

Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.)

Alexander A. KORTT¹, J. Bruce CALDWELL¹, Glenn G. LILLEY¹ and Thomas J. V. HIGGINS²

¹ Division of Biomolecular Engineering, Parkville, Melbourne, Australia

² Division of Plant Industry, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia

(Received February 19/June 11, 1990) — EJB 90 0178

The amino acid sequence of a methionine-rich 2S seed protein from sunflower (*Helianthus annuus* L.) and the sequence of a cDNA clone which codes for the entire primary translation product have been determined. The mature protein consists of a single polypeptide chain of 103 amino acids (molecular mass 12133 Da) which contains 16 residues of methionine and 8 residues of cysteine. The cDNA sequence established that the protein is synthesized as a precursor of 141 residues with a typical hydrophobic signal sequence of 25 residues followed by a further 13-residue hydrophobic pro-sequence which is presumably removed by post-translational cleavage. The sequence of the mature protein and that deduced from the cDNA were identical with no evidence of processing at the C-terminus. Comparison of the sunflower methionine-rich protein sequence with sequences of other seed 2S proteins from dicotyledons and monocotyledons showed limited but distinct sequence similarities; in particular the arrangement of the cysteine residues was conserved. The sunflower protein shows 34% identity with the methionine-rich Brazil nut 2S protein and the prepro regions of the precursors of these two proteins show about 50% identity. This similarity indicates that these methionine-rich 2S proteins have diverged as a subclass of the 2S superfamily of proteins which contain only 2–3% methionine. While the related 2S proteins from other dicotyledons are processed to a small and large subunit, the sunflower protein is not cleaved in this way.

The seed storage proteins of most dicotyledonous plants contain two major protein classes, globulins and albumins, which are distinguished on the basis of solubility [1]. These proteins are expressed during seed development as precursors which undergo co- and post-translational modification prior to deposition in protein bodies [2]. The globulins, which are soluble in high-salt buffers and have sedimentation coefficients of 7S and 11S, are characterized by high levels of arginine, glutamine and asparagine, which provide a source of nitrogen for the developing seedling [2].

The albumins are a more diverse group of proteins, usually soluble in water, with sedimentation coefficients of approximately 2S and some are rich in cysteine. While some 2S proteins appear to have only a storage function, some others such as lectins and protease and α -amylase inhibitors have biological activity. As a consequence of this these proteins are nutritionally undesirable. The sequences of a number of 2S seed storage proteins from different plants have shown that some of these proteins share sequence similarities and they have been grouped into a 2S superfamily of related sulfur-

rich seed proteins [3]. The 2S proteins from castor bean [4], rapeseed [5], mustard seed [6] Brazil nut [7] and lupin seed [8] have the same structural organization and consist of two subunits of different sizes linked by a disulfide bridge. The number and location of the cysteine residues in the sequence are conserved. Sequence analysis of cDNA clones coding for the rapeseed [5, 9] and Brazil nut [10] 2S proteins has shown that these proteins are synthesized as larger single-chain precursors which undergo post-translational modification to yield the two-chain structures of the mature proteins. The 2S protein of Brazil nut is of particular interest as it is a methionine-rich protein with 18% of residues methionine. This class of storage protein is particularly important from the point of view of animal nutrition because it contains higher levels of sulfur-containing amino acids than do the other major seed-storage proteins.

The sunflower (*Helianthus annuus* L.) 2S protein fraction contains at least eight distinct proteins and two of these are sulfur-rich with 8% of residues cysteine and 16% methionine [11]. In contrast to the 2S proteins of castor bean, rapeseed, mustard seed, Brazil nut and lupin seed, the sunflower 2S proteins are each composed of a single polypeptide chain. It has been shown that the sunflower 2S proteins are resistant to degradation by rumen bacteria [12]. This indicates that these sulfur-rich proteins may be ideal candidates for improving the nutritive content, with respect to the sulfur amino acids, of the seed and vegetative tissue of plants destined for ruminant feeding.

The elucidation of the structure of the methionine-rich 2S protein of sunflower was undertaken in order to isolate the

Correspondence to A. A. Kortt, Commonwealth Scientific and Industrial Research Organisation, Division of Biomolecular Engineering, Parkville Laboratory, 343 Royal Parade, Parkville, Victoria 3052, Australia

Abbreviation: SFA8, sunflower (methionine-rich) albumin 8.

Enzymes. Chymotrypsin (EC 3.4.21.2); trypsin (EC 3.4.21.4); *Staphylococcus aureus* V8 protease (EC 3.4.21.19).

