COVALENT STRUCTURE OF THE GLYCOPROTEIN HORSERADISH PEROXIDASE (EC 1.11.1.7)

Karen Gjesing WELINDER

Institute of Biochemicals Genetics, The University of Copenhagen, Ø Farimagsgade 2A, DK-1353 Copenhagen K, Denmark

Received 27 September 1976

1. Introduction

In the present communication the complete amino acid sequence of a plant peroxidase is presented for the first time. The amino acid sequence of the dominating cathodic horseradish peroxidase isoenzyme is shown in fig.1. It consists of 308 amino acid residues, a hemin group, and 8 neutral carbohydrate side chains attached through asparagine residues. The molecular weight has been reported between 40 000 and 45 000, and the carbohydrate moiety constitutes about 18% [1–3]. The polypeptide chain has a molecular weight of 33 890 as calculated from its sequence.

Sequencing data have previously appeared on 25

```
                5                  10         carb    15                  20                  25
Glu-Leu-Thr-Pro-Thr-Phe-Tyr-Asp-Asn-Ser-Cys-Pro-Asn-Val-Ser-Asn-Ile-Val-Arg-Asp-Thr-Ile-Val-Asn-Glu-

               30                  35                  40                  45                  50
Leu-Arg-Ser-Asp-Pro-Arg-Ile-Ala-Ala-Ser-Ile-Leu-Arg-Leu-His-Phe-His-Asp-Cys-Phe-Val-Asn-Gly-Cys-Asp-

               55      carb        60                  65                  70                  75
Ala-Ser-Ile-Leu-Leu-Asp-Asn-Thr-Thr-Ser-Phe-Arg-Thr-Glu-Lys-Asp-Ala-Phe-Gly-Asn-Ala-Asn-Ser-Ala-Arg-

               80                  85                  90                  95                  100
Gly-Phe-Pro-Val-Ile-Asp-Arg-Met-Lys-Ala-Ala-Val-Glu-Ser-Ala-Cys-Pro-Arg-Thr-Val-Ser-Cys-Ala-Asp-Leu-

               105                 110                 115                 120                 125
Leu-Thr-Ile-Ala-Ala-Gln-Gln-Ser-Val-Thr-Leu-Ala-Gly-Gly-Pro-Ser-Trp-Arg-Val-Pro-Leu-Gly-Arg-Arg-Asp-

               130                 135                 140                 145                 150
Ser-Leu-Gln-Ala-Phe-Leu-Asp-Leu-Ala-Asn-Ala-Asn-Leu-Pro-Ala-Pro-Phe-Phe-Thr-Leu-Pro-Gln-Leu-Lys-Asp-

               155         carb    160                 165                 170                 175
Ser-Phe-Arg-Asn-Val-Gly-Leu-Asn-Arg-Ser-Ser-Asp-Leu-Val-Ala-Leu-Ser-Gly-Gly-His-Thr-Phe-Gly-Lys-Asn-

               180                 185 carb            190                 195         carb    200
Gln-Cys-Arg-Phe-Ile-Met-Asp-Arg-Leu-Tyr-Asn-Phe-Ser-Asn-Thr-Gly-Leu-Pro-Asp-Pro-Thr-Leu-Asn-Thr-Thr-

               205                 210             carb 215            220                 225
Tyr-Leu-Gln-Thr-Leu-Arg-Gly-Leu-Cys-Pro-Leu-Asn-Gly-Asn-Leu-Ser-Ala-Leu-Val-Asp-Phe-Asp-Leu-Arg-Thr-

               230                 235                 240                 245                 250
Pro-Thr-Ile-Phe-Asp-Asn-Lys-Tyr-Tyr-Val-Asn-Leu-Glu-Glu-Gln-Lys-Gly-Leu-Ile-Gln-Ser-Asp-Gln-Glu-Leu-

          carb                     260                 265         carb    270                 275
Phe-Ser-Ser-Pro-Asn-Ala-Thr-Asp-Thr-Ile-Pro-Leu-Val-Arg-Ser-Phe-Ala-Asn-Ser-Thr-Gln-Thr-Phe-Phe-Asn-

               280                 285                 290                 295                 300
Ala-Phe-Val-Glu-Ala-Met-Asp-Arg-Met-Gly-Asn-Ile-Thr-Pro-Leu-Thr-Gly-Thr-Gln-Gly-Gln-Ile-Arg-Leu-Asn-

               305
Cys-Arg-Val-Val-Asn-Ser-Asn-Ser

Disulfide bridges: 11-91, 44-49, 97-301, 177-209.
```

