Volume 229, number 1, 111–118 FEB 05609 February 1988

The amino acid sequence of a cereal Bowman-Birk type trypsin inhibitor from seeds of Jobs' tears (*Coix lachryma-jobi* L.)

Maria B. Ary*, Peter R. Shewry+ and Michael Richardson

*Department of Biochemistry and Molecular Biology, Science Centre, Federal University of Ceará, 60,001 Fortaleza (CE), Brazil, +Biochemistry Department, Rothamsted Experimental Station, Harpenden AL5 2JQ and Department of Botany, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, England

Received 23 December 1987

The major trypsin inhibitor from seeds of Jobs' tears (*Coix lachryma-jobi*) was purified by heat treatment, fractional precipitation with $(NH_4)_2SO_4$, ion-exchange chromatography on DEAE-Sepharose, gel-filtration on Sephadex G-75 and preparative reverse-phase HPLC. The complete amino acid sequence was determined by analysis of peptides derived from the reduced and *S*-carboxymethylated protein by digestion with trypsin, chymotrypsin and the *S. aureus* V8 protease. The polypeptide contained 64 amino acids with a high content of cysteine. The sequence exhibited strong homology with a number of Bowman-Birk inhibitors from legume seeds and similar proteins recently isolated from wheat and rice.

Trypsin inhibitor; Amino acid sequence; Sequence homology; Bowman-Birk; (Cereal, *Coix lachryma-jobi*, Gramineae)

1. INTRODUCTION

Cereal grains contain numerous protein inhibitors of enzymes. It has been suggested that such inhibitors may be classified into a number of different families on the basis of their amino acid sequence homologies, the position of disulphide bonds and the location of the reactive sites [1]. Current evidence suggests that the cereal proteins of this type may belong to at least six distinct families. For example the α-amylase inhibitors from wheat [2–4], and rye [5], proteinase inhibitors from rye [5], barley [6] and maize [7], and a bifunctional α-amylase/trypsin inhibitor from ragi [8] belong to a so-called superfamily [9] which also includes the cereal chloroform-methanol (CM) soluble proteins and certain 2 S storage proteins from dicotyledonous plants. Other bifunctional inhibitors from barley [10,11], wheat [12,13] and rice [14] have sequence homology with the legume Kunitz inhibitor family, while proteinase inhibitors from barley [15,16] clearly belong to the potato inhibitor 1 family. The sequences of probable α-amylase/proteinase inhibitors from barley [17,18] and rice [19] suggest that they should be included with the ragi I-2 α-amylase inhibitor [20] as a separate family. The diversity of structures to be found amongst the enzyme inhibitors has been further emphasized by the recent report that a maize inhibitor of trypsin and insect α-amylases [21] has strong homology with the sweet protein thaumatin and a pathogenesis-related protein induced in tobacco by infection with tobacco mosaic virus, and the finding that trypsin inhibitors from wheat germ [22] and rice bran [23] have strong structural affinities with the Bowman-Birk family of proteinase inhibitors from legumes.

In wheat germ [22] these inhibitors of the Bowman-Birk type exist in two forms, inhibitor I (14.5 kDa) and inhibitor II (7 kDa), with inhibitor I having a duplicated structure of inhibitor II. The rice bran inhibitor on the other hand appears to ex-

Correspondence address: M. Richardson, Department of Botany, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, England

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

ist only in the inhibitor I (14.5 kDa duplicated) form [23]. Other workers have recently reported the isolation of a 12 kDa trypsin inhibitor from seeds of the cereal crop Jobs' tears (*Coix lachryma-jobi*) [24], which they suggest might be similar to the inhibitor I forms in wheat germ and rice bran.

We now report that the major trypsin inhibitor present in the seeds of *Coix* is a 7 kDa protein with an amino acid sequence of 64 residues which is homologous with the Bowman-Birk family found in legumes and the type II inhibitors of wheat germ.

2. MATERIALS AND METHODS

2.1. *Isolation and purification*

The preliminary steps of the purification of the trypsin inhibitor from defatted flour of seeds of Jobs' tears (extraction with 50 mM NaH_2PO_4-K_2HPO_4 buffer, pH 7.5, containing 0.1 M NaCl; heat treatment; precipitation with $(NH_4)_2SO_4$ (40–95%); ion-exchange chromatography on DEAE-Sepharose CL6B; and gel-filtration on Sephadex G-75) were carried out as described by Ohtsubo et al. [24].