Note. The novel nucleotide sequence data published here have been deposited with the EMBL sequence data bank and are available under the accession number X56686. The novel amino acid sequence data have been deposited with the EMBL sequence data bank.

gene for this protein and to exploit its potential use for the improvement of nutritive value of pasture plants by genetic engineering. It was also of interest to establish the relationship of this sunflower protein to the two-chain 2S proteins of other seeds, in particular the methionine-rich protein of Brazil nut.

In this paper we describe the complete amino acid sequence of sunflower albumin 8 (SFA8), the major methionine-rich 2S protein of sunflower seed and the isolation and sequence of a cDNA clone which codes for this protein. The relationship of the sunflower protein to the other 2S seed storage proteins is discussed, with particular reference to the methionine-rich protein from Brazil nut.

MATERIALS AND METHODS

Preparation of SFA8 and amino acid sequence analysis

The major sulfur-rich albumin (SFA8) was isolated from the sunflower albumin fraction by reversed-phase HPLC as described [11] and 1 mg protein was reduced and alkylated with iodoacetic acid as in [13]. The *S*-carboxymethylated protein was dissolved in 0.1 M ammonium bicarbonate pH 8.0 and aliquots (≈ 12 nmol) were digested with trypsin (Worthington), α -chymotrypsin (Worthington) or *Staphylococcus aureus* V8 protease (Pierce) at 37°C for 4 h at an enzyme/substrate ratio of 1:50 (by mass). The digests were dried in a Savant drier at 50°C, dissolved in 0.5 ml 0.1% (by vol.) trifluoroacetic acid and the soluble peptides separated by reversed-phase HPLC on a Vydac 218TP54 column (4.6 $\times$ 250 mm) at 45°C using a linear gradient of acetonitrile in 0.1% (by vol.) trifluoroacetic acid. Routinely, a gradient of 7–28% acetonitrile over 30 min was used for the separation of peptides. Peptides which co-eluted or were poorly resolved were purified by rechromatography using a shallower gradient or by isocratic elution.

Amino acid sequence analysis of the intact protein and peptides, identification of phenylthiohydantoin and amino acid analyses were performed as described [14].

cDNA library and molecular cloning

Total RNA was isolated [15] from mid-maturation (19-day post-anthesis) sunflower embryos and passaged through Sigmacell (Sigma) and mRNA was isolated using an oligo(dT)-cellulose column (Collaborative Research Inc., USA). Double-stranded cDNA was prepared from total mRNA using an Amersham kit (RPN 1285Y) and cDNA larger than 0.4 kb (size selected on a Pharmacia Superose 12B column) was cloned into λ phage gt11. Approximately 10^5 recombinant phage were screened with a mixture of the oligonucleotide probes and eight positive clones were identified and purified. Northern analysis was performed using a conventional formaldehyde gel system [16]. Probes for Northern analyses and for the identification of recombinant cDNA clones coding for SFA8 were prepared by end-labelling synthetic oligonucleotides [17]. Two synthetic oligonucleotide probes (5'-GG(A/G)CACAT(G/A)CA(T/C)TT(C/T)TC(G/A)TC-3' and 5'-GG(T/C)TC(A/G)TT(A/G/T/C)AA(T/C)-CATCATCAT-3'), complementary to the sequences of residues 59–65 and 68–74 of SFA8 (see Fig. 2), were synthesized as the reverse complement for use in Northern analysis of mRNA. The hybridization conditions for the oligonucleotide probes were as described in [18] and hybridization temperatures were optimized by trial and error. Positive clones were isolated, purified and amplified to prepare DNA for sub-

cloning into pUC119 and M13-mp19 phage DNA. Template DNA was sequenced using the dideoxy-chain termination method of Sanger [19]. Primers were either standard universal primer or appropriate synthetic oligonucleotides.

The length of the mRNA coding for SFA8 and the sequence of the 5' region of the near full-length cDNA clone were determined by dideoxy sequencing [20] with primers 5'-GGAGGAGTACTCCTGCTGC-3' and 5'-ATTCAGTTCTACCTCGGCCATAGGG-3' (the reverse complement of bases 71–89 and 158–182 respectively). In separate experiments, the primers were labelled with [γ - ^{32} P]dATP and hybridised at 45°C to 5 μ g total poly(A)-rich mRNA. The oligonucleotides were extended at 50°C for 45 min with avian myoblastoma virus reverse transcriptase. The reaction was stopped with a formamide/dye mix and the products analysed in a polyacrylamide sequencing gel in the presence of 8 M urea.

