Amino acid sequence of conglutin δ, a sulfur-rich seed protein of *Lupinus angustifolius* L.

Sequence homology with the C-III α-amylase inhibitor from wheat

Glenn G. Lilley and Adam S. Inglis

CSIRO Division of Protein Chemistry, 343 Royal Parade, Parkville (Melbourne), 3052 Victoria, Australia

Received 29 October 1985

Complete amino acid sequences and disulfide cross-link arrangements have been determined for the two subunit polypeptides (M_r 9401 and 4597) of conglutin δ, a helix-rich seed protein from *Lupinus angustifolius* cv. Uniwhite. There are two intrachain disulfide bonds and a free sulfhydryl group within the large chain and two interchain disulfide bonds to the small chain. The sequences show regions enriched in glutamine-glutamic acid and serine residues which were correlated by a predictive method to the high measured level of α-helix (~38%). Homology was found between a cystine-rich region of conglutin δ and the C-III α-amylase inhibitor from wheat suggesting that these proteins originated from a common ancestral gene.

Legume storage protein *(Lupinseed)* *Disulfide cross-link* *Free sulfhydryl group* *Sequence homology* *α-Amylase* *Enzyme inhibitor*

1. INTRODUCTION

Acceleration in our understanding of the structure and expression of genes, which control protein composition and synthesis in grain crops, may soon allow the sulfur-amino acid deficiency problem in legumes [1] to be solved by direct manipulation of the appropriate genes. A prerequisite for identification of genes suitable for manipulation is the need to identify and characterize the major storage proteins expressed. In so doing, a basis for the changes to be introduced by selection or by engineering of genes may be established.

Gillespie and Blagrove [2] recognized that conglutin γ, a minor lupinseed globulin which contains relatively high levels of 1/2 cystine, methionine and other essential amino acids, had the potential to improve the nutritional quality of lupinseed. However, the importance of the 2S sulfur-rich lupinseed protein, conglutin δ, was not recognized [3–5] in spite of its high level of cystine (~8 res.%). Recent studies have described the isolation [6] and physicochemical characterization [7] of conglutin δ, which represents 20–30% of the globulin fraction of *Lupinus angustifolius* cv. Uniwhite seed and contains about 70% of the seed sulfur. It is a small helical protein (subunit M_r 14000, α-helix ~38%) quite unlike the major 7–11S legume seed storage proteins. The aim of this study was to establish the primary structures, including disulfide cross-links, of the subunit polypeptides of conglutin δ. The relationship of conglutin δ to the proposed [8] superfamily of seed proteins, which includes the prolamins of the monocotyledons and 2S globulins of the dicotyledons, is explored.

2. EXPERIMENTAL

Conglutin δ_2 (the major protomeric form responsible for the 2S classification of conglutin δ) was isolated from mature seeds of *L. angustifolius* cv. Uniwhite and reduced and *S*-alkylated as described [6]. SCM subunit-polypeptide chains were separated by anion-exchange chromatog-

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

raphy on DEAE-Sephacel or a Waters DE-5PW HPLC column and purified by reversed phase (RP) HPLC on a Vydac 218TP54 column. SCM-polypeptides were judged to be homogeneous by the absence of contaminating bands on SDS-PAGE and by the presence of a single N-terminal amino acid for each purified chain.

Separate aliquots of SCM-polypeptide chains were fragmented in 0.05 M ammonium bicarbonate buffer at 37°C with one of the following enzymes at an enzyme:substrate ratio of 1:50 (w/w): *Armillaria mellea* protease (a generous gift from the Pharmaceuticals Division, ICI, England); trypsin (Worthington); chymotrypsin (Worthington); *Staphylococcus aureus* V8 protease (Pierce). Digestion time was 4 h except for *A. mellea* protease which was ≧24 h. Digests were dried under vacuum at 50°C, dissolved in 0.1% (v/v) trifluoroacetic acid (TFA) containing up to 10% (v/v) acetonitrile (CH_3CN) and fractionated by RP-HPLC on a Vydac 218TP54 column in 0.1% (w/v) TFA with gradient of CH_3CN.

The principal method of peptide sequencing was by a modified [9] manual Edman technique [10] although some peptides were sequenced with a modified Edman sequenator [11] and a gas-phase sequenator (Applied Biosystems, USA) [12]. PTH-amino acids were identified after separation by RP-HPLC on a Zorbax Gold C18 column (Du Pont). Peptides were hydrolysed under vacuum at 108°C in 6 N HCl (24 h) and analysed for amino acids using a modified Beckman 120C analyser or a Waters amino acid analyser.