The fractions containing trypsin inhibitory activity after gel filtration on Sephadex G-75 were pooled, desalted and lyophilized. The partially purified trypsin inhibitor (35 mg) was dissolved in 6 M guanidine HCl in 0.1% trifluoroacetic acid (1.5 ml) and injected onto a Vydac preparative C18 wide-pore reverse-phase column (25 cm × 22 mm, 218TP1022; Technicol, Stockport), in a Varian 5000 HPLC apparatus. The column was eluted with a gradient of acetonitrile (HPLC grade S, Rathburn) in 0.1% trifluoroacetic acid (flow rate 10 ml/min) as shown in fig.1.

2.2. *Assay of trypsin inhibitor activity*

The inhibitory activity of all samples against bovine trypsin were determined by measuring the hydrolysis of *N*-benzoyl-DL-arginine-*p*-nitroanilide as described [8].

2.3. *Estimation of molecular mass*

The molecular mass was estimated by high-performance gel filtration on an Ultropac TSK-G 2000SW column (0.75 × 60 cm, LKB Produkter AB) equilibrated and eluted with 6 M guanidine HCl in 50 mM phosphate buffer, pH 6.5, containing 1 mM EDTA. The column was calibrated using chymotrypsin (25.3 kDa), trypsin (24.5 kDa), soybean Kunitz inhibitor (20.1 kDa), lysozyme (14 kDa), wheat α-amylase inhibitor (13.4 kDa), soybean Bowman-Birk inhibitor (7.8 kDa) and insulin (5.8 kDa).

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out in 17% acrylamide gels at pH 8.8 as described by Laemmli [25].

2.4. *Sequence determination*

The trypsin inhibitor was reduced and *S*-carboxymethylated as in [8]. Samples (2.5 mg) of the reduced and *S*-carboxymethylated protein were digested separately with trypsin, chymotrypsin and *S. aureus* V8 protease as in [26]. The mixtures of peptides produced by these methods were fractionated and purified by reverse-phase HPLC on a Vydac analytical reverse-phase column (25 cm × 4.6 mm, 218TP54, Technicol) in a Varian 5000 HPLC using variable gradients of 0–50% acetonitrile in 0.1% trifluoroacetic acid.

The intact inhibitor and the peptides derived from it were sequenced using the 4-*N,N*-dimethylaminoazobenzene-4′-isothiocyanate (DABITC)/phenylisothiocyanate (PITC) double coupling method as in [26]. Amino acid compositions were determined using a Waters Pico-Tag system.

2.5. *Comparison of amino acid sequences*

The amino acid sequence of the *Coix* trypsin inhibitor was compared with those of other proteins stored in a databank (US National Biomedical Research Foundation 1986) and the sequences of other inhibitors more recently obtained, by computer analysis using the FASTP and RDF programmes as described in [27].

2.6. *Predictions of hydropathy and secondary structure*

The profile of hydropathy of the *Coix* trypsin inhibitor was predicted from the amino acid sequence by the method of Kyte and Doolittle [28] using the computer programme of Krystek et al. [29]. The secondary structure of the protein was predicted by the computer method in [30].

3. RESULTS AND DISCUSSION

Application of the preliminary steps of the purification method of Ohtsubo et al. [24] followed by preparative reverse-phase HPLC yielded two peaks of trypsin inhibitory activity, one major and one minor, together with a number of other protein peaks (fig.1). The protein in the major peak (TI1) was eluted at an acetonitrile concentration of 24% which is very similar to the value reported for the wheat germ trypsin inhibitors of the 7 kDa type II [22].

During gel filtration on Sephadex G-75 the elution volume of inhibitor TI1 was very similar to that observed for the 7.8 kDa soybean Bowman-Birk proteinase inhibitor. When subjected to high-performance gel filtration on a TSK-G 2000 column in 6 M guanidine HCl the purified protein TI1 was homogeneous and was shown to have a molecular mass of 7 kDa. However, pretreatment with either dithiothreitol or 2-mercaptoethanol appeared to convert the protein to a form of 14.5 kDa and to larger aggregated forms. The inhibitor also behaved anomalously on SDS-PAGE giving a single narrow band of apparent molecular mass 11 kDa in the absence of reducing agent, and