Secondary structures

The circular dichroic spectrum of SFA8 was measured in a Dichrographe III (Jobin Yvon) using a 2-mm path-length cell containing a 0.009% aqueous solution of the albumin in water. The ratios of α -helix, β -structure and random coil were estimated by the method of Provencher and Glockner [21] and compared with those predicted from the amino acid sequence data [22].

RESULTS

Amino acid sequence of SFA8

The sequence determination showed that SFA8 contains 103 amino acid residues corresponding to a molecular mass of 12133 Da. Gas-phase sequencer analysis of intact SFA8 yielded the sequence to residue 26 and the complete amino acid sequence of SFA8 was established by sequencing the peptides derived from staphylococcal protease, tryptic and chymotryptic digests of the protein as shown in Fig. 1. The staphylococcal protease peptides, isolated by reversed-phase HPLC accounted for all the residues of the molecule. Dipeptide S3 was not isolated from the digest but incomplete cleavage of the Glu19-Met20 bond yielded a peptide containing both S3 and S4. Tryptic digestion of the *S*-carboxymethylated protein produced a number of non-specific cleavages at several methionine residues including Met53-Gln54, Met68-Met69, Met69-Met70, Met75-Trp76 and Met86-Ala87. Chymotrypsin also cleaved at some of the methionine residues (Fig. 1). The single tryptophan residue at position 76 was confirmed by spectral analysis of peptide S7 using a Hewlett Packard 1040A photodiode array detector as described [23].

The carboxyl-terminal amino acid was identified as a methionine residue from the sequence data and compositions of C-terminal overlapping peptides S8, T12 and C11 (Fig. 1). The amino acid compositions of the staphylococcal peptides S1–S8 (Table 1) were consistent with the sequence data and the sequence composition is in agreement with the amino acid composition of the protein (Table 1). The sequence analysis showed that the 16 methionine residues were distributed throughout the molecule with only a single cluster of three methionines (residues 68–70).

Cloning, identification and sequencing of cDNA of SFA8

Northern analysis of total seed RNA and purified mRNA with a mixture of ^{32} P-end-labelled oligonucleotide probes

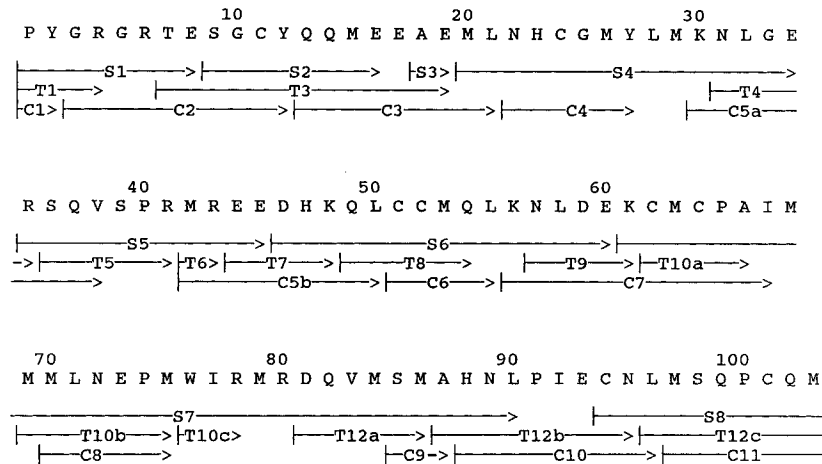


Fig. 1. Amino acid sequence and peptide alignments of sunflower methionine-rich albumin, SFA8. The vertical bar indicates the amino terminus of the peptide and the residues sequenced in each peptide are indicated by the solid line. Peptides are numbered from the N-terminus of the protein; S, T and C signify staphylococcal protease, tryptic and chymotryptic peptides, respectively

Table 1. Amino acid compositions of SFA8 and SFA8 staphylococcal protease peptides

Samples were hydrolysed in 6.0 M HCl for 24 h at 110°C. Values in parentheses are those expected from the sequence data. Tryptophan was identified by spectral analysis of peptide S7 and as its phenylthiohydantoin during sequencing