To locate disulfide cross-links and free sulfhydryl groups, protein was extracted [6] in the presence of *N*-ethylmaleimide (NEM) and digested in 0.05 M ammonium acetate/acetic acid (pH 7) with trypsin (37°C) for 3 h at an enzyme:substrate ratio, 1:25 (w/w). A small portion of this digest was oxidized at 0°C (0.5 min) with ice-cold performic acid, diluted with ice-cold water and freeze-dried. After fractionation (by RP-HPLC) of a sample of digest before and after oxidation, elution profiles were compared. Peaks in the digest which were not identified in the oxidized sample were assumed to be peptides containing disulfide links. These linked peptides were purified, partially sequenced and the amino acid compositions determined. They were also oxidized and the component peptides purified and sequenced. To identify NEM-blocked sulfhydryl groups, PTH-Cys-NEM was prepared from NEM-cysteine [10] for use as a standard in the PTH analysis.

3. RESULTS AND DISCUSSION

The large and small subunit-polypeptide chains of conglutin δ_2 (SCM form) were fractionated by anion-exchange HPLC and each size class showed a family of chains (fig.1). Polypeptide fractions 3 and 9, representing the major small and large chains, were sequenced entirely and their sequences are shown in fig.2a,b. The less abundant small polypeptide (fraction 4) lacked the Phe^1-Arg^2- N-terminus but was otherwise identical. N-terminal sequences (to residue 8) of the less abundant large chain fractions 7, 8 and 10 were identical although fraction 10 lacked the N-terminal Arg^1. Amino acid analysis and elution profiles of tryptic digests (not shown) suggested only minor substitutions elsewhere. The differences found in both families of chains suggested different degrees of post-translational processing of proteins that are the result of the expression of a multi-gene family. However, post-extraction proteolysis, although not otherwise evident, cannot be ruled out.

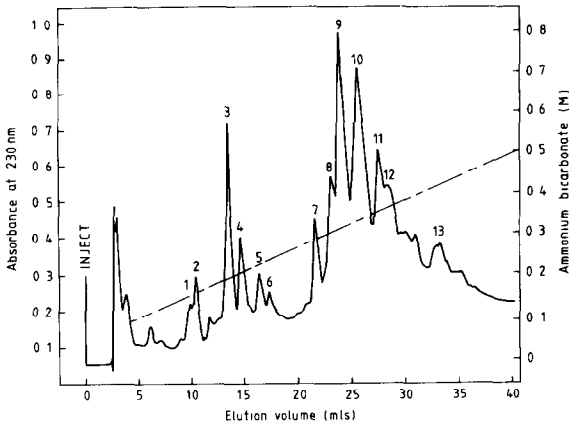

Fig.1. Chromatography of SCM-conglutin δ_2 subunit polypeptides (2.5 mg) on a Waters DE-5PW anion-exchange HPLC column (10 × 0.4 cm) equilibrated in 0.1 M ammonium bicarbonate at 1.0 ml·min^{-1}. (– – –) Ammonium bicarbonate gradient. Fractions 1–6 and 7–13 are the small and large subunit-polypeptide families, respectively.