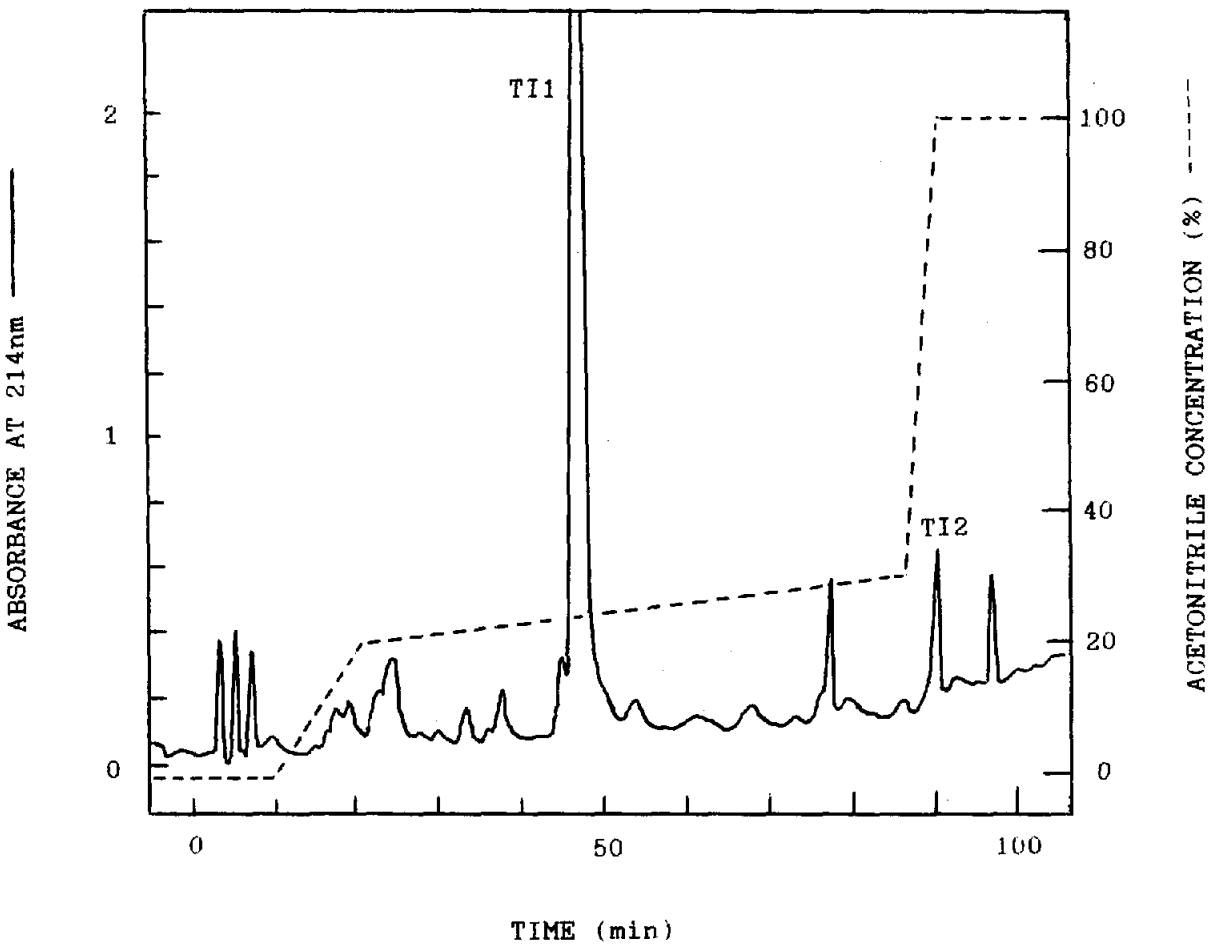


Fig.1. Reverse-phase chromatography of 35 mg partially purified trypsin inhibitor from seeds of *Coix* on a preparative C18 column (25 cm × 22 mm, Vydac 218TP1022), equilibrated in 0.1% trifluoroacetic acid and eluted with a gradient of acetonitrile in 0.1% trifluoroacetic acid at a flow rate of 10 ml/min. Only the proteins in peaks TI1 and TI2 inhibited trypsin.

broader more diffuse bands (17–20 kDa) after treatment with 2-mercaptoethanol. The higher values of molecular mass observed after treatment with reducing agents probably resulted from re-oxidation to form inter-chain disulphide bonds and confirm the observations of Ohtsubo et al. [24] that the *Coix* inhibitor had a similar mobility to cytochrome *c* (12.5 kDa) on SDS-PAGE, but eluted after this protein during gel filtration on Sephadex G-100.

