Sequence homology between barley endosperm protein Z and protease inhibitors of the α_1-antitrypsin family

J. Hejgaard, S.K. Rasmussen*, A. Brandt* and I. Svendsen+

*Department of Biochemistry and Nutrition, Technical University of Denmark, B 224, DK-2800 Lyngby, *Department of Physiology, Carlsberg Laboratory, Gl Carlsbergvej 10, DK-2500 Valby, Copenhagen,* and *+Department of Chemistry, Carlsberg Laboratory, Gl Carlsbergvej 10, DK-2500 Valby, Copenhagen, Denmark*

Received 8 October 1984

Six cDNA clones encoding parts of protein Z, a major barley endosperm albumin, have been identified. Nucleotide and amino acid sequencing have established a 180 residues long C-terminal amino acid sequence of protein Z as well as two minor amino acid sequences (14 and 7 residues) These sequences show that barley protein Z is homologous with human α_1-antitrypsin, human α_1-antichymotrypsin, human antithrombin III, mouse contrapsin and chicken ovalbumin (26–32% of the 180 residues in the C-terminal sequence in identical positions). The sequence homology and specific cleavage of protein Z at a bond corresponding to the reactive site of the inhibitors indicate a possible inhibitory function Inhibition of microbial or pancreatic serine proteases could, however, not be associated with protein Z

Barley seed *Amino acid sequence* *Nucleotide sequence* *Serine proteinase inhibitor* *Ovalbumin* *Polyadenylation signal*

1. INTRODUCTION

Protein Z is a major barley endosperm albumin of ~43 kDa, obtained by ultracentrifugation [1]. The protein has been isolated and partially characterized but not associated with any distinct physiological function [1]. Like barley β-amylase, protein Z is present in 'free' salt-soluble molecular forms as well as thiol- or papain-extractable 'bound' forms in the mature grain [2]. During grain filling, protein Z acts like a storage protein and contributes a substantial part of the grain lysine [3–5]. Protein Z is resistant to thermal and proteolytic modification during production of beer from barley and appears to be the only protein of barley origin present in significant amounts in beer [6].

Recently, cDNA clones encoding parts of protein Z have been identified [7]. The C-terminal amino acid sequence, now obtained by nucleotide and amino acid sequencing has revealed an unexpected homology between this plant protein and members of the physiologically important α_1-antitrypsin family of plasma protease inhibitors, which also includes ovalbumin [8,9].