Amino acid	Amount in								sequence analysis
	S2	S2	S4	S5	S6	S7	S8	SFA8	
	mol/mol peptide								
Lys			0.9 (1)		2.0 (2)	0.7 (1)		4.1	4
His			0.9 (1)		1.2 (1)	0.7 (1)		3.0	3
Arg	2.6 (2)			3.7 (3)		1.5 (2)		6.9	7
Asp			2.0 (2)		3.0 (3)	2.4 (3)	0.9 (1)	8.9	3
Asn									6
Thr	0.9 (1)							1.0	1
Ser		1.0 (1)		1.7 (2)		1.1 (1)	0.8 (1)	4.7	5
Glu	1.1 (1)	4.6 (4)	1.2 (1)	3.0 (3)	3.8 (3)	3.1 (3)	2.0 (2)	19.9	10
Gln									8
Pro	1.0 (1)			1.0 (1)		2.1 (3)	0.8 (1)	5.8	6
Gly	2.0 (2)	1.0 (1)	2.7 (2)					5.0	5
Ala						1.7 (2)		3.1	3
Cys		1.0 (1)	1.1 (1)		2.1 (2)	1.7 (2)	1.7 (2)	6.9	8
Val				1.0 (1)		0.7 (1)		2.0	2
Met		0.9 (1)	3.0 (3)	0.9 (1)	0.9 (1)	5.4 (8)	1.7 (2)	15.5	16
Ile						2.5 (3)		3.0	3
Leu			2.6 (3)		3.0 (3)	2.0 (2)	1.0 (1)	9.0	9
Tyr	1.0 (1)	1.0 (1)	1.0 (1)					3.0	3
Phe								0	0
Trp						(1)		1.0	1
Total	8	9	15	11	15	33	10		103

showed a dominant component (0.7 kb) which hybridized strongly at 40°C (data not shown). The ribosomal RNAs (24S and 18S) also showed weak hybridization to the probes at 40°C but this was eliminated at higher hybridization stringency. Preparations of mRNA, from which the cDNA library was constructed, only showed the 0.7-kb component and ribosomal RNAs were not observed under any stringency conditions.

An *Eco*RI insert (0.56 kb) from one of the positive cDNA clones (λ SF16) was subcloned into pUC119 and M13-mp19 in both orientations, and sequenced. The deduced amino acid

sequence from this clone represented the entire sequence of mature SFA8. The 0.56-kb insert, SF16, was labelled by incorporation of [^{32} P]dATP and used to screen the λ library. Several larger cDNA clones were identified and cloned into M13 and shown to contain the entire translated region of the SFA8 cDNA (Fig. 2). cDNA clone pSF8 consisted of 600 bases 5' to and including the poly(A) tail. An open reading frame of 423 bases starting at base number 44 (Fig. 2) coding for a protein of 141 amino acids was observed.

To provide more sequence 5' to the initiating Met and to obtain a better estimate of the length of the mRNA for SFA8,

