Volume 239, number 1, 147–150 FEB 06434 October 1988

Signal peptide homology between the sweet protein thaumatin II and unrelated cereal α-amylase/trypsin inhibitors

Ana Lázaro, Pablo Rodriguez-Palenzuela, Carmen Maraña, Pilar Carbonero and Francisco Garcia-Olmedo

Cátedra de Bioquímica y Biología Molecular, ETS Ingenieros Agrónomos-UPM, E-28040 Madrid, Spain

Received 18 August 1988

A cDNA clone (pUP-23) corresponding to a member of a protein family that includes inhibitors of trypsin and of heterologous α-amylases has been selected from a library derived from developing barley endosperm and its sequence has been determined. A stretch of 95 nucleotides that included the signal peptide and the first 8 residues of the mature protein was found to be homologous to an exactly equivalent region of the nucleotide sequence encoding the sweet protein thaumatin II. Evolutionary implications of this finding are discussed.

α-Amylase/trypsin inhibitor; Thaumatin II; Cereal; (Barley, *Thaumatococcus danielli*)

1. INTRODUCTION

Plant proteinaceous inhibitors of serine proteinases belong to at least eight protein families, one of which also includes inhibitors of endogenous α-amylases, while members of two others inhibit either trypsin, heterologous α-amylases, or both (see [1]). The sequence of a maize protein which is a potent inhibitor of bovine trypsin and insect α-amylase has recently been determined and found to be homologous with the sweet protein thaumatin II, present in the fruits of *Thaumatococcus danielli*, and with a pathogenesis-related protein induced in tobacco plants following infection with tobacco mosaic virus [2]. A second, unrelated protein family includes the subunits of tetrameric, dimeric and monomeric inhibitors of heterologous α-amylases, trypsin inhibitors, and bifunctional ones present in various cereals [1,3–14]. We report here that the signal peptides of some members of the second inhibitor family are homologous to the signal peptide of thaumatin II.

Correspondence address: F. Garcia-Olmedo, Cátedra de Bioquímica y Biología Molecular, ETS Ingenieros Agrónomos-UPM, E-28040 Madrid, Spain

2. MATERIALS AND METHODS

A cDNA library, derived from poly(A)$^+$ RNA collected at 20 days after anthesis from developing barley endosperm (*Hordeum vulgare* cv. Bomi) as described by Paz-Ares et al. [10], was screened by standard procedures [15], using as probe the insert in plasmid pUP-13 [10], radioactively labeled by nick-translation. Restriction digestion, agarose gel electrophoresis and Southern blotting to nylon membranes (Hybond N, Amersham) were performed according to Maniatis et al. [15] and to the manufacturer's instructions. Hybridization to nick-translated probes was in 5 × SSPE (0.9 M NaCl, 0.05 M NaH_2PO_4, pH 7.4, 0.005 M EDTA), 2 × Denhardt's (0.04% polyvinylpyrrolidone, 0.04% BSA, 0.04% Ficoll), 0.2% SDS, 100 μg/ml salmon sperm DNA, at 60°C. Nucleotide sequences were determined by the method of Maxam and Gilbert [16], after subcloning in the *Pst*I site of plasmids pUC-12 and pUC-13 [17].

3. RESULTS

A barley endosperm cDNA library was screened for new clones belonging to the family of α-amylase/trypsin inhibitors using as probe the insert in plasmid pUP-13, which had been previously reported as corresponding to a member of this protein family [10]. Clone pUP-23 was selected because it hybridized weakly with the probe and had a different partial restriction map.

