Primary Structure of Barwin: A Barley Seed Protein Closely Related to the C-Terminal Domain of Proteins Encoded by Wound-Induced Plant Genes

Birte Svensson,[‡] Ib Svendsen,[‡] Peter Højrup,[§] Peter Roepstorff,[§] Svend Ludvigsen,[‡,∥] and Flemming M. Poulsen*[,‡]

Carlsberg Laboratorium, Kemisk Afdeling, Gamle Carlsberg Vej 10, DK-2500 Valby, Copenhagen, Denmark, and Institut for Molekylær Biologi, Odense Universitet, Campusvej 55, DK-5230 Odense M, Denmark

Received November 27, 1991; Revised Manuscript Received June 9, 1992

ABSTRACT: Barwin is a basic protein with p*I* above 10 and molecular mass 13.7 kDa isolated from aqueous extracts of barley seed. The complete amino acid sequence of 125 residues has been determined by a combination of conventional protein sequencing, plasma desorption mass spectrometry, and ^{1}H nuclear magnetic resonance spectroscopy. Three disulfide bridges have been localized as Cys31–Cys63, Cys52–Cys86, and Cys66–Cys123 both by ^{1}H nuclear magnetic resonance spectroscopy and by plasma desorption mass spectrometry. The N-terminal residue was identified as pyroglutamate. Barwin is closely related to a peptide segment of 122 residues at the C-terminal region of the proteins encoded by two wound-induced genes in potato plants, *win1* and *win2*, and a protein encoded by the hevein gene of rubber tree. In 77 sequence positions of 125 the barwin, *win1*, *win2*, and hevein protein sequences have amino acid sequence identity, when two gaps—one of two residues allowing for the insert of Gly23 and Ala24 and one allowing for the insert of Thr97 in the barwin sequence—are introduced in the latter three. The close sequence similarity with the proteins encoded by the wound-induced potato and rubber tree genes and the ability of the protein to bind saccharides suggest that barwin might belong to a group of proteins involved in a common defense mechanism in plants.

Plants produce a number of proteins that can be directly or indirectly involved in their defense against pathogenic microorganisms, fungi, and mechanical stress (Chrispeels & Raikhel, 1991; Darvill & Albersheim, 1984; Kahl, 1978). There are several examples that the synthesis of certain proteins is stimulated by the stress factor. For example, in potato plants wounding induces two specific genes, *win1* and *win2* (Stanford et al., 1989), and in rubber tree the closely related hevein gene is induced by wounding (Broekaert et al., 1990). These observations suggest that such proteins may play a role in the protection of the plant by mechanisms, however, that remain to be discovered. The induced genes have an open reading frame corresponding to peptides of 200, 211, and 204 amino acid residues, respectively. C-Terminal to the putative cleavage site these proteins have a region of 43 amino acid residues, often referred to as the hevein domain, with partial identity to several chitin-binding domains in other proteins (Stanford et al., 1989; Broekaert et al., 1990). The present paper reports on a basic 125-residue protein from barley seeds that is closely related to a 122 amino acid residue peptide segment in the C-terminal part of the proteins encoded by the wound-induced genes. Furthermore, it has been shown that this protein binds weakly a chitin analog, the tetrameric β-(1,4) oligosaccharide of *N*-acetylglucosamine (Ludvigsen & Poulsen, 1992b). We have named this protein barwin. The biological function of this protein is not yet known; given, however, that the synthesis of closely related proteins in potato and rubber tree is stimulated by wounding, it is interesting to characterize. This paper describes the purification of barwin and the amino acid composition and sequence as well as the identification of the three disulfide bridges. This chemical characterization is a prerequisite for a study of the three-dimensional structure of the protein in solution by ^{1}H NMR spectroscopy as described in the following papers (Ludvigsen & Poulsen, 1992a,b).