1 n nnn nnn nnn nnn nca tac cta tta cat caa aca cat tct aca

44 ⁻³⁸ M A R F S I V F ⁻³⁰ A A A G V L L L
ATG GCA AGG TTT TCG ATC GTG TTT GCA GCA GCA GGA GTA CTC CTC CTG

92 ⁻²⁰ V A M A P V S E A ⁻¹⁰ S T T T I I T
GTA GCC ATG GCG CCA GTT TCT GAG GCT TCC ACC ACA ACC ATC ATC ACC

140 T I I E E N P Y G R G R T E S 10 G
ACC ATC ATA GAG GAG AAC CCC TAT GGC CGA GGT AGA ACT GAA TCC GGA

188 C Y Q Q M E E A E M L N H C G M
TGC TAT CAG CAG ATG GAG GAG GCG GAG ATG CTC AAC CAC TGC GGG ATG

236 ³⁰ Y L M K N L G E R S Q V S P R M
TAC CTC ATG AAA AAT TTA GGA GAA AGG AGC CAA GTT AGT CCC AGA ATG

284 ⁵⁰ R E E D H K Q L C C M Q L K N L
AGA GAA GAG GAT CAC AAG CAA CTT TGC TGC ATG CAA CTG AAG AAC CTA

332 ⁶⁰ D E K C M C P A I M M 70 M L N E P
GAT GAG AAG TGC ATG TGT CCG GCA ATC ATG ATG ATG TTG AAC GAA CCA

380 ⁸⁰ M W I R M R D Q V M S M A H N L
ATG TGG ATA AGG ATG CGA GAT CAA GTG ATG AGC ATG GCT CAT AAC CTG

428 ¹⁰⁰ P I E C N L M S Q P C Q M *
CCC ATA GAA TGC AAT CTC ATG TCC CAA CCA TGC CAA ATG taa acc agc

476 tgg ttc taa act gca tga cta gct aag tcc cct gtg gaa taa aac gag

524 att cac tat cac atc gtc gtg tgt tgt tgt ttc aat gta ata tcg tcc

572 atc gtg tgt ctc gtt att aat aaa atc gga ctt gtc ttt tta taa aaa

620 aaa aaa aaa a

Fig. 2. Nucleotide sequence and deduced amino acid sequence derived from analysis of clone pSF8 and mRNA. The continuous underline denotes the signal sequence, the arrowhead indicates the predicted signal cleavage site, the amino acid sequence of mature SFA8 is shown in bold type, * shows the translation termination codon, n denotes the nucleotides not identified in the 5' region of the mRNA. The deduced protein sequence was identical to that determined by protein sequencing. The two left arrows show the regions to which the oligonucleotide probes were synthesized and the double underline indicates the putative polyadenylation signal sequence

sequencing experiments were performed on mRNA directly. The nucleotide sequence thus obtained agreed with that determined from cDNA clones (Fig. 2) with the exception of cytidine for thymidine changes at base positions 38 and 55. Although bases 1–14 could not be determined unequivocally the length of the mRNA was determined to be 629 bases (Fig. 2), in reasonable agreement with the length estimate from Northern analysis.

Secondary structure

Estimation of the secondary structure of SFA8 from its CD spectrum gave approximately 12% α -helix and 45% β -structure. Prediction of the secondary structure of SFA8 from its amino acid sequence, according to the method of Garnier et al. [22], predicted approximately 18% α -helix and 39% β -sheet in agreement with the values determined from the CD spectrum.

DISCUSSION

The protein and nucleotide sequence data show that SFA8 is synthesized as a precursor molecule of 141 residues which is cleaved probably twice to yield a mature single-chain molecule of 103 residues. The sequence encoded by the cDNA was identical to the directly determined amino acid sequence beginning at base 158 and terminating with a stop codon at base 467 (Fig. 2). In the SFA8 cDNA sequence there are two

possible methionine initiation codons 5' to the N-terminus of mature SFA8 at base positions 44 and 98. The ATG codon at base position 44 is preceded by an A at -3 and followed by a G at $+4$, consistent with Kozak's rule for an initiation codon [24] and is followed by a deduced sequence of 25 amino acids which has the characteristics of a signal sequence typical of proteins destined for transmembrane transport [25]. Of the 25 residues in the putative signal sequence 19 are hydrophobic, with a positive residue (Arg -36) before, and a negative residue (Glu -15) after the hydrophobic region (Fig. 2). This putative signal sequence is followed by a hydrophobic sequence of 13 residues (pro-region), containing a four-residue repeat of Thr-Thr-Ile-Ile, which must be removed during or after translation. There are no data to establish where in the prepro sequence the cleavage of the signal sequence occurs: however, the cleavage site may be terminated by the sequence Pro(-18)Val-Ser-Glu-Ala(-14) after which co-translational cleavage is predicted as highly probable [26]. At the site where cleavage presumably occurs to yield the N-terminus of mature SFA8 the residues at the C-terminus of the pro-region are identical to those in the Brazil nut precursor in this region (Fig. 3).

Although the precursor of SFA8 is synthesized in a prepro form as found for the precursors of the 2S proteins of rapeseed [8] and the Brazil nut [10] the processing of SFA8 is limited to the removal of the signal peptide and the pro-region with no further endo- or exoproteolytic cleavage of the polypeptide chain as found for other 2S proteins of this family [9, 10].