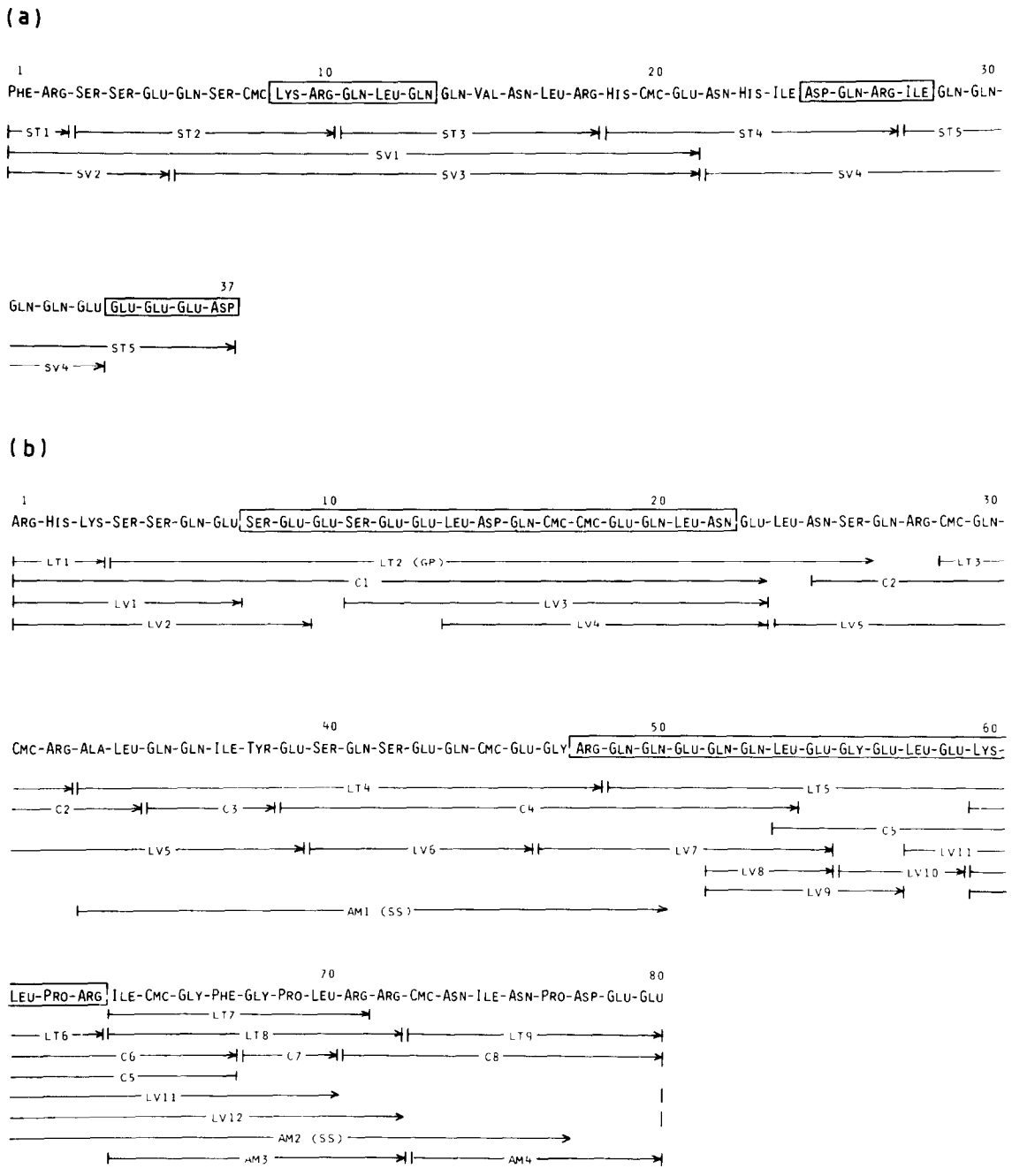

Fig.2. Primary structure of the major (a) small polypeptide chain and (b) large polypeptide chain of conglutin δ_2. Residues sequenced are indicated by the arrows; ST and LT, tryptic peptides from small and large chains, respectively; C, chymotryptic peptides; SV and LV, *Staphylococcus aureus* V8 peptides from the small and large chains, respectively; AM, *Armillaria mellea* protease peptides; GP, sequence from gas-phase sequenator; SS, sequence from modified Edman sequenator (stationary stainless-steel cup). Other peptides were sequenced manually. Boxed sequences represent regions of predicted α-helix.

Table 1

Amino acid composition of *L. angustifolius* cv. Uniwhite conglutin δ_2 and constituent subunit-polypeptide chains

Amino acid	Large polypeptide		Small polypeptide		Conglutin δ_2 analysis
	Analysis	Sequence	Analysis	Sequence	
Lys	1.9	2	1.0	1	3.0
His	0.8	1	1.9	2	2.8
Arg	6.3	7	3.6	4	10.7
Cmc[a]	7.0	7	1.9	2	9.4
Met	0.0	0	0.0	0	tr
Thr	0.0	0	0.0	0	tr
Asx	6.1	6	3.7	4	10.5
Glx	28.6	29	14.5	15	43.8
Ser[a]	6.9	7	3.0	3	9.4
Pro	2.8	3	0.0	0	4.3
Gly	4.1	4	0.0	0	4.3
Ala	1.1	1	0.0	0	1.4
Val	0.0	0	1.0	1	1.4
Leu	8.2	8	2.1	2	10.5
Ile	2.9	3	1.9	2	5.0
Tyr	0.9	1	0.0	0	1.4
Phe	1.0	1	1.0	1	1.7
Total	78.6	80	35.6	37	120.0

[a] Values obtained by extrapolation to zero hydrolysis time

Data are expressed in residues/molecule

The major small polypeptide chain was sequenced by manual methods. A combination of tryptic and *S. aureus* peptides provided the necessary overlaps (see fig.2a). The identity of the C-terminus of the small polypeptide was inferred from the C-terminus of tryptic peptide ST5 and confirmed by the agreement of the summed compositions of each tryptic peptide with the composition of the intact chain (table 1); carboxypeptidase Y did not release amino acids from this molecule.

