a feature of the native Sin a 1 allergen [43]. Trypsin resistance can be considered a highly conformation-dependent property in this protein, since its amino acid sequence is very rich in lysine and arginine residues (15.5%) leading to the existence of many putative trypsin cleavage sites. Trypsin resistance of rSin a 1, considered together with the antigenic similarity of the recombinant and natural mustard allergens, indicates that the three-dimensional structure of the segments corresponding to the large and small chains on the soluble rSin a 1 isolated from *E. coli* cells should be closely similar to that of the mature protein obtained from yellow mustard seeds.

In conclusion, these results confirm the validity of our strategy, which consists of the expression of the Sin a 1 allergen as a precursor form, since this form is able to fold properly in *E. coli*, and the IPF does not seem to interfere with the antigenicity or the allergenicity of the protein. The recombinant production of allergens should make it possible to obtain well defined and homogeneous molecules which will allow a more detailed immunological characterization of these proteins. Moreover, the recombinant allergens could permit an easy standardization for accurate diagnosis, as well as the development of highly specific anti-allergic treatments, instead of the crude allergenic extracts currently used. Finally, the employed expression strategy may be of interest for the production of other recombinant 2S albumins.

This work was supported by grants PB89/087 and PB92/0195 from the Dirección General de Investigación Científica y Técnica (Ministerio de Educación y Ciencia, Spain). We are grateful to Dr C. López-Otín for oligonucleotide synthesis, and Dr J. G. Gavilanes for critical reading of the manuscript.

REFERENCES

1. Metcalfe, D. D. (1985) Food allergens, *Clin. Rev. Allergy* 3, 331–349.
2. Lemanske, R. F. & Taylor, S. L. (1987) Standardized extracts, foods, *Clin. Rev. Allergy* 5, 23–26.
3. Chapman, M. D. (1989) Purification of allergens, *Curr. Opin. Immunol.* 1, 647–653.
4. Huang, Q., Coleman, J. W. & Stanworth, D. R. (1985) Investigation of the allergenicity of β -lactoglobulin and its cleavage fragments, *Int. Arch. Allergy Appl. Immunol.* 78, 337–344.
5. Malik, Z., Bottomley, R. & Austen, B. (1988) Allergenic properties of the genetic variants A and B of bovine β -lactoglobulin, *Int. Arch. Allergy Appl. Immunol.* 86, 245–248.
6. Catterall, J. F., Stein, J. P., Kristo, P., Means, A. R. & O'Malley, B. W. (1980) Primary sequence of ovomucoid messenger RNA as determined from cloned complementary cDNA, *J. Cell Biol.* 87, 480–487.
7. Nisbet, A. D., Saundry, R. H., Moir, A. J. G., Fotherhill, L. A. & Fotherhill, J. E. (1981) The complete amino acid sequence of hen ovalbumin, *Eur. J. Biochem.* 115, 335–345.
8. Elsayed, S. & Bennich, H. (1975) The primary structure of allergen M from cod, *Scand. J. Immunol.* 4, 203–208.
9. Elsayed, S. & Apold, J. (1983) Immunological analysis of cod fish allergen M: locations of the immunoglobulin binding sites as demonstrated by the native and synthetic peptides, *Allergy* 38, 449–459.
10. Burks, A. W., Williams, L. W., Helm, R. M., Connaughton, C., Cockrell, G. & O'Brien, T. (1991) Identification of a major peanut allergen, *Ara h I*, in patients with atopic dermatitis and positive peanut challenges, *J. Allergy Clin. Immunol.* 88, 172–179.
11. Burks, A. W., Williams, L. W., Connaughton, C., Cockrell, G., O'Brien, T. & Helm, R. M. (1992) Identification and characterization of a second major peanut allergen, *Ara h II*, with use of the sera of patients with atopic dermatitis and positive peanut challenge, *J. Allergy Clin. Immunol.* 90, 962–969.
12. van Toorenbergen, A. W. & Dieges, P. H. (1987) Demonstration of spice-specific IgE in patients with suspected food allergies, *J. Allergy Clin. Immunol.* 79, 108–113.
13. Niinimäki, A. & Hannuksela, M. (1981) Immediate skin test reactions to spices, *Allergy* 36, 487–493.
14. Menéndez-Arias, L., Monsalve, R. I., Gavilanes, J. G. & Rodríguez, R. (1987) Molecular and spectroscopic characterization of a low molecular mass seed storage protein from yellow mustard (*Sinapis alba* L.), *Int. J. Biochem.* 10, 899–907.
15. Menéndez-Arias, L., Moneo, I., Domínguez, J. & Rodríguez, R. (1988) Primary structure of the major allergen of yellow mustard (*Sinapis alba* L.) seed, *Sin a I*, *Eur. J. Biochem.* 177, 159–166.
16. Menéndez-Arias, L., Domínguez, J., Moneo, I. & Rodríguez, R. (1990) Epitope mapping of the major allergen from yellow mustard seeds, *Sin a I*, *Mol. Immunol.* 27, 143–150.
17. Crouch, M. L., Tenbarger, K. M., Simon, A. E. & Ferl, R. (1983) cDNA clones from *Brassica napus* seed storage proteins: evidence from nucleotide sequence analysis that both subunits of napin are cleaved from a precursor polypeptide, *J. Mol. Appl. Genet.* 2, 273–283.
18. Terras, F. R. G., Torreken, S., Van Leuven, F., Osborn, R. W., Vanderleyden, J., Cammue, B. P. A. & Broekaert, W. I. (1993) A new family of basic cysteine-rich plant antifungal proteins from Brassicaceae species, *FEBS Lett.* 316, 233–240.
19. Scheiner, O. (1992) Recombinant allergens: biological, immunological and practical aspects, *Int. Arch. Allergy Immunol.* 98, 93–96.
20. Murray, H. G. & Thompson, W. F. (1980) Rapid isolation of high molecular mass DNA, *Nucleic Acids Res.* 8, 4321–4325.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, 2nd edn, Cold Spring Harbor Laboratory, Cold Spring Harbor NY.
22. Ericson, M. L., Rodin, J., Lenman, M., Glimelius, K., Josefsson, L.-G. & Rask, L. (1986) Structure of the rapeseed 1.7 S storage protein, napin, and its precursor, *J. Biol. Chem.* 261, 14576–14581.
23. Josefsson, L.-G., Lenman, M., Ericson, M. L. & Rask, L. (1987) Structure of a gene encoding the 1.7 S storage protein, napin, from *Brassica napus*, *J. Biol. Chem.* 262, 12196–12201.