METHODS AND MATERIALS

The isolation of barwin followed the procedure previously used for the isolation of the barley seed serine proteinase inhibitor CI-2 (Jonassen, 1982). The barley seeds (15 kg) were milled and extracted with a solution (150 L) of 0.1 M NaCl, 2.5 mM ascorbic acid, and 2.5 mM EDTA at pH 6.3 at 4 °C for 18 h. The extract was filtered and concentrated to 10 L on a DDS 600 membrane filter and diafiltered until the conductivity was lower than that of the starting buffer of the subsequent ion-exchange step on DEAE-cellulose, a 0.3 M glycine buffer at pH 8.0. At each step of the purification the presence of barwin was monitored by SDS–PAGE. Barwin was eluted from the DE-52 column with 0.15 M NaCl and 0.3 M glycine. The eluted protein was dialyzed and subsequently bound on a CM-52 column equilibrated at pH 4.5 in 0.02 M sodium acetate. The protein was eluted in a gradient from 0 to 0.5 M NaCl. The fractions that contained a mixture of the serine proteinase inhibitor CI-2 and barwin were desalted; subsequently, the two proteins were separated on a P-10 column, and barwin was finally purified on a Fractogel CM-650(S) column at pH 4.5 in 0.02 M sodium acetate using a gradient from 0 to 0.5 M NaCl. Typically 75–90 mg of the protein was isolated from 15 kg of barley seeds.

Reduction and 2-pyridylethylation (Svensson et al., 1983a), hydroxylamine cleavage (Bornstein & Balian, 1977), *o*-iodosobenzoic acid cleavage (Mahoney et al., 1981), cleavage with endoproteinase Lys C in urea (Asano et al., 1986; Runswick & Walker, 1983), and cleavage with clostripain in urea (Asano et al., 1986) were all accomplished according to protocols applied routinely in our laboratory. The same applies to desalting of reaction mixtures, the separation of peptides, amino acid sequence analysis, determination of amino acid compositions, and molecular mass determination by SDS–

* Author to whom correspondence should be addressed.
‡ Carlsberg Laboratorium.
§ Odense Universitet.
∥ Present address: Department of Insulin Research, Novo Nordisk, Novo Allé, DK-2880 Bagsværd, Copenhagen, Denmark.

0006-2960/92/0431-8767$03.00/0

Table I: Amino Acid Composition of Barwin

	amino acid analysis	sequence
Asn, Asp	16.7 (17)	18
Thr	8.3 (8)	9
Ser	4.1 (4)	4
Glu, Gln	9.1 (9)	9
Pro	6.3 (6)	6
Gly	12.1 (12)	12
Ala	15.0 (15)	16
Cys	5.0 (5)	6
Val	6.4 (7)	7
Met	0.0 (0)	0
Ile	3.6 (4)	4
Leu	6.0 (6)	6
Tyr	6.5 (7)	7
Phe	3.0 (3)	3
His	2.1 (2)	2
Lys	4.1 (4)	4
Arg	7.0 (7)	7
Trp	ND[a]	5
total		125

[a] ND, not determined.

PAGE (Asano et al., 1986; Johanssen et al., 1979; Svendsen et al., 1980; Svensson et al., 1983a,b). The tryptic cleavage was carried out on native barwin in 50 mM pyridinium acetate at pH 6.5 [4 h, 37 °C at an enzyme/substrate ratio of 1:20 (w/w)] to avoid disulfide interchange. The peptides were separated according to standard procedures (Mikkelsen et al., 1987).

Mass spectra were recorded on a BioIon 10 mass spectrometer (BioIon AB, Uppsala, Sweden). The samples were applied on nitrocellulose covered targets (Johnson et al., 1986) using the spin technique for sample application as reported previously (Nielsen et al., 1988). The mass spectra were analyzed in relation to the known linear sequence by the GPMA program package, thereby excluding the possibility of misassignments (Sørensen et al., 1990). The strategy for the assignments of the disulfide bridges as well as the use of methylation for peptide identification has been described previously (Højrup, 1990).

The protein sequence database MIPS of the Max-Planck-Institute für Biochemie, Martinsried, and the associated programs Search, Relate, and Align were used in the search for amino acid sequence identities in other proteins.

Barley seeds (*Hordeum vulgare*, cv. Hiproly) were obtained from Carlsberg Plant Breeding, Copenhagen. DEAE- and CM-cellulose were DE-52 and CM-52 types from Whatman, Maidstone, U.K.; Bio-Gel, P-10, and P-6 were from Bio-Rad, Richmond, CA. Sephadex G-25 superfine was from Pharmacia, Uppsala, Sweden, and 2-vinylpyridine was from Aldrich, Steinheim, FRG. Clostripain, hydroxylamine hydrochloride, Fractogel CM-650(S), and tyramine hydrochloride were from E. Merck, Darmstadt, FRG. *o*-Iodosobenzoic acid was from Pierce, Rockford, IL, endoproteinase Lys C was from Boehringer-Mannheim, Mannheim, FRG, and trypsin was a gift from Novo Nordisk A/S, Denmark.