The sequence of the large subunit polypeptide chain (fig.2b) was tentatively assigned from the alignment of a complete set of tryptic and chymotryptic peptides except for the overlaps of C1–C2 and LV4–LV5 which required a major effort to prove. The large peptide, LT2, was both difficult to prepare and to sequence manually. Furthermore, cleavages after the leucines at positions 14, 21 and 24 with chymotrypsin were not achieved, also giving large chymotryptic fragments in this sequence region. Sequenator analysis yielded the sequence Arg^1-His^2-Lys^3-Ser^4-Ser^5-Gln^6-- and located one of the 2 lysine residues. *A. mellea* protease, which cleaves before lysine [13], was used to produce large peptides. The peptide beginning Lys^3 from the N-terminus was not recovered although peptide AM2 (from the C-terminus) was recovered and partly sequenced in the automatic sequenator [11]. Extended *A. mella* protease digestion (36 h at 37°C) gave the peptides AM1, AM3 and AM4, their sequences suggesting a residual trypsin-like activity [13] under the conditions used. The gas-phase sequenator was used successfully to extend the sequence of peptide LT2 and overlapped peptides C1–C2 and LV4–LV5. The sequences of C1 and LT2 in this overlap region were strongly supported by the sequences of peptide LV3, LV4 and LV5. Other peptides from *S. aureus* V8 protease confirmed the overall sequence. Compositions for LT9, C8 and AM4, together with the

release of Glu and Asp by carboxypeptidase Y digestion of the intact chain, confirmed Glu^{80} as the large chain C-terminus.

An examination of the RP-HPLC elution profiles (fig.3) from a digest of NEM-conglutin δ_2 before and after oxidation with performic acid showed 2 major peaks in the pre-oxidation control

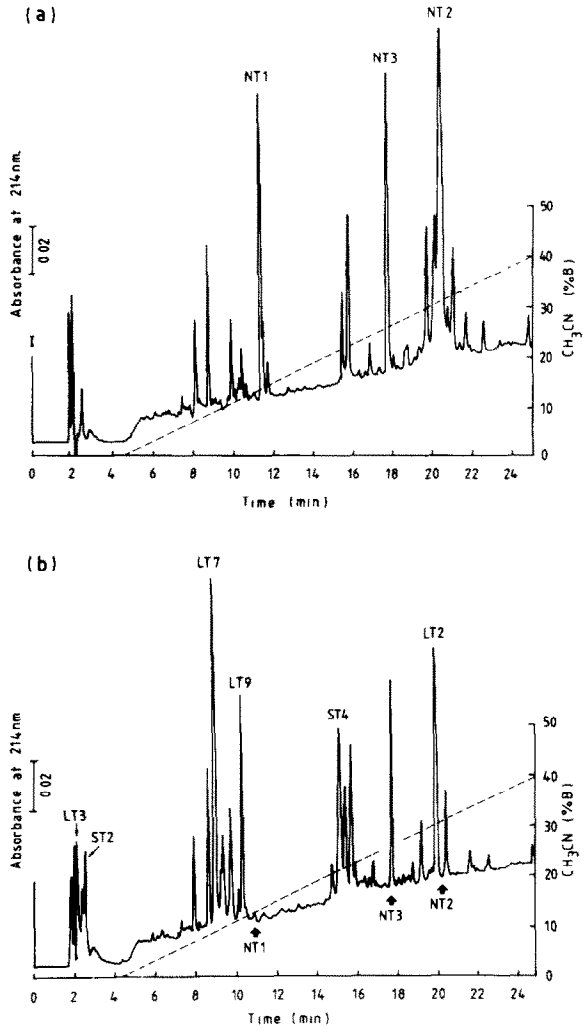

Fig.3. RP-HPLC profiles of (a) tryptic digest of native (N) NEM-treated conglutin δ_2 and (b) the same tryptic digest after oxidation with performic acid. The heavy arrows in (b) show the pre-oxidation elution positions of peptides NT1–3 for comparison. The Vydac 218TP54 column was equilibrated in 0.1% (v/v) TFA (buffer A) at 45°C and peptides eluted with a gradient of 0.1% (v/v) TFA-70% (v/v) CH_3CN (buffer B).