The complete amino acid sequence of the *Coix* trypsin inhibitor is shown in fig.2 together with details of the overlapping peptides from which it was deduced. It should be noted that the N-terminal glycine residue was missing in about 30% of the chains of the intact polypeptide, but no other microheterogeneity was detected. The sequence contained 64 amino acid residues, corresponding to 7262 Da, and was in good agreement with the amino acid composition of the protein. The composition calculated from the sequence is also very similar to that reported for the *Coix* 12 kDa trypsin inhibitor by Ohtsubo et al. [24] when the latter is recalculated for a protein of 7 kDa.

Fig.3 shows the clear sequence homology which exists between the *Coix* trypsin inhibitor, rice bran inhibitor [23], wheat germ inhibitors [22] and Bowman-Birk proteinase inhibitors from a number of legumes [31–37]. The legume inhibitors have

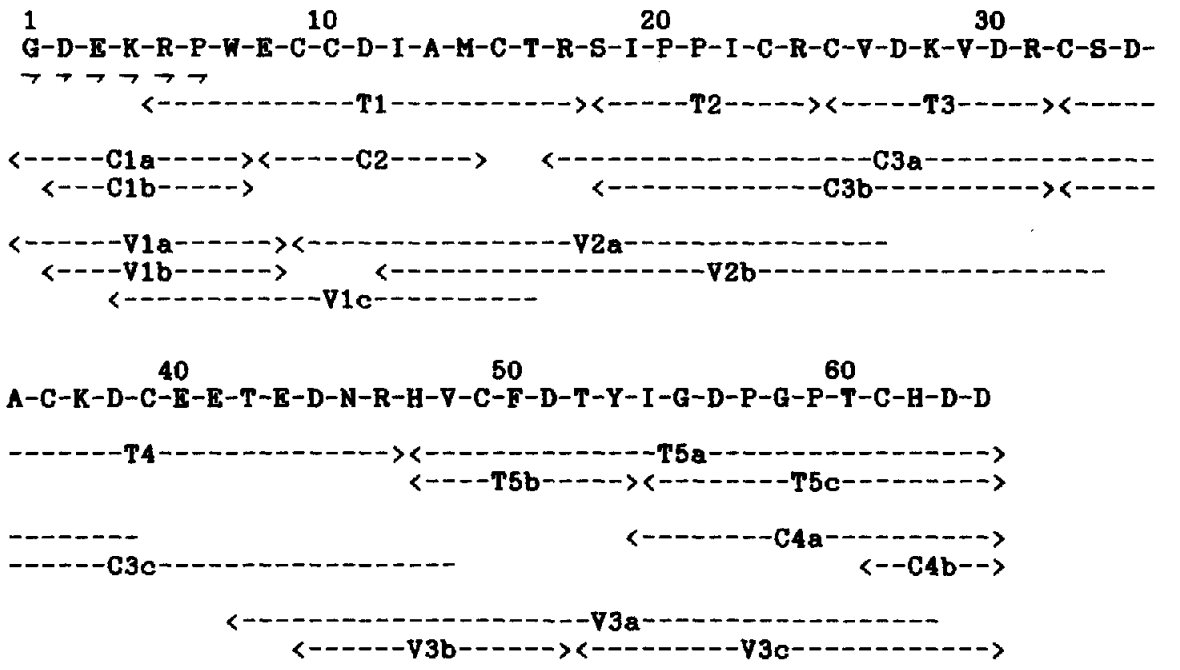
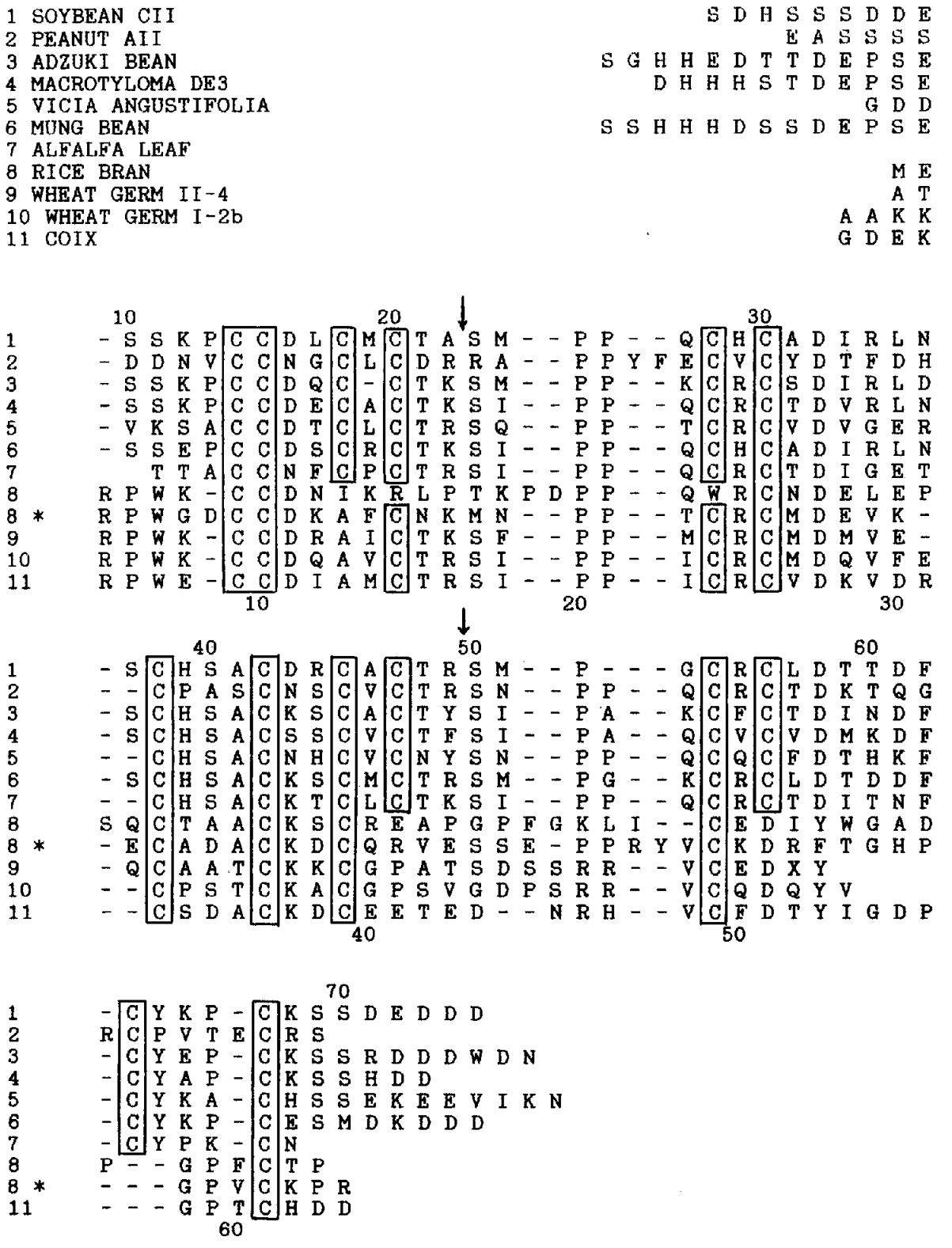