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

The sequence of the insert in clone pUP-23 is presented in fig.1. The longest open reading frame in the nucleotide sequence coded for a protein whose N-terminal sequence had the typical features of a signal peptide and was followed by the sequence of a putative mature protein that was clearly homologous to the previously described inhibitors (fig.1, table 1). The sequence was closer to

```
                                                                    38
                                    AAGAGATTGAACCAACGACCAATAAACTAGTATCAACA

                                                                    95
ATG GCA TCC GAC CAT CGT CGC TTC GTC CTC TCC GGC GCC GTC TTG CTC TCG GTC CTC
|Met Ala Ser Asp His Arg Arg Phe Val Leu Ser Gly Ala Val Leu Leu Ser Val Leu
1  signal peptide
                                                                   152
GCC GTC GCC GCC GCC ACC TTG GAG AGC GTC AAG GAC GAG TGC CAA CTA GGG GTG GAC
Ala Val|Ala Ala Ala Thr Leu Glu Ser Val Lys Asp Glu Cys Gln Leu Gly Val Asp
20       mature protein
                                                                   209
TTC CCG CAT AAC CCG TTA GCC ACC TGC CAC ACC TAC GTG ATA AAA CGG GTC TGC GGC
Phe Pro His Asn Pro Leu Ala Thr Cys His Thr Tyr Val Ile Lys Arg Val Cys Gly
39
                                                                   266
CGC GGT CCC AGC CGG CCC ATG CTG GTG AAG GAG CGG TGC TGC CGG GAG CTG GCG GCC
Arg Gly Pro Ser Arg Pro Met Leu Val Lys Glu Arg Cys Cys Arg Glu Leu Ala Ala
58
                                                                   323
GTC CCG GAT CAC TGC CGG TGC GAG GCG CTG CGC ATC CTC ATG GAC GGG GTG CGC ACG
Val Pro Asp His Cys Arg Cys Glu Ala Leu Arg Ile Leu Met Asp Gly Val Arg Thr
77
                                                                   380
CCG GAG GGC CGC GTG GTT GAG GGA CGG CTC GGT GAC AGG CGT GAC TGC CCG AGG GAG
Pro Glu Gly Arg Val Val Glu Gly Arg Leu Gly Asp Arg Arg Asp Cys Pro Arg Glu
96
                                                                   437
GAG CAG AGG GCG TTC GCC GCC ACG CTT GTC ACG GCG GCG GAG TGC AAC CTA TCG TCC
Glu Gln Arg Ala Phe Ala Ala Thr Leu Val Thr Ala Ala Glu Cys Asn Leu Ser Ser
115
                                                                   497
GTC CAG GCG CCG GGA GTA CGC TTG GTG CTA CTG GCA GAT GGA TGA CGATGCAAATGCGCC
Val Gln Glu Pro Gly Val Arg Leu Val Leu Leu Ala Asp Gly ter
134
                                                                   572
AAGGTAATGAAGCGGAGTACTGTATACAGAATAAAAGTACTCGAGTGAAAACAAACTCATAAATAAACCTTGTGA
                              poly A                          poly A

                                                                   647
GATGTATGCGTATGATCTATGGTGTGGACAGTTAAATTGTGGCCGATTGATGAATAAAAAAGGTTGGAACAAATT
                                                    poly A

                          672
AAATTGTTGTGGGTTCATATACTAT
```

Fig.1. Nucleotide sequence and deduced amino acid sequence corresponding to the longest open reading frame of the insert in clone pUP-23. The beginning of the signal peptide and of the mature protein, as well as the polyadenylation signal (poly A) are indicated.

Table 1

Homology (% identical positions) of the amino acid sequence deduced for the mature protein from the nucleotide sequence of the insert in clone pUP-23 and those of members of the cereal α-amylase/trypsin inhibitor family

Inhibitor	Species	%	Ref.
Bifunctional inhibitor I	ragi	49	11
Trypsin inhibitor	maize	47	12
Trypsin inhibitor CMe	barley	42	4,6,7
Trypsin inhibitor CMc[a]	barley	35	4,6,7
α-Amylase inhibitors			
tetrameric CMd[a]	barley	35	5
CMa[a]	barley	32	5
CMb[a]	barley	26	5
CM1[a]	wheat	32	3,4
CM2[a]	wheat	32	3,4
CM3[a]	wheat	32	3,4
CM16[a]	wheat	22	3,4
CM17[a]	wheat	25	3,4
dimeric BDAI-1	barley	24	14
0.19	wheat	26	8
0.53	wheat	26	8
monomeric 0.28	wheat	25	9