which were absent following oxidation (peptides NT1 and NT2, respectively). In contrast, tryptic peptide NT3 identified with the cystine-containing peptide LT4 from the large subunit polypeptide because of absorption at 280 nm (from tryptophan) was not affected by oxidation. Sequencing of NT3 confirmed the presence of NEM blocked cysteine at residue 45 showing that LT4 was not involved in cross-linking. Peptides NT1 and NT2 (fig.3a) were oxidized and fragments separated by RP-HPLC. Peptide NT1 was composed of peptides LT3, LT9 and ST2 (fig.3b), named thus because of their correspondence to the large subunit polypeptides LT3, LT9 (fig.2b) and peptide ST2 (fig.2a). Peptide NT2 was also composed of 3 polypeptides: LT2, LT7 and ST4. Although the residues involved in disulfide bonding were established, in both cases there remained the problem of the orientation. Peptides LT2 and LT3 contained -Cys-Cys- and -Cys-Gln-Cys- sequences, respectively, which could not be cleaved enzymatically. Attempts to identify the arrangement by digestion of conglutin δ with other enzymes such as pepsin did not resolve the orientation problem.

The orientation of the cross-links between ST2, LT9 and LT3 was established by a novel, non-enzymic approach. After one degradation cycle of

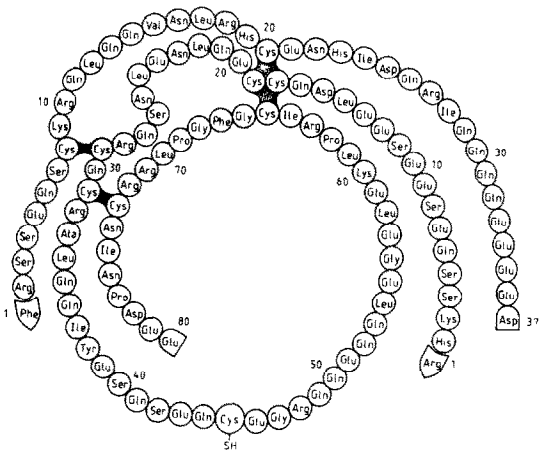

Fig.4. Schematic representation of the assembled subunit of conglutin δ_2. Disulfide linkages determined unequivocally are shown as solid regions; cross-links not unequivocally assigned are shown shaded. Circular symbols represent the large chain and oval symbols are the small chain.

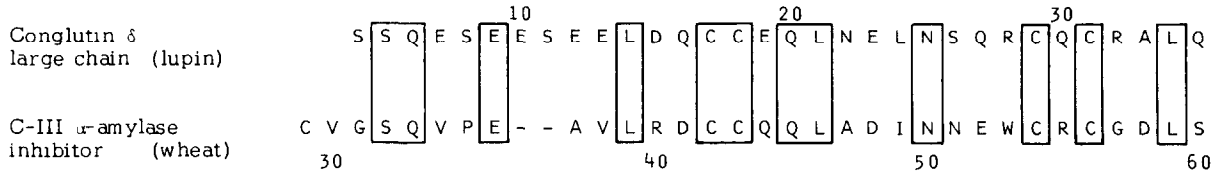

Fig.5. Sequence homology between conglutin δ (large chain) and the major C-III α-amylase inhibitor from wheat. Residues enclosed by the boxes are identical.

NT1 by the Edman procedure [9], the products (ST2 and LT9) were isolated by RP-HPLC and identified by N-terminal sequencing. The second degradation of ST2 showed only serine and established the orientation shown in the sequence schematic (fig.4); the alternative orientation would have given glutamine with serine in the second step from ST2. Peptide NT2 was not amenable to analysis by this procedure; thus the exact orientation remains unknown.

The major large chain has a sequence M_r of 9801 (SCM form), in good agreement with the value obtained [7] in the ultracentrifuge (M_r 9600). That earlier physiochemical study also suggested a 1:1 stoichiometry for the large and small polypeptides in the conglutin δ subunit, an assumption that has been validated by the cross-linking study above. The sum of the sequence M_r values (less the contribution of the SCM groups) for each major polypeptide (M_r 9401 and 4597) gave an M_r of 13998, in close agreement with the value obtained in the ultracentrifuge [7] for the intact conglutin δ_2 subunit (M_r 14000). The interpretation of stoichiometry of the conglutin δ_2 subunit is therefore confirmed. Furthermore, the compositions determined for each polypeptide by sequencing were in close agreement with their respective amino acid analyses and with the analysis for the intact subunit (table 1) for a 1:1 stoichiometry.