Fig.2. Amino acid sequence of the major trypsin inhibitor TI1 from seeds of Jobs' tears (*Coix lachryma-jobi* L.). T, tryptic peptides; C, chymotryptic peptides; V, peptides from digestion with protease from *S. aureus* V8. Regions of peptides sequenced by the DABITC method (---); blank spaces, residues were not sequenced or yielded unsatisfactory results; —→, residues determined by direct DABITC sequencing of the intact reduced and *S*-carboxymethylated protein.

been shown to be 'double-headed', with one inhibitory site in each of two structural domains. Similar domains are present in the *Coix* inhibitor, where they correspond to residues 1–31 and 32–64, respectively. The *Coix* inhibitor TI1 has greatest homology with the cereal proteins (48–55% identity of residues), but there is also significant homology with the legume Bowman-Birk inhibitors (27–37% identity). Ten out of the 14 cysteine residues which are characteristic of the Bowman-Birk family are conserved in the *Coix* inhibitor, and the sequence homology is particularly high in the region around the first reactive site of the legume inhibitors (in domain I). This homology suggests that the trypsin reactive site in the *Coix* TI1 is Arg(17)–Ser(18) as this peptide bond is located in the exactly homologous position as the first reactive sites of the legume inhibitors, and moreover obeys the tentative rules for such sites [1]. There is no such identifiable peptide bond in the region of the *Coix* sequence which corresponds with the location of the second reactive site in domain II of the double-headed legume inhibitors. We suggest therefore that the *Coix* pro-