Fig.1. The amino acid sequence of horseradish peroxidase. carb = site of carbohydrate attachment.

of the 28 tryptic peptides of the protein [4,5] and on 120 thermolytic peptides which accounted for the complete sequence except for 5 residues (14–16 and 235–236) [6]. These data also established the sequences around the 4 disulfide bridges, and of 8 carbohydrate attachment sites which were all Asn (carbohydrate)–X–Ser/Thr, X being any residue. The α-amino terminal was found blocked in form of a pyrrolidonecarboxylyl residue L-Glu, and the C-terminal peptides were obtained in two forms, with and without C-terminal serine, suggesting lability of the asparaginyl–serine peptide bond [6]. These results have been confirmed, except that 'tryptic' peptide T5 [4] had arisen from a chymotryptic-like cleavage of a phenylalanyl–phenylalanine bond (142–143). All tryptic peptides have now been obtained, and a complete tryptic peptide map of this horseradish isoperoxidase appeared recently [7]. Here the procedures of cyanogen bromide cleavage, traditional tryptic cleavage, and cold tryptic cleavage leading to the complete primary structure of horseradish peroxidase will be summarized.

2. Materials and methods

The studied horseradish peroxidase (lot U1929, Mann Inc.) was previously characterized, and preparation of apo-peroxidase and *S*-carboxymethylated peroxidase was also described [4]. Reagents were analytical grade, Ampholine ® buffer, pH 3–10, was obtained from LKB, and urea was deionized before use. Trasylol ® trypsin inhibitor was obtained from Bayer. Digestions with trypsin (TPCK-trypsin, Worthington), α-chymotrypsin (Novo), elastase (Whatman) and carboxypeptidase A and B (DFP-treated, Worthington) were carried out in 0.2 M *N*-ethylmorpholine acetate buffer, pH 8.0, at 37°C unless otherwise stated. Pepsin was obtained from Sigma. Pyrrolidone-carboxylyl peptidase was kindly donated by Dr Doolittle [8] yeast carboxypeptidase by Dr Hayashi [9] and staphylococcal protease by Dr Drapeau [10,11].

The methods have been described previously: paper electrophoresis, peptide detection and determination of net charges in [7], gel and ion exchange chromatography in [6], details on amino acid and sequence analyses in [12].

3. Results

3.1. *Traditional tryptic digestion*

Tryptic digestion of peroxidase [4] followed by gel filtration permitted complete sequence analysis of the large half-cystine containing peptides except for one [5].

3.2. *Cyanogen bromide cleavage*

Cyanogen bromide fragments were prepared from reduced and *S*-carboxymethylated peroxidase and separated into four pools by gel filtration on Sephadex G-50 fine, in 0.05 M acetic acid. The last pool contained the pure tripeptide CN4 (residues 282–284) in 40% yield. CN5 (residues 285–308) was obtained from pool three by paper electrophoresis. The large fragments CN1 (residues 1–83) and CN2 (residues 84–181) were purified from pool two and CN3 (residues 182–281) from pool one by gel filtration on Sephadex G-100 in 8 M urea and column isoelectric focusing in 4 M urea. Table 1 lists amino acid composition, N-terminus and p*I* of each cyanogen bromide fragment and of the entire protein. The sum of the residues of the five cyanogen bromide fragments is in agreement with the amino acid composition of the protein.