RESULTS

The purification of barwin resulted in a highly purified sample as observed, not only by SDS–PAGE but also by mass spectrometry, and when the sequence became known by amino acid analysis (Table I). Similarly, the complete analysis of the NMR spectra gave no indication of any protein contaminants in the purified barwin (Ludvigsen & Poulsen, 1992a).

Barwin was found to have pyroglutamate at the N-terminal as shown by mass spectrometry and amino acid analysis of an

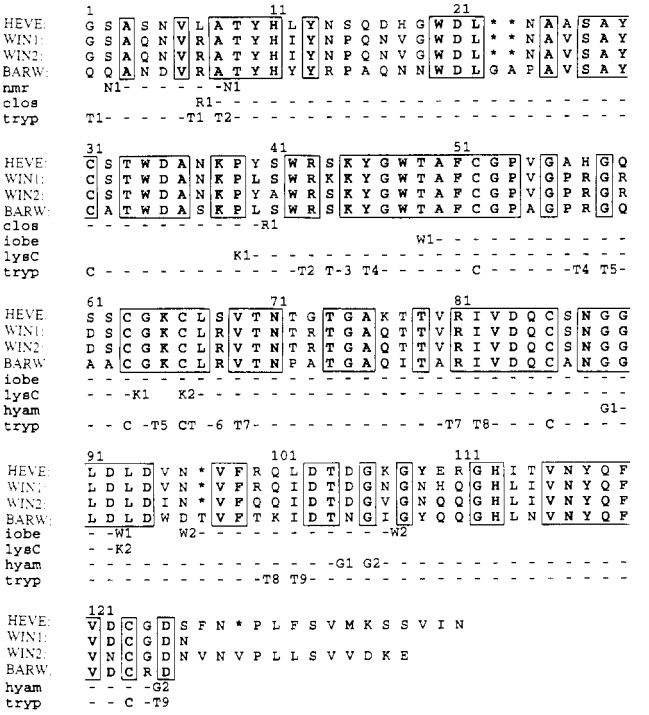

FIGURE 1: Amino acid sequence of barwin (BARW) and comparison with the *win1* and *win2* and hevein gene encoded sequences (WIN1, WIN2, and HEVE, respectively). Sequence identities with barwin in the three proteins are marked with bold letters, and sequence identities in all four sequences are framed. "clos" marks the line for the peptide obtained by the clostripain digestion, and the beginning and end of the peptide sequencing are labeled with R1. Similarly, the peptides obtained by endoproteinase Lys C (lys C) are labeled with and K1 and K2; by *o*-iodosobenzoic acid (iobe), W1 and W2; by hydroxylamine (hyam), G1 and G2; and by trypsin (tryp), T1, T2, ..., and T9. [The tryptic peptides T3 (S44, K45) and T6 (C66, L67, R68) have been labeled differently for typographical reasons.] The cysteines in the sequence are also indicated in the lines of the tryptic fragments with C. The N-terminal residue in barwin is pyroglutamate. N1 has been sequenced by NMR.

N-terminal peptide. Relevant peptides for complete sequence determination were obtained from cleavage with endoproteinase Lys C, clostripain, *o*-iodosobenzoic acid, and hydroxylamine followed by desalting and, subsequently, purification by reversed-phase HPLC occasionally preceded by molecular sieve chromatography. A comparison of the amino acid composition determined by amino acid analysis and that derived from the sequence is shown in Table I. The major part of the sequence was determined from sequencing of seven peptides (R1, K1, K2, W1, W2, G1, and G2) that were sufficient to establish the necessary overlap as seen in Figure 1. The N-terminal nine residues were sequenced on the basis of the amino acid composition of the N-terminal heptapeptide and by use of sequential assignment of the ^{1}H NMR spectrum (Ludvigsen & Poulsen, 1992a). The sequence assignment of the protein was confirmed by the molecular mass of the reduced tryptic peptides as deduced from the mass spectrometrical data (Table IIA). Apart from the sequencing of the N-terminal peptide, the NMR study has provided an almost complete sequential assignment of the protein, and this has provided an independent confirmation of the chemical sequencing.