Fig.3. Homology of the amino acid sequences of the Bowman-Birk family of proteinase inhibitors from legumes and cereals. (1) Inhibitor CII from soybean (*Glycine max*) [31]; (2) inhibitor AII from peanut (*Arachis hypogaea*) [32]; (3) adzuki bean (*Phaseolus angularis*) [33]; (4) *Macrotyloma axillare* [34]; (5) *Vicia angustifolia* [35]; (6) mung bean (*Phaseolus radiata*) [36]; (7) alfalfa leaf (*Medicago sativa*) [37]; (8) rice bran (*Oryza sativa*) residues 1–69 [23], residues 70–133 are an internal duplication homologous with residues 3–69, and are shown as 8*; (9) wheat germ (*Triticum aestivum*) II-4 incomplete sequence [22]; (10) wheat I-2b incomplete sequence [22]; (11) Jobs' tears (*Coix lachryma-jobi*). On the numbering of the soybean (top line) and the *Coix* (bottom) sequences are shown. The sequences are aligned for maximum homology, resulting in some gaps (–) which may represent insertions/deletions. Arrows indicate the position of the reactive sites determined for the legume inhibitors. The cysteine residues are boxed to facilitate comparisons.


1 SOYBEAN CII
2 PEANUT AII
3 ADZUKI BEAN
4 MACROTYLOMA DE3
5 VICIA ANGUSTIFOLIA
6 MUNG BEAN
7 ALFALFA LEAF
8 RICE BRAN
9 WHEAT GERM II-4
10 WHEAT GERM I-2b
11 COIX
10
20
30
40
50
60
70