2. MATERIALS AND METHODS

2.1. *Materials*

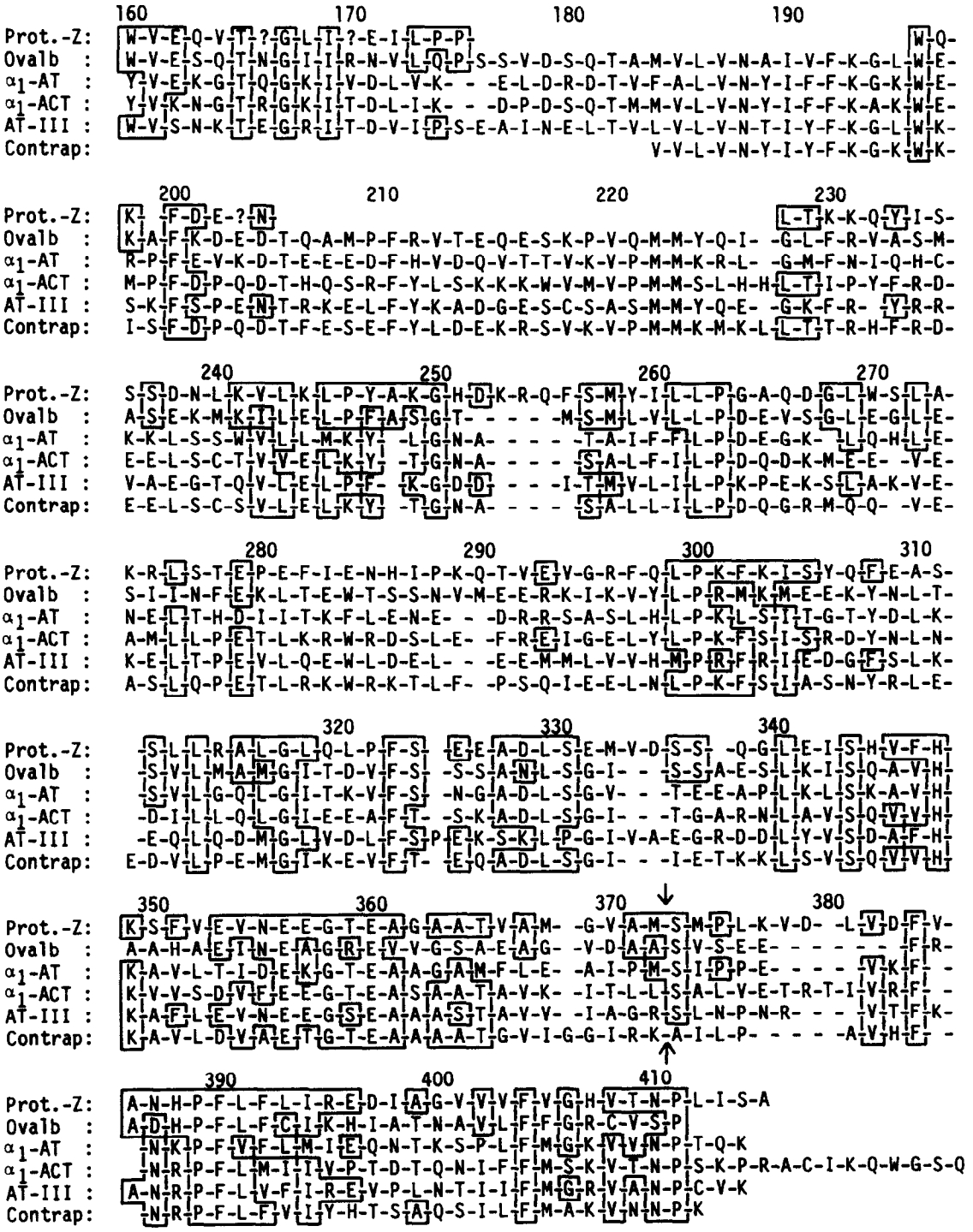
Isolation of protein Z from barley grains has been described in detail [1]. Restriction endonucleases, DNA polymerase (Klenov fragment), T4 DNA ligase, and primers for sequencing and hybridization were obtained from Boehringer, and deoxyadenosine 5′-[α-^{32}P]triphosphate was from New England Nuclear. All chemicals used in the amino acid sequencer were from Rathburn, except ethylacetate, which was from Merck, and Polybrene, from Pierce. Clostripain was obtained from Sigma, and iodosobenzoic acid from Pierce. Bio-gels were from Bio-Rad Laboratories.

2.2. *Nucleotide sequence determination and data analysis*

cDNA inserts were purified and sequenced as described in [10,11]. The nucleotide and amino

Published by Elsevier Science Publishers B V (Biomedical Division)
00145793/85/$3 30

acid sequence data were analysed using a computer programme obtained from R. Staden, Medical Research Centre (Cambridge, UK) which was extended and adapted to fit a Honeywell Bull DSP-8 computer.



2.3. *Amino acid sequence determination*

Amino acid sequencing was done in a liquid phase sequencer (Beckman model 890C) as in [12]. In most cases, Polybrene (hexadimethrine) was added together with the peptide. Identification of the phenylthiohydantoin-amino acids was made by reverse-phase high performance liquid chromatography [13]. Cleavage with cyanogen bromide was made with a 10-fold excess of reagent over the methionine content, and the resulting fragments were separated by gel filtration in 30% acetic acid on a Bio-gel P60 column. Cleavage with clostripain was made in 0.1 M ammonium hydrogen carbonate containing 3 M urea. The resulting peptides were separated by gel filtration in 30% acetic acid on Bio-gel P30 and Bio-gel P10. Cleavage with iodosobenzoic acid was done as in [14] and the resulting peptides separated on Bio-gel P60.

3. RESULTS AND DISCUSSION

A barley endosperm cDNA collection was screened for protein Z cDNA clones using the previously identified clone pc paz1-1 [7] as hybridization probe. The six cDNA clones have identical nucleotide sequences in their overlapping regions (fig.1) and, tentatively the sequences of pc paz1-2 and pc paz1-3 are combined to give a 695 bp long nucleotide sequence (fig.2), corresponding to 40% of the protein Z mRNA size [7].