24. Scofield, S. R. & Crouch, M. L. (1987) Nucleotide sequence of a member of the napin storage protein family from *Brassica napus*, *J. Biol. Chem.* 262, 12202–12208.
25. Baszczynski, C. L. & Fallis, L. (1990) Isolation and nucleotide sequence of a genomic clone encoding a new *Brassica napus* napin gene, *Plant Mol. Biol.* 14, 633–635.
26. Raynal, M., Depigny, D., Grellet, F. & Delseny, M. (1991) Characterization and evolution of napin-encoding genes in radish and related crucifers, *Gene (Amst.)* 99, 77–86.
27. Ericson, M. L., Muren, E., Gustavsson, H.-O., Josefsson, L.-G. & Rask, L. (1991) Analysis of the promoter region of napin genes from *Brassica napus* demonstrates binding of nuclear protein *in vitro* to a conserved sequence motif, *Eur. J. Biochem.* 197, 741–746.
28. Scharf, S., Horn, G. T. & Erlich, H. A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences, *Science* 233, 1076–1078.
29. Gibson, T. J. (1984) *Studies on the Epstein-Barr virus genome*, Ph. D. Thesis, Cambridge University, England.
30. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors, *Proc. Natl Acad. Sci. USA* 74, 5463–5467.
31. Smith, D. B. & Johnson, K. S. (1988) Single step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase, *Gene (Amst.)* 67, 31–40.
32. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids, *J. Mol. Biol.* 166, 557–580.
33. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature* 227, 680–685.
34. Batanero, E., Villalba, M. & Rodríguez, R. (1994) Glycosylation site of the major allergens from olive tree pollen. Allergenic implications of the carbohydrate moiety, *Mol. Immunol.* 31, 31–37.
35. Monsalve, R. I., González de la Peña, M. A., Menéndez-Arias, L., López-Otín, C., Villalba, M. & Rodríguez, R. (1993) Characterization of a new oriental-mustard (*Brassica juncea*) allergen, *Br j I*: detection of an allergenic epitope, *Biochem. J.* 293, 625–632.
36. Monsalve, R. I., Villalba, M., López-Otín, C. & Rodríguez, R. (1991) Structural analysis of the small chain of the 2S albumin, napin III, from rapeseed, *Biochim. Biophys. Acta* 1078, 265–272.
37. González de la Peña, M. A., Villalba, M., García-López, J. L. & Rodríguez, R. (1993) Cloning and expression of the major allergen from yellow mustard seeds, *Sin a I*, *Biochem. Biophys. Res. Commun.* 190, 648–653.
38. Krebbers, E., Herdies, L., De Clercq, A., Seurinck, J., Leemans, J., Van Damme, J., Segura, M., Gheysen, G., Van Montagu, M. & Vandekerckhove, J. (1988) Determination of the processing sites of an *Arabidopsis* 2S albumin and characterization of the complete gene family, *Plant Physiol. (Bethesda)* 87, 859–866.
39. D'Hondt, K., Bosch, D. & Van Damme, J. (1993) An aspartic proteinase present in seeds cleaves *Arabidopsis* 2S albumin precursors *in vitro*, *J. Biol. Chem.* 268, 20884–20891.
40. Marston, F. A. O. (1986) The purification of eukaryotic polypeptides synthesized in *Escherichia coli*, *Biochem. J.* 240, 1–12.
41. Wilkinson, D. L. & Harrison, R. G. (1991) Predicting the solubility of recombinant proteins in *Escherichia coli*, *BioTechnology* 9, 443–448.
42. Pal, M. & Biswas, B. B. (1995) Expression of the *Arabidopsis thaliana* 2S albumin gene 3 in *Saccharomyces cerevisiae*, *Gene (Amst.)* 153, 175–178.
43. Domínguez, J., Cuevas, M., Ureña, V., Muñoz, T. & Moneo, I. (1990) Purification and characterization of an allergen of mustard seeds, *Ann. Allergy* 64, 352–357.