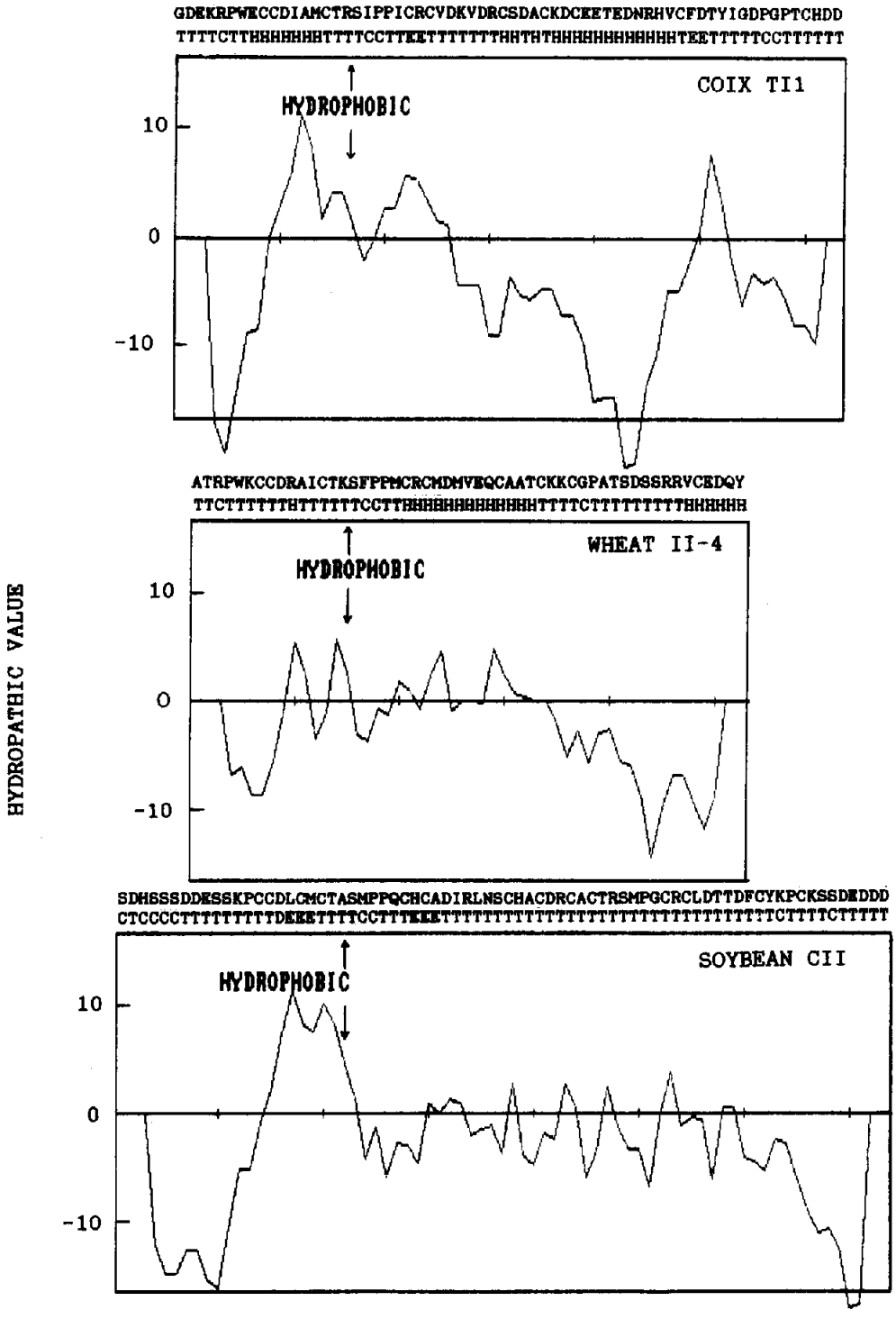


Fig.4. Comparison of the hydropathy profiles and secondary structures predicted from the amino acid sequences of the *Coix* TI1 (top) wheat germ II-4 (middle) and soybean CII (bottom) inhibitors. The Kyte and Doolittle [28] plots of hydropathy were computed as described in [29]. The secondary structures were predicted using the computer method as in [30]. H, α-helix; E, β-sheet; T, β-turn; and C, random coil. Arrows indicate the positions of the putative reactive sites.

tein is a single-headed inhibitor like the type II inhibitor in wheat germ [22], and is not double-headed as suggested by Ohtsubo et al. [24] who based their calculations of the stoichiometry of inhibition on an erroneous estimate of the molecular mass of the inhibitor protein. The larger type I inhibitor of rice bran [23] differs in the presence of an internal duplication, residues 70–133 (8* in fig.3) corresponding to residues 3–69 (8 in fig.3). The sequences reported for the type I and II inhibitors of wheat are incomplete, but an internal duplication also appears to be present in the former [22].

Using the FASTP and RDF computer programmes [27], limited homology of low significance can also be detected between a short region of the *Coix* inhibitor (residues 9–25) and sections of the sequences of the trypsin inhibitors from rye (residues 44–57) [5] and barley (residues 43–56) [6], which are classified in the cereal superfamily. Such weak affinities between the Bowman-Birk and cereal families have been noted previously [38].

Fig.4 shows a comparison of the hydropathy profiles of the *Coix* trypsin inhibitor, wheat germ type II inhibitor and soybean Bowman-Birk CII inhibitor. As might be expected the two cereal inhibitors show the greatest similarities, but all three proteins are similar in possessing hydrophilic regions near their N- and C-termini, and more hydrophobic regions around the putative reactive sites. Also shown in fig.4 are the secondary structures predicted from the amino acid sequences. It can be seen that the only regions showing much similarity are those around the reactive sites. It should be noted however that the *Coix* and wheat inhibitors have similar contents of α-helix (34 and 37%, respectively) and β-turn (55 and 52%, respectively), whereas the soybean CII has a higher content of β-turn (81%) and no α-helix. Such levels of β-turns and the frequency of disulphide bridges probably account for the unusual thermal stability of these proteins.

Acknowledgements: We are grateful to the Science and Engineering Research Council and CAPES (Brazil) for financial support, and to Dr H. Nakano (Prefectural Agricultural Experimental Station, Okayama, Japan) for supplying the seeds of *Coix*. We also thank J. Gilroy for valuable technical assistance and Professor D. Boulter for the provision of certain facilities.