The nucleotide sequence 1–540 preceding the termination codon TAA could be translated into a 180 residue long amino acid sequence, which may account for the C terminal half of protein Z (fig.2). In addition, a 140 bp long 3′-noncoding part between the termination codon and the poly(A) tail was found. Two possible polyadenylation signals [15] are present at 22 and 30 bp upstream from the polyadenylation site. Clone pc

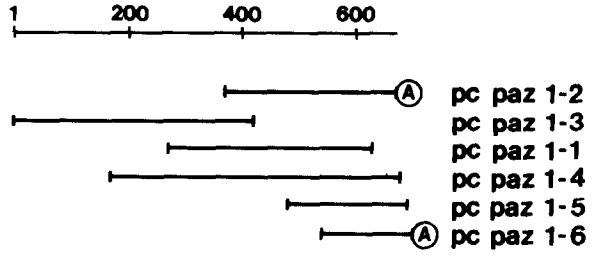


Fig.1. Alignment of the 6 overlapping cDNA clones encoding protein Z. Those having a poly(A) tail are marked with A.

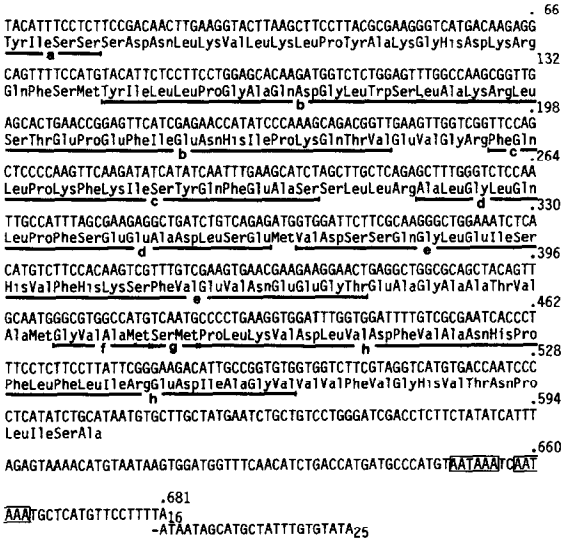


Fig.2. Nucleotide sequence of pc paz1-2 (base 371–695) and pc paz1-3 (base 1–409). The entire non-coding region is shown including the additional 23 nucleotides found in pc paz1-6. The two putative polyadenylation signals are boxed. The predicted amino acid sequence is shown below the nucleotide sequence. Eight sequences confirmed by amino acid sequencing or amino acid analyses are underlined and marked a–h.

←

Fig.3. Comparison of the C-terminal amino acid sequence of barley protein Z (prot.-Z) with the corresponding sequences of chicken ovalbumin (ovalb) [20], human α_1-antitrypsin (α_1-AT) [8,17], human α_1-antichymotrypsin (α_1-ACT) [18], human antithrombin III (AT-III) [19], and mouse contrapsin (Contrap) [21]. ? denotes unidentified residues in protein Z. Residues which are identical in protein Z and one or more of the animal proteins are boxed. Gaps are introduced to maximize homology. Residues are arbitrarily numbered, starting at position 160 in α_1-antitrypsin [8]. Arrows indicate the position of the reactive site of the inhibitors, which is also a specific cleavage site in ovalbumin and protein Z (see text).

paz1-6 is polyadenylated 23 bp downstream to that of pc paz1-2 indicating a polymorphism in the polyadenylation site of protein Z mRNA.

Amino acid sequencing of protein Z confirmed the sequence deduced from the cDNA nucleotide sequences and provided additional amino acid sequence information (sequences a–h, fig.2; see also fig.3). Protein Z was cleaved with cyanogen bromide, and the resulting peptides were separated on Bio-gel P60. The first peak eluted contained one N-terminally blocked peptide (CNBr-I, ~20 kDa) corresponding to the N-terminal half of protein Z. The following peaks contained peptides which were sequenced in part to give the sequences b, e, and h (fig.2). Two small peptides identified by their amino acid compositions (Gly_1, Val_1, Ala_1, Met_1) and (Ser_1, Met_1) were compatible with the sequences f and g, respectively.

The peptide with the N-terminal sequence b was specifically cleaved at arginine residues with clostripain. Four peptides which coeluted from the Bio-gel P60 column were sequenced as a mixture. Two of the sequences could be predicted from sequence b. The remaining amino acids identified in each cycle (sequences c,d, fig.2) always corresponded to those predicted from the nucleotide sequence. Finally, a peptide isolated from a clostripain digest of protein Z had the N-terminal sequence: Leu-Thr-Lys-Lys-Gln-Tyr-Ile-Ser-Ser- which overlaps the N-terminal sequence deduced from nucleotide sequencing (peptide a, fig.2). Thus, when Met and Arg residues are inferred at the N-terminal ends of cyanogen bromide or clostripain peptides, respectively, more than 70% of the deduced amino acid sequence has been confirmed by amino acid sequencing, and no deviations were observed.