The positions of the disulfide bonds in barwin were identified by mass spectrometric analysis of the purified peptides obtained by tryptic digestion at nonreducing conditions (Table IIB). The following disulfide bonds were established: Cys31–Cys63, Cys52–Cys86, and Cys66–Cys123, accounting for all cysteine residues in the protein sequence. The assignments were further

Table II: Mass Analysis of Barwin Peptides

(A) Tryptic Peptides of the Reduced Protein

peptide	residue no.	theor mass	obsd mass reduced	cysteine
intact	1...125	13713	13712	
T1	1...7	811.9	812.4	
T2	8...43	4131.5	4134.1	Cys33
T3	44...45	233.3	230.4[a]	
T4	46...58	1382.5	1382.3	Cys53
T5	59...65	633.7	632.9[a]	Cys63
T6	66...68	390.5	387.6[a]	Cys66
T7	69...81	1299.4	1300.5	
T8	82...101	2210.4	2211.8	Cys86
T9	102...125	2770.0	2767.8	Cys123

(B) Summary of Disulfide Location in Barwin

tryptic peptide	obsd mass unreduced[b]	after DTT treatment	methylation analysis	peptide ID	cystine bridges
1	812.4	ND[c]	826.7	T1	
2	1300.5	1301.2	1315.4	T7	
3	3161.1		3232.2	T6–SS–T9	Cys66–Cys123
	(2765.9)	2775.5		T9	
4	4439.0	ND	4513.7	(T6 + T7)–SS–T9	Cys66–Cys123
	(2776.3)		(2824.4)	T9	
	(1672.5)		(1685.7)	T6 + T7	
5	4767.7		4817.8	(T2–SS–T5)	Cys33–Cys63
		4136.8		T2	
6	3808.4		3909.2	(T3 + T4)–SS–T8	Cys52–Cys86
	(2203.8)		(2296.2)	T8	
	(1598.7)	1598.9	(1613.0)	T3+T4	
7	3596.0		3693.1	T4–SS–T8	Cys52–Cys86
			(1382.3)	T8	
	(1382.3)	1383.5	(1397.1)	T4	

[a] The mass has been determined indirectly from measurements of fragments in which the sequence concerned is included (see B). [b] Numbers in parentheses are the masses of fragment ions arising from cleavage of the S–S bond. The precision of mass measurements is in general 0.1%. [c] ND, not determined.

confirmed by in situ reductions and methylation analysis. Independently of the chemical identification of the disulfide bridges, these were localized by NMR spectroscopy and relied on the sequence-specific assignment of the protein ^{1}H NMR spectrum and the structure determination to be described in the succeeding papers (Ludvigsen & Poulsen, 1992a,b). The mass determinations of the tryptic fragments were in full agreement with the chemical determinations of the sequence, providing yet another independent measure for the correctness of the amino acid sequence.

DISCUSSION

The observation that barwin can interact specifically with the oligosaccharide (β-D-Glc$_p$NAc1-4)$_4$ (Ludvigsen & Poulsen, 1992b) makes the protein a potential candidate for being a plant lectin. Barwin, however, has no sequence identity to any of the known plant lectins. In particular, the partial sequence identity between several of the known chitin-binding plant proteins, chitinase, wheat germ agglutinin, and rice and nettle lectin (Stanford et al., 1989), is not observed for barwin. However, in the N-terminal parts of their amino acid sequences these chitin-binding proteins show partial sequence identity with the hypothetical proteins encoded by the closely related wound-induced genes from potato, *win1* and *win2*, and the hevein gene from rubber tree. In this work we have observed that the C-terminal region of these closely related proteins encoded by the three genes is closely related to barwin. In 77 positions of the 125 amino acid residues in the barley seed protein, identity was found in the C-terminal hevein and the two *win* sequences, when two gaps—one of two residue positions to allow for the insert of Gly23 and Ala24 and one to allow for the insert of Thr97—are introduced in the three sequences. Furthermore, several of the amino acid sequence substitutions are structurally conservative according to the classification of Dayhoff et al. (1983). The highly preserved sequences of these proteins in three different families of higher plants suggest that the four proteins are synthesized by the plants to serve similar functions, presumably in the defense of the plants against various pathogens.