[a] Comparisons based on partial, N-terminal sequences

inhibitors showing antitrypsin activity (35–49% identical amino acid residues) than to the rest (22–35%). A search for homology of the present sequence with previously described ones, using the Microgenie program (Beckman), led to the alignment shown in fig.2, in which a stretch of 95 nucleotides that included the signal peptide and the first 8 residues of the mature protein was found to be homologous to an exactly equivalent region of the nucleotide sequence encoding the sweet protein thaumatin II [18]. The JUMTEST z score [19] for the alignment of the peptides is more than five standard deviations above the mean obtained for randomly generated sequences having the same base composition. In contrast, no homology can be shown between the sequences of the mature proteins, thaumatin II and that encoded by plasmid pUP-23, whose sequences are of quite different length (214 versus 125 residues). When the two proteins are aligned, the z score [19] for the alignment does not differ significantly from that expected for randomly generated sequences with the same base composition.

4. DISCUSSION

Pertinent binary comparisons of signal peptides and of mature proteins from the α-amylase/trypsin inhibitor family and from thaumatin II, as well as of the mature maize bifunctional inhibitor, which are presented in fig.3, allow a discussion of the possible evolutionary pathway leading to the described situation. A faster rate of divergence of the mature proteins as compared with the signal peptides, which would have pushed homology beyond recognition, does not seem likely for the following reasons: (i) signal peptides seem to be diverging at a faster rate than the mature sequences within the α-amylase/trypsin inhibitor family; (ii) divergence between thaumatin II (*Thaumatococcus danielli*, order Zingiberales) and the maize bifunctional inhibitor (*Zea mays*, order Poales) is smaller than that observed between the signal peptides of pUP-23 (*Hordeum vulgare*, order Poales) and thaumatin II; (iii) these signal peptides are diverging at a fast rate, as judged from the excess of first base differences between the two sequences (fig.2). These observations imply that an extremely fast divergence would have had to take place within the order Poales for this evolutionary pathway to have occurred. On the other hand, the signal peptides corresponding to clones pUP-13 and pUP-23 seem to be homologous among

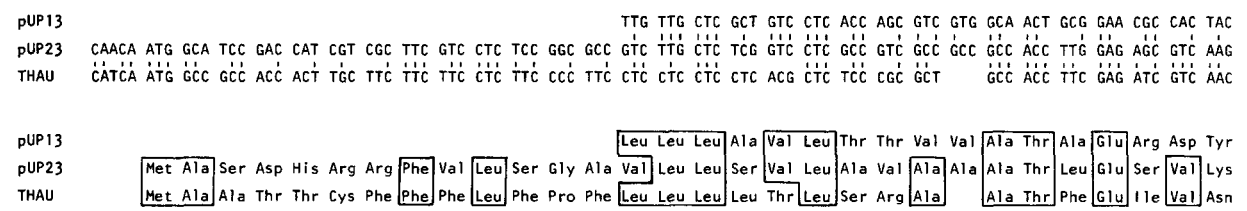

Fig.2. Alignment of nucleotide and deduced amino acid sequences in the region of homology between the insert in clone pUP-23 and thaumatin II cDNA [18]. The 5′-terminal part of the probe used in the screening (insert in clone pUP-13, [10]) has also been aligned.

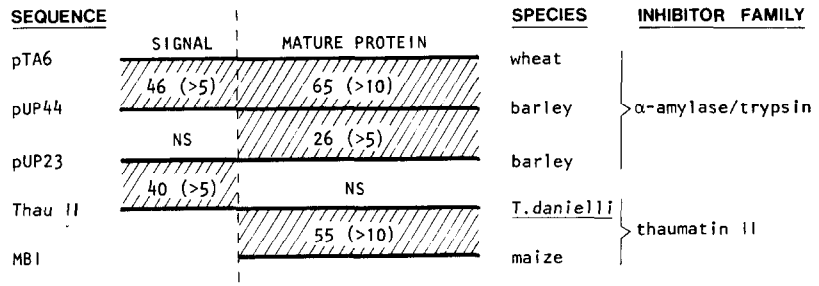

Fig.3. Homology relationships (% identical amino acid residues) of signal peptide and mature protein sequences among members of the two indicated families of inhibitors. Clone pUP-44 corresponds to a dimeric inhibitor of heterologous α-amylases from barley [14] and pTA6 is a related sequence from wheat (Maraña, C., unpublished). Thau II represents the deduced amino acid sequence of the thaumatin II precursor [18]. MBI represents the amino acid sequence of a maize bifunctional inhibitor [2]. The numbers appearing between each pair of sequence segments respectively represent the percentage of identical amino acid residues and the number of standard deviations (in parentheses) that the JUMTEST z score [19] for the alignment differs from that expected for random sequences with the same base composition. NS indicates ≤ 1 standard deviation.