When the C-terminal 180 residues of protein Z are compared with the C-terminal sequences of proteins which are members of the α_1-antitrypsin family, a high degree of sequence homology with these proteins is observed (fig.3). Parts of the angiotensinogen [16] sequence are also related to amino acid sequences of the α_1-antitrypsin family but, for clarity, this sequence was not included in the comparison. The homology (fig.3) is not confined to a smaller domain of the molecules, e.g., the reactive site region of the inhibitors, as seen from the presence of Gly, Pro and bulky hydrophobic (aromatic or alifatic) side chains in essentially all positions in protein Z where these residues are conserved in the 5 other proteins. With introduction of only two additional gaps to maximize homology, the percentages of amino acids in identical positions are 26%, 28%, 32%, 32%, and 30%, when protein Z is compared with α_1-antitrypsin [8,17], α_1-antichymotrypsin [18], antithrombin III [19], ovalbumin [20] and contrapsin [21], respectively. The same degree of similarity was found when the four completed primary structures of animal proteins were compared with each other [8], except that α_1-antitrypsin and α_1-antichymotrypsin were more closely related (42% of the amino acids in identical positions [18]).

When the large N-terminally blocked cyanogen bromide fragment (CNBr-I) was cleaved with iodosobenzoic acid on the carboxyl side of tryptophan residues, two of the peaks obtained after gel filtration each contained one major peptide. Sequencing of these two peptides gave the sequences: Val-Glu-Gln-Val-Thr-?-Gly-Leu-Ile-?-Glu-Ile-Leu-Pro-Pro- and Gln-Lys-Phe-Asp-Glu-?-Asn-. The two peptides could be aligned with the inhibitor sequences starting with the highly conserved aromatic residues at positions 160 and 196, respectively (fig.3). Clearly, homology is not confined to the C-terminal half of the molecules, and the present results strongly suggest that barley protein Z is a member of the α_1-antitrypsin (super) family of homologous proteins.

The previously published N-terminal sequence starting Pro-Leu-Lys- [7], which was obtained by direct sequencing of the protein Z preparation, corresponds to the sequence h (fig.2) at the C-terminal of the molecule. Gel filtration of protein Z in 30% acetic acid followed by amino acid and sequence analyses showed that the protein Z preparation contained N-terminally blocked protein as well as three coeluting peptide fragments corresponding to a partial cleavage of the protein at a bond near the beginning of sequence h. In addition to the major sequence: Pro-Leu-Lys-, minor amounts of peptides with N-terminals: Ser-Met-Pro-Leu- and Met-Pro-Leu- were also present (sequence g followed by sequence h, fig.2). These observations clearly indicate that a part of the N-terminally blocked protein Z has been cleaved proteolytically at a Met–Ser bond at peptide g prior to or during purification. Activity of amino pep-

tidases, which are relatively abundant in mature barley, may explain a further modification during purification to give a peptide with an N-terminal proline as the major component (sequence h).

The proposed cleavage site of protein Z (Met_{372}–Ser_{373}, fig.3) matches exactly the proposed reactive site bonds of the protease inhibitors (fig.3), suggesting that protein Z may also be a serine protease inhibitor. Measurements with serine proteases, including subtilisin, Aspergillus protease, pancreas elastase, chymotrypsin and trypsin did not show any inhibitory activity of the purified protein Z. Cleavage of the reactive site bond as well as exposure to a pH below 5 (during purification of protein Z [1]) are conditions which, characteristically, inactivate inhibitors of the α_1-antitrypsin family [9]. However, attempts to associate protein Z with inhibition of one of the above tested enzymes in gel filtration experiments using freshly prepared barley extracts were also negative. Protein Z may still be an inhibitor of an unidentified target enzyme, e.g., protein Z may have a regulatory or protective inhibitory function during grain filling or germination. It is, however, also well-known that α_1-antitrypsin is sensitive to other classes of proteases cleaving the active site peptide bond [9], and alternatively, cleavage by metallo- or thiol proteases followed by aminopeptidases, may explain a modification during purification of protein Z.

Ovalbumin has no known inhibitory function, but the transformation of ovalbumin into plakalbumin occurs in the presence of subtilisin [22,23] by initial cleavage of a specific bond identified as Ala_{372}–Ser_{373} (fig.3). Thus, protein Z and ovalbumin are both sensitive to proteolytic modification at a position corresponding to the reactive site of the inhibitors, and both proteins are present in high amounts in tissues surrounding a growing embryo. These tissues are rich in a number of enzyme inhibitors with apparent protective functions. One may speculate that proteins with an original physiological function as inhibitors may have acquired an important alternative function as amino acid storage during evolution. Recently, amino acid sequence homology has been shown between a 13.3-kDa barley trypsin inhibitor and a castor bean protein with typical storage functions consisting of two disulfide-linked polypeptide chains of ~6.6 kDa and ~4.3 kDa, respectively [24].