Previous spectral studies showed [7] that conglutin δ_2 contained a significant proportion (~38%) of highly stable α-helix with stability directly dependent on the presence of the disulfide bridges. The level of helix predicted [14] from the amino acid sequence agrees closely with the measured value. The predicted helical domains are positioned (fig.2) in 2 regions of the large chain and 3 shorter regions of the small chain.

A sequence region of the conglutin δ large subunit polypeptide is homologous (fig.5) to the 0.28 α-amylase inhibitor (iso-inhibitor C-III) from wheat, described by Kashlan and Richardson [15]. Furthermore, wheat α-amylase inhibitors 0.19, 0.28 and 0.39 (designated thus from their mobility relative to bromophenol blue on gel electrophoresis) are likewise helix-rich and stable to both thermal and chemical denaturation [16,17]. However, conglutin δ does not show inhibitory action towards porcine α-amylase although activity against insect α-amylases [18] cannot be ruled out. Nevertheless, it seems likely that the structural gene(s) coding for conglutin δ have evolved from the ancestral gene(s) which gave rise to other seed protease and α-amylase inhibitors, the 2S seed storage globulins (such as those of castor bean and oilseed rape) and the non-repetitive domains of some prolamins in grass seeds [8]. Whereas the level of glutamine in the prolamins has been raised by the more recent (in an evolutionary sense) insertion of sequence repeats rich in proline, glutamine and aromatic amino acids [8], the high levels of glutamine in conglutin δ correspond to regions also rich in serine.

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to the Analytical Services group at the CSIRO Division of Protein Chemistry (Mr N. Bartone, Mr P. Sommers and Mr M. Rubira for PTH and amino acid analyses). This report forms part of PhD studies undertaken by G.G.L. whilst enrolled at Deakin University (Australia).

REFERENCES

[1] Kelly, J.F. in: Nutritional Improvement of Food Legumes by Breeding, pp.179–184, Protein Advisory Group of the United Nations System, New York.

[2] Gillespie, J.M. and Blagrove, R.J. (1974) Search 5, 600–601.
[3] Blagrove, R.J. and Gillespie, J.M. (1975) Aust. J. Plant Physiol. 2, 13–27.
[4] Gerritsen, T.H. (1956) Biochim. Biophys. Acta 22, 269–273.
[5] Duranti, M., Restani, P., Poniatowska, M. and Cerletti, P. (1981) Phytochemistry 20, 2071–2075.
[6] Lilley, G.G. (1985) J. Sci. Food Agric., in press.
[7] Lilley, G.G. (1985) J. Sci. Food Agric., in press.
[8] Kreis, M., Forde, B.G., Rahman, S., Miflin, B.J. and Shewry, P.R. (1985) J. Mol. Biol. 183, 499–502.
[9] McKern, N.M., O'Donnell, I.J., Inglis, A.S. and Clarke, B.L. (1983) FEBS Lett. 164, 149–153.
[10] Peterson, J.D., Nehrlich, S.D. and Oyer, P.E. (1977) J. Biol. Chem. 247, 4866–4871.
[11] Woods, E.F. and Inglis, A.S. (1984) Int. J. Biol. Macromol. 6, 277–283.
[12] Hunkapiller, M.W., Hewick, R.M., Dreyer, W.J. and Hood, L.E. (1981) Methods Enzymol. 91, 399–413.
[13] Doonan, S., Doonan, H.J., Hanford, R., Vernon, C.A., Walker, J.M., Airoldi, L.P. da S., Bossa, F., Barra, D., Carloni, M., Fasella, P. and Riva, F. (1975) Biochem. J. 149, 497–506.
[14] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97–120.
[15] Kashlan, N. and Richardson, M. (1981) Phytochemistry 20, 1781–1784.
[16] Petrucci, T., Rab, A., Tomasi, M. and Silano, V. (1976) Biochim. Biophys. Acta 420, 288–297.
[17] Petrucci, T., Sannia, G., Parlamenti, R. and Silano, V. (1978) Biochem. J. 173, 229–235.
[18] Buonocore, V., Gramenzi, F., Pace, W., Petrucci, T., Poerio, E. and Silano, V. (1980) Biochem. J. 187, 637–645.