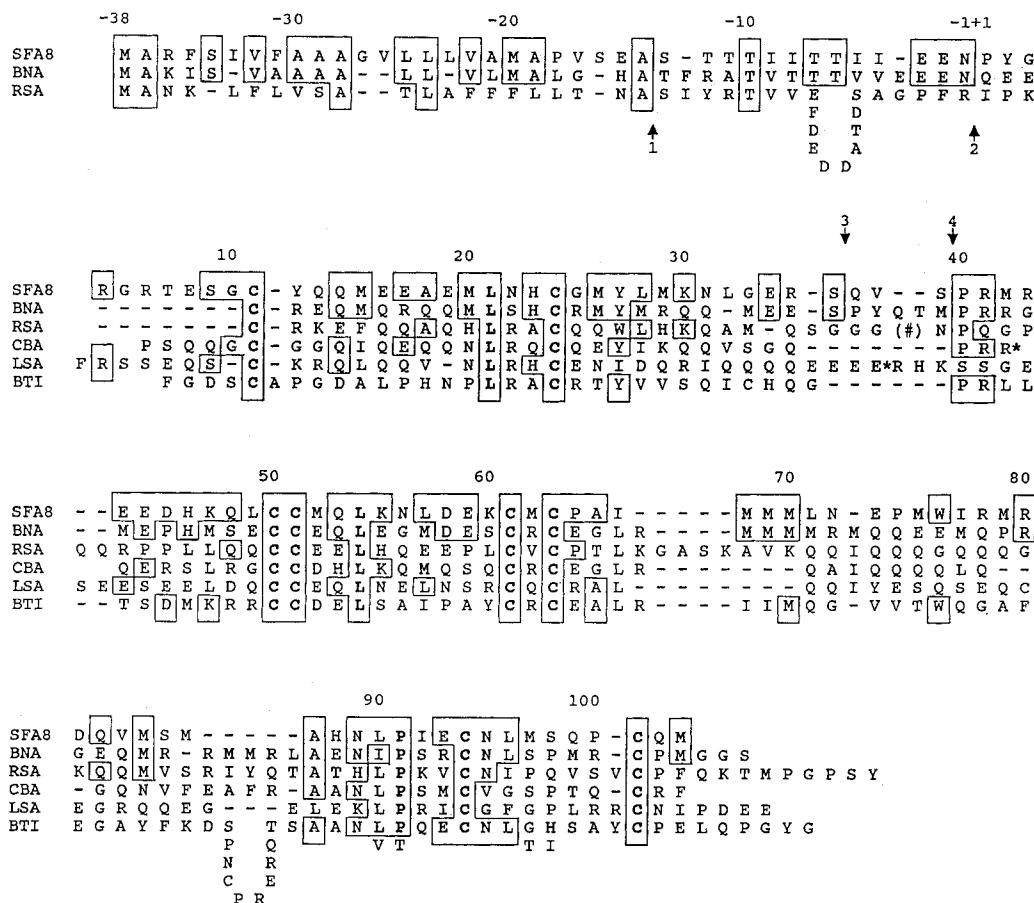


Fig. 3. Comparison of the amino acid sequences of mature SFA8 and related 2S seed proteins. BNA, Brazil nut 2S protein [7]; RSA, rapeseed 2S protein [5]; CBA, castor bean 2S protein [4]; LSA, lupin seed 2S protein [8]; BTI, barley trypsin inhibitor [28]. Conserved residues between SFA8 and the other proteins are boxed and the residues conserved in all six proteins are in bold-face type. The asterisk indicates the C-terminal residue of the small polypeptide chain of the two-chain 2S proteins. The sequences of the prepro-regions of the precursors of SFA8, BNA [33] and RSA [9] are also shown. Arrow 1 denotes the predicted signal cleavage site in the three precursor proteins and arrow 2 denotes the cleavage site adjacent to the N-terminus of the mature protein. The additional residues of the pro-region of rapeseed 2S protein have been accommodated in a loop in order to align the second cleavage site. The residues of BNA [33] and RSA [9] which are excised during processing are also shown (residues 35–40). Arrows 3 and 4 indicate the cleavage sites in the precursors of BNA and RSA which yield the C-terminal residue of the small chain and the N-terminal residue of the large chain of the mature molecules. (#) denotes 16 additional residues in this region of the RSA precursor which are not shown. Gaps (–) introduced to maximize similarity.

SFA8 is the only 2S protein isolated to date from a dicotyledon that is not a two-chain structure; however, related 2S proteins from monocotyledons such as the barley trypsin inhibitor [27] and the wheat α -amylase inhibitors [28, 29] are single-chain molecules.

Comparison of the amino acid sequence of SFA8 with the sequences of other members of the seed 2S protein superfamily [3] (Fig. 3) showed that SFA8 contained both the small and large polypeptide chains of the two-chain 2S proteins in a single polypeptide chain and that there was considerable sequence similarity between these proteins, particularly in the complete conservation of the cysteine residues. While SFA8 showed only 17–24% sequence similarity to the low-methionine 2S protein members of this family, it showed greater similarity to the methionine-rich Brazil nut 2S protein, with 34% identity (Fig. 3) and 8 of the 16 methionine positions being conserved. This indicates that these two methionine-rich proteins from such widely divergent species are more closely related to each other than SFA8 is to the other 2S seed proteins of this family.