Finally it should be noted that comparison of nucleotide sequences corresponding to the C-terminal parts of protein Z and α_1-antichymotrypsin [18] by dot-matrix analysis [25] did not reveal any obvious homology. Work is in progress to isolate and characterize genomic clones coding for protein Z. Comparison of the protein Z gene(s) with those encoding the animal protease inhibitors and ovalbumin may provide new information concerning the evolution of the α_1-antitrypsin gene family.

ACKNOWLEDGEMENTS

This work was supported in part by the Biomolecular Engineering Programme of the Commission of the European Communities with contract no.GB 1-4-024-DK to professor D. von Wettstein. The technical assistance of T. Bektved, N. Rasmussen, B. Corneliussen, L. Sørensen and P. Breddam is gratefully acknowledged.

REFERENCES

[1] Hejgaard, J (1982) Physiol. Plant. 54, 174–182.
[2] Hejgaard, J. (1976) Physiol Plant. 38, 293–299.
[3] Hejgaard, J. and Boisen, S. (1980) Hereditas 93, 311–320.
[4] Giese, H. and Hejgaard, J. (1984) Planta 161, 172–177
[5] Giese, H. and Hopp, H.E. (1984) Carlsberg Res Commun. 49, 365–383.
[6] Hejgaard, J and Kaersgaard, P (1983) J. Inst. Brew. 89, 402–410.
[7] Rasmussen, S.K , Hopp, H.E., Brandt, A., Svendsen, I. and Hejgaard, J (1984) Carlsberg Res. Commun. 49, 385–390.
[8] Carrell, R W., Jeppsson, J -O , Laurell, C -B., Brennan, S O., Owen, M.C., Vaughan, L. and Boswell, D.R. (1982) Nature 298, 329–334
[9] Travis, J. and Salvesen, G.S (1983) Annu. Rev. Biochem 52, 655–709.
[10] Sanger, F., Nicklen, S and Coulson, A.R (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[11] Rasmussen, S.K., Hopp, H.E and Brandt, A. (1983) Carlsberg Res. Commun. 48, 187–199.
[12] Johansen, J.T , Overballe-Petersen, C., Martin, B., Hasemann, V. and Svendsen, I. (1979) Carlsberg Res. Commun. 44, 201–217.
[13] Svendsen, I., Martin, B. and Jonassen, I. (1980) Carlsberg Res. Commun. 45, 79–85.

[14] Mahoney, W C., Smith, P.K. and Hermodson, M A (1981) Biochemistry 20, 443–448
[15] Proudfoot, N.J and Brownlee, G.G. (1976) Nature 263, 211–214
[16] Kageyama, R , Ohkubo, H. and Nakanishi, S. (1984) Biochemistry 23, 3603–3609
[17] Kurachi, K , Chandra, T., Friezner Degen, S J , White, T.T , Marchioro, T L., Woo, S L C and Davie, E W. (1981) Proc Natl Acad Sci USA 78, 6826–6830.
[18] Chandra, T., Stackhouse, R , Kidd, V J., Robson, K J.H and Woo, S L C (1983) Biochemistry 22, 5055–5060
[19] Petersen, T E , Dudek-Wojciechowska, G , Sottrup Jensen, L and Magnusson, S (1979) in The Physiological Inhibitors of Blood Coagulation and Fibrinolysis (Collen, D et al. eds) pp 43–54, Elsevier, Amsterdam, New York
[20] McReynolds, L., O'Malley, B.W , Nisbet, A D , Fothergill, J E , Givol, D., Fields, S , Robertson, M. and Brownlee, G G (1978) Nature 273, 723–728
[21] Hill, R.E , Shaw, P H , Boyd, P.A., Baumann, H and Hastie, N D (1984) Nature 311, 175–177
[22] Ottesen, M. (1958) Comp. Rend. Trav Lab Carlsberg, Sér Chim 30, 211–270
[23] Thompson, E O P , Sleigh, R W and Smith, M B (1971) Aust J Biol Sci 24, 525–534
[24] Odani, S., Koide, T , Ono, T and Ohnishi, K (1983) Biochem. J 213, 543–545
[25] Maizel, J.W and Lenk, R G. (1981) Proc Natl Acad. Sci. USA 78, 7665–7669