The structure determination of barwin has shown that this protein is a single domain (Ludvigsen & Poulsen, 1992a,b), suggesting that the proteins encoded by the potato and hevein gene are two-domain proteins, in which the N-terminal residues from the putative cleavage site form one chitin-binding domain of the 43-residue hevein type and the C-terminal domain that is closely related to barwin forms yet another domain. This is in agreement with the finding of Lee et al. (1991), who have shown that in rubber tree the two domains are present as posttranslational cleavage products.

ACKNOWLEDGMENT

We thank Ms. Pia Mikkelsen for preparation of barwin and for preparation of the peptides used for sequencing. We also thank Ms. Pia Breddam, Ms. Bodil Corneliussen, and Ms. Lone Sørensen for operating the protein sequencers and the amino acid analyzer. P.R. and F.M.P. are members of the Danish Protein Engineering Research Centre (PERC).

REFERENCES

Asano, K., Svensson, B., Svendsen, I., Poulsen, F. M., & Roepstorff, P. (1986) *Carlsberg Res. Commun. 51*, 129–141.

Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 47*, 132–145.

Broekaert, W., Lee, H., Kush, A., Chua, N.-H., & Raikhel, N. V. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 7633–7637.

Chrispeels, M. J., & Raikhel, N. V. (1991) *Plant Cell 3*, 1–9.

Darvill, A. G., & Albersheim, P. (1984) *Annu. Rev. Plant Physiol. 35*, 243–298.

Dayhoff, M. O., Schwartz, R. M., & Orcott, B. L. (1983) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, pp 345–352, National Biomedical Research Foundation, Washington, DC.

Højrup, P. (1990) in *Ion Formation from Organic Solvents* (Hedin, A., Sundquist, B. U. R., & Benninghoven, A., Eds.) pp 61–66, Wiley, Chichester, U.K.

Johansen, J. T., Overballe-Petersen, C. Martin, B., Hasemann, V., & Svendsen, I. (1979) *Carlsberg Res. Commun. 44*, 201–217.

Johnson, G. A., Hedin, A. B., Håkansson, P. L., Sundqvist, B. U. R., Säve, G., Nielsen, P. F., Roepstorff, P., Johansson, K. E., Kamensky, I., & Lindberg, M. S. L. (1986) *Anal. Chem. 58*, 1084–1087.

Jonassen, I. (1982) *Carlsberg Res. Commun. 45*, 47–58.

Kahl, G. (1978) *Biochemistry of the wounded plant tissue*; De Gruyter, Berlin.

Lee, H., Broekaert, W. F., & Raikhel, N. V. (1991) *J. Biol. Chem. 266*, 15944–15948.

Ludvigsen, S., & Poulsen, F. M. (1992a) *Biochemistry* (second of three papers in this issue).

Ludvigsen, S., & Poulsen, F. M. (1992b) *Biochemistry* (third of three papers in this issue).

Mahoney, W. C., Smith, P. K., & Hermodson, M. A. (1981) *Biochemistry 20*, 443–448.

Mikkelsen, J., Højrup, P., Nielsen, P. F., Roepstorff, P., & Knudsen, J. (1987) *Biochem. J. 245*, 857–861.

Nielsen, P. F., Klarskov, K., Højrup, P., & Roepstorff, P. (1988) *Biomed. Environ. Mass. Spectrom. 17*, 355–362.

Sørensen, H. H., Thomsen, J., Bayne, S., Højrup, P., & Roepstorff, P. (1990) *Biomed. Environ. Mass Spectrom. 19*, 713–720.

Stanford, A., Bevan, M., & Northcote, D. (1989) *MGG, Mol. Gen. Genet. 215*, 200–208.

Svendsen, I., Martin, B., & Jonassen, I. (1980) *Carlsberg Res. Commun. 45*, 79–85.

Svensson, B., Larsen, K., & Svendsen, I. (1983a) *Carlsberg Res. Commun. 48*, 517–527.

Svensson, B., Larsen, K., Svendsen, I., & Boel, E. (1983b) *Carlsberg Res. Commun. 48*, 529–544.