Further evidence that the methionine-rich proteins of sunflower and Brazil nut are closely related comes from a com-

parison of their prepro-region sequences (Fig. 3). The signal peptides show a high degree of similarity with 52% identity as do the pro-regions (46% identity) (Fig. 3). However, the rapeseed 2S protein signal and pro-sequences show essentially no sequence similarity to those of the two methionine-rich proteins (Fig. 3). The monocotyledon members of the 2S superfamily of proteins, such as the barley trypsin inhibitor (121 residues) [27] and the wheat α -amylase inhibitor (124 residues) [29] are longer than SFA8. In the alignment shown in Fig. 3, the additional residues of the barley trypsin inhibitor are located in a loop in the C-terminal region of the molecule. These proteins show only 13–17% sequence similarity with SFA8 but the characteristic arrangement of cysteines and other conserved residues is retained (Fig. 3). In contrast, the methionine-rich proteins isolated from maize (*Zea mays* L.) seed [30, 31] show no sequence similarity to SFA8 or Brazil nut 2S protein and they do not contain the conserved cysteine residues; consequently, these methionine-rich proteins do not belong to the seed 2S protein superfamily [3].

Comparison of the sequence of SFA8 (residues 35–40) with the deduced precursor sequences of rapeseed [9] and Brazil nut [32] proteins, in the region where processing occurs

to yield the two-chain structure, is also shown in Fig. 3. The precursor of the Brazil nut 2S protein contains an additional two residues in this region compared to SFA8 and five residues are cleaved from the precursor to produce the mature two-chain protein [32]; the precursor of the rapeseed protein contains an extra 19 residues which are excised to yield the mature two-chain protein [9]. It is not clear why the SFA8 molecule is not processed like the other dicotyledon members of this family as the residues at which cleavage could occur in this region of SFA8 are similar to those found in Brazil nut 2S protein precursor (Fig. 3). The lack of internal processing in SFA8 may be attributed to either the absence of an appropriate processing protease in sunflower seed or to structural differences in this region of SFA8. The shorter sequence in SFA8 may preclude the formation of an exposed surface loop susceptible to proteolytic cleavage. Comparison of SFA8 with the barley trypsin inhibitor, also a single-chain molecule, shows that the latter also contains fewer residues in this region (Fig. 3).

SFA8 shows no tryptic, chymotryptic or α -amylase inhibitory activity (A. A. Kortt and J. B. Caldwell, unpublished data) although a putative reactive-site sequence (Pro40-Arg-Met42) similar to that of the barley trypsin inhibitor reactive-site sequence occurs in SFA8 (Fig. 3). The reactive-site sequence in the barley trypsin inhibitor (Pro40-Arg-Leu42 in Fig. 3) aligns near the internal cleavage site in the 2S protein precursor molecules and the Pro-Arg-Xaa motif is found in several of these molecules. The absence of inhibitory activity in these molecules has been attributed to the cleavage of the precursor at or near this potential inhibitory reactive site. Compared with barley trypsin inhibitor SFA8 contains three additional residues N-terminal to the putative reactive site sequence (Fig. 3) which may perturb the formation of a functional reactive site in SFA8. This observation suggests that modification of the SFA8 sequence in this region by the removal of the three additional residues, using site-directed mutagenesis, may produce a molecule with inhibitory activity. Experiments to test this proposal are in progress.

It is interesting that although the cysteine residues have been conserved and are involved in disulfide bridge formation [8], CD spectra of the proteins in this superfamily suggest distinct differences in secondary structure. Some, such as the lupinseed 2S protein [33] and the wheat 0.28 α -amylase inhibitor [28] are highly helical while SFA8 was found to have 45% β -structure with only about 12% α -helix, not unlike the barley trypsin inhibitor [29].

While some members of the 2S superfamily have identified biological activities (e.g. protease inhibitors, allergens) their main role is thought to be that of storage proteins as they are characterized by high contents of glutamine and arginine. These proteins with their relatively high cysteine contents may also function as stores of sulfur for the seed. Two members of this superfamily are unique in containing unusually high amounts of methionine making them ideal candidates for transfer and expression into food and forage plants in order to increase their nutritional value. The sunflower protein SFA8 is particularly suited for this role as it is a single polypeptide chain and is resistant to *in vitro* rumen digestion; expression of SFA8 in legumes for example would facilitate the improvement of nutritional quality of legume seeds which are deficient in sulfur-containing amino acids. The cloning of cDNA encoding the sulfur-rich protein represents a first step in altering the content of sulfur containing amino acids of seeds and vegetative parts of agriculturally important plants.