REFERENCES

[1] Laskowski, M. jr and Kato, I. (1980) Annu. Rev. Biochem. 49, 593.
[2] Kashlan, N. and Richardson, M. (1981) Phytochemistry 20, 1781–1784.
[3] Maeda, K., Hase, T. and Matsubara, H. (1983) Biochim. Biophys. Acta 743, 52–57.
[4] Maeda, K., Kakabayashi, S. and Matsubara, H. (1985) Biochim. Biophys. Acta 828, 213–221.
[5] Lyons, A., Richardson, M., Tatham, A.S. and Shewry, P.R. (1987) Biochim. Biophys. Acta 915, 305–313.
[6] Odani, S., Koide, T. and Ono, T. (1983) J. Biol. Chem. 258, 7998–8003.
[7] Mahoney, W.C., Hermodson, M.A., Jones, B., Powers, D.D., Corfman, R.S. and Reeck, G.R. (1984) J. Biol. Chem. 259, 8412–8416.
[8] Campos, F.A.P. and Richardson, M. (1983) FEBS Lett. 152, 300–304.
[9] Shewry, P.R., Lafiandra, D., Salcedo, G., Aragoncillo, C., Garcia-Olmedo, F., Lew, E.J.L., Dietler, M.D. and Kasarda, D.D. (1984) FEBS Lett. 175, 359–363.
[10] Hejgaard, J., Svendsen, I. and Mundy, J. (1983) Carlsberg Res. Commun. 48, 91–94.
[11] Svendsen, I., Hejgaard, J. and Mundy, J. (1986) Carlsberg Res. Commun. 51, 43–50.
[12] Mundy, J., Hejgaard, J. and Svendsen, I. (1984) FEBS Lett. 167, 210–214.
[13] Maeda, K. (1986) Biochim. Biophys. Acta 871, 250–256.
[14] Kato, I., Tominaga, N. and Kihara, F. (1972) Proc. Conference on Protein Structure (Maebashi) 23, 53–56.
[15] Svendsen, I., Jonassen, I., Hejgaard, J. and Boisen, S. (1980) Carlsberg Res. Commun. 45, 389–395.
[16] Svendsen, I., Boisen, S. and Hejgaard, J. (1982) Carlsberg Res. Commun. 47, 45–53.
[17] Mundy, J. and Rogers, J.C. (1986) Planta 169, 51–63.
[18] Svensson, B., Asano, K., Jonassen, I., Poulsen, F.M., Mundy, J. and Svendsen, I. (1986) Carlsberg Res. Commun. 51, 493–500.
[19] Yu, Y.G., Chung, C.H., Fowler, A. and Suh, S.W. (1987) Biochemistry, in press.
[20] Campos, F.A.P. and Richardson, M. (1984) FEBS Lett. 167, 221–225.
[21] Richardson, M., Valdes-Rodriguez, S. and Blanco-Labra, A. (1987) Nature 327, 432–434.
[22] Odani, S., Koide, T. and Ono, T. (1986) J. Biochem. 100, 975–983.
[23] Tashiro, M., Hashino, K., Shiozaki, M., Ibuki, F. and Maki, Z. (1987) J. Biochem. 102, 297–306.
[24] Ohtsubo, K., Yanagi, S.O. and Yanase, H. (1985) Agric. Biol. Chem. 49, 1985–1991.
[25] Laemmli, U.K. (1970) Nature 227, 680–685.
[26] Richardson, M., Campos, F.A.P., Moreira, R.A., Ainouz, I.L., Begbie, R., Watt, W.B. and Pusztai, A. (1984) Eur. J. Biochem. 144, 101–111.
[27] Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435–1441.
[28] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[29] Krystek, S.R., Reichert, L.E. and Andersen, T.T. (1985) Endocrinology 117, 1110–1124.

[30] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 97–120.
[31] Odani, S. and Ikenaka, T. (1977) J. Biochem. 82, 1523–1531.
[32] Norioka, S. and Ikenaka, T. (1983) J. Biochem. 94, 589–599.
[33] Yoshikawa, M., Kiyohara, T., Iwasaki, T., Ishii, Y. and Kimura, N. (1979) Agric. Biol. Chem. 43, 787–796.
[34] Joubert, F.J., Kruger, H., Townshend, G.S. and Botes, D.P. (1979) Eur. J. Biochem. 97, 85–91.
[35] Shimokawa, Y., Kuromizu, K., Araki, T., Ohata, J. and Abe, O. (1984) Eur. J. Biochem. 143, 677–684.
[36] Wilson, K.A. and Chen, J.C. (1983) Plant Physiol. 71, 341–349.
[37] Brown, W.E., Takio, K., Titani, K. and Ryan, C.A. (1985) Biochemistry 24, 2105–2108.
[38] Kreis, M., Shewry, P.R., Forde, B.G., Forde, J. and Miflin, B.J. (1985) Oxf. Surv. Plant Mol. Cell Biol. 2, 253–317.