themselves (fig.2) and unrelated to the pair of homologous signal peptides in clones pTA6 and pUP-44, which respectively encode α-amylase inhibitors from wheat and barley belonging to the same family. In conclusion, some sort of intragenic recombination (exon shuffling?) involving ancestral genes from the two families of inhibitors might have occurred, although a common origin combined with markedly different divergence rates cannot be excluded.

Acknowledgements: Advice from R. Blasco and technical assistance from L. Lamoneda and C. Rojas are acknowledged. Financial support was from the Comision Asesora de Investigación Científica y Técnica (grant no.0193/85).

REFERENCES

[1] García-Olmedo, F., Salcedo, G., Sanchez-Monge, R., Gomez, L., Royo, J. and Carbonero, P. (1987) Oxford Surveys Plant Mol. Cell Biol. 4, 275–334.
[2] Richardson, M., Valdes-Rodriguez, S. and Blanco-Labra, A. (1987) Nature 327, 432–434.
[3] Shewry, P.R., Lafiandra, D., Salcedo, G., Aragoncillo, C., García-Olmedo, F., Lew, E.J.-L., Dietler, M.D. and Kasarda, D.D. (1984) FEBS Lett. 175, 359–363.
[4] Barber, D., Sanchez-Monge, R., García-Olmedo, F., Salcedo, G. and Mendez, E. (1986) Biochim. Biophys. Acta 873, 147–151.
[5] Barber, D., Sanchez-Monge, R., Mendez, E., Lazaro, A., García-Olmedo, F. and Salcedo, G. (1986) Biochim. Biophys. Acta 869, 115–118.
[6] Odani, S., Koide, T. and Ono, T. (1983) J. Biol. Chem. 258, 7998–8003.
[7] Lazaro, A., Barber, D., Salcedo, G., Mendez, E. and García-Olmedo, F. (1985) Eur. J. Biochem. 149, 617–623.
[8] Maeda, K., Kakabayashi, S. and Matsubara, H. (1985) Biochim. Biophys. Acta 828, 213–221.
[9] Kashlan, N. and Richardson, M. (1981) Phytochemistry 20, 1781–1784.
[10] Paz-Ares, J., Ponz, F., Rodriguez-Palenzuela, P., Lazaro, A., Hernandez-Lucas, C., García-Olmedo, C. and Carbonero, P. (1986) Theor. Appl. Genet. 71, 842–846.
[11] Campos, F.A.P. and Richardson, M. (1983) FEBS Lett. 152, 300–304.
[12] Mahoney, W.C., Hermondson, M.A., Jones, B., Powers, D.D., Corfman, R.S. and Reeck, G.R. (1984) J. Biol. Chem. 259, 8412–8416.
[13] Higgins, T.J.V., Chandler, P.M., Randall, P.J., Spencer, D., Beach, L.R., Blagrove, R.J., Kortt, A.A. and Inglis, A.S. (1986) J. Biol. Chem. 261, 11124–11130.
[14] Lazaro, A., Sanchez-Monge, R., Salcedo, G., Paz-Ares, J., Carbonero, P. and García-Olmedo, F. (1988) Eur. J. Biochem. 172, 129–134.
[15] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[16] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[17] Vieira, J. and Messing, J. (1982) Gene 19, 259–268.
[18] Edens, L., Heslinga, L., Klok, R., Ledeboer, A.M., Maat, J., Toonen, M.Y., Visser, C. and Verrips, C.Th. (1982) Gene 18, 1–12.
[19] Doolittle, R.F. (1986) Of urfs and orfs. University Science Books.