We thank Mr J. E. Burns for performing the manual sequencing experiments, Mr N. Bartone for the amino acid analyses and synthesis

of oligonucleotides, and Mr P. Strike for gas-phase sequencer analyses and phenylthiohydantoin analyses. We also thank Dr L. C. Gruen for analysis of the circular dichroic spectrum.

REFERENCES

- Derbyshire, E., Wright, D. J. & Boulter, D. (1976) *Phytochemistry* **15**, 3–24.
- Higgins, T. J. V. (1984) *Annu. Rev. Plant Physiol.* **35**, 191–221.
- Kreis, M., Forde, B. G., Rahman, S., Milfin, B. J. & Shewry, P. R. (1985) *J. Mol. Biol.* **183**, 499–502.
- Sharief, F. & Li, S. S.-L. (1982) *J. Biol. Chem.* **257**, 14573–14759.
- Ericson, M. L., Rodin, J., Lenman, M., Glimelius, K., Josefsson, L. G. & Rask, L. (1986) *J. Biol. Chem.* **261**, 14576–14581.
- Menendaz-Arias, L., Moneo, I., Dominguez, J. & Rodriguez, R. (1988) *Eur. J. Biochem.* **177**, 159–166.
- Ampe, C., Van Damme, J., de Castro, L. A. B., Sampaio, M. J. A. M., Van Montagu, M. & Vandekerckhove, J. (1986) *Eur. J. Biochem.* **159**, 597–604.
- Lilley, G. G. & Inglis, A. S. (1986) *FEBS Lett* **195**, 235–241.
- Scofield, S. R. & Crouch, M. L. (1987) *J. Biol. Chem.* **262**, 12202–12208.
- Sun, S. S. M., Altenbach, S. B. & Leung, F. W. (1987) *Eur. J. Biochem.* **162**, 477–482.
- Kortt, A. A. & Caldwell, J. B. (1990) *Phytochemistry* **29**, 2805–2810.
- Spencer, D., Higgins, T. J. V., Freer, M., Dove, H. & Coombe, J. (1988) *Br. J. Nutr.* **60**, 241–247.
- Kortt, A. A., Inglis, A. S., Fleming, A. I. & Appleby, C. A. (1988) *FEBS Lett.* **231**, 341–346.
- Kortt, A. A., Strike, P. M. & de Jersey, J. (1989) *Eur. J. Biochem.* **181**, 403–408.
- Chandler, P. M., Higgins, T. J. V., Randall, P. J. & Spencer, D. (1984) *Plant Physiol.* **71**, 47–54.
- Gonda, T. J., Sheiness, D. K. & Bishop, J. M. (1982) *Mol. Cell Biol.* **2**, 617–624.
- Richardson, C. C. (1971) *Proc. Nucleic Acids Res.* **2**, 815.
- Woods, D. (1984) *Focus*, vol. 6, no. 3, pp. 1–3, BRL Research Laboratories, Maryland, USA.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl Acad. Sci. USA* **74**, 5463–5467.
- Hudson, P., Haley, J., Cronk, M., Shine, J. & Niall, H. (1981) *Nature* **291**, 127–131.
- Provencher, S. W. & Glockner, J. (1981) *Biochemistry* **20**, 33–37.
- Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
- Gale, D. J., Logan, R. I. & Rivett, D. E. (1987) *Text. Res. J.* **57**, 539–542.
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
- Perlman, D. & Halvorson, H. O. (1983) *J. Mol. Biol.* **167**, 391–409.
- Folz, R. J. & Gordon, J. I. (1987) *Biochem. Biophys. Res. Commun.* **146**, 870–877.
- Odani, S., Koide, T., Ono, T. & Ohnishi, K. (1983) *Biochem. J.* **213**, 543–545.
- Kashlau, N. & Richardson, M. (1981) *Phytochemistry* **20**, 1781–1784.
- Maeda, K., Kakatayashi, S. & Matsubara, H. (1985) *Biochim. Biophys. Acta* **828**, 213–221.
- Kirihara, J. A., Petri, J. B. & Messing, J. (1988) *Gene* **71**, 359–370.
- Pederson, K., Argos, P., Naravana, S. V. L. & Larkins, B. A. (1986) *J. Biol. Chem.* **261**, 6279–6284.
- Altenbach, S. B., Pearson, K. W., Leung, F. W. & Sun, S. S.-M. (1987) *Plant Mol. Biol.* **8**, 239–250.
- Lilley, G. G. (1986) *J. Sci. Food Agric.* **37**, 895–907.