CN1 was digested with trypsin, and the peptides identified by specific staining of peptide maps, amino acid composition and N-terminal analysis. The observed tryptic peptides accounted for the data of CN1 and could only be arranged in one way from the corresponding thermolytic peptides. Glutamic acid specific digestion of CN1 with staphylococcal protease provided further evidence.

Tryptic peptides of CN2 were analysed similarly and supplemented by sequence analysis of chymotryptic peptides accounting for all of CN2. Further overlaps were needed within tryptic peptide T15 (residues 125–149) and at the C-terminus.

CN3 was analysed in the same way as CN1, but two tryptic peptides (corresponding with T9a-ox2 and T14, see paragraph 4.2.) were lost as hitherto. Staphylococcal protease digestion was incomplete, but sequence analyses on small peptides provided useful overlaps. CN5 was analysed by trypsin, pepsin and carboxypeptidase A. The C-terminal serine was only partially present.

Table 1

Amino acid compositions, isoelectric points and N-termini of horseradish peroxidase and its cyanogen bromide fragments

	Cyanogen bromide fragments						Horseradish isoperoxidases		
	CN1[d]	CN2	CN3	CN4	CN5	ΣCN	U1929	C	IIIb
							Welinder et al. [4]	Shannon et al. [2,13]	Paul et al. [3]
Amino acid residues per molecule determined by amino acid analysis and sequence analysis (in brackets):									
Asx	15.6 (16)	10.3 (10)	16.8 (17)	1.0 (1)	4.1 (4)	(48)	46.5	50	46.7
Thr[a]	6.3 (6)	5.0 (5)	10.4 (11)		2.8 (3)	(25)	25.3	23	25.4
Ser[a]	6.1 (7)	8.5 (9)	6.7 (7)		1.4 (1-2)	(25)	25.7	22	25.9
Glx	4.1 (3)	6.2 (6)	9.2 (9)		1.9 (2)	(20)	19.9	19	20.7
Pro	4.0 (4)	6.1 (6)	5.6 (6)		1.0 (1)	(17)	17.1	16	17.2
Gly	3.2 (3)	7.1 (7)	4.2 (4)		3.5 (3)	(17)	17.0	15	16.9
Ala	6.0 (6)	11.7 (12)	5.3 (5)			(23)	23.0	23	22.3
Cys/Cmc[b]	1.9 (3)	3.0 (3)	0.9 (1)		0.4 (1)	(8)	7.8-7.9	4	
Val[c]	4.9 (5)	5.9 (6)	4.1 (4)		1.9 (2)	(17)	17.6	17	16
Met/Hse	0.9 (1)	1.0 (1)	1.0 (1)	1.0 (1)		(4)	3.3-4.2	3	3.9
Ile[c]	5.7 (6)	2.1 (2)	3.2 (3)		1.9 (2)	(13)	12.3	13	13.4
Leu	7.1 (6)	12.8 (13)	13.4 (14)		2.0 (2)	(35)	35.0	36	34.4
Tyr	1.4 (1)	0.0 (0)	3.3 (4)			(5)	5.9	5	5.5
Phe	6.1 (6)	6.0 (6)	7.4 (8)			(20)	20.2	23	20.2
His	1.5 (2)	1.0 (1)				(3)	2.9	3	3.0
Lys	1.2 (1)	3.0 (3)	2.0 (2)			(6)	5.8	6	6.2
Arg	6.4 (7)	7.0 (7)	4.1 (4)	1.0 (1)	2.0 (2)	(21)	19.7	21	18.4
Trp		(1)				(1)			
Total	83	98	100	3	24	308			
pI[e]	4.4	9.0	4.5				9	8.8	8.8
N-terminus	blocked	Lys	Asx	Asx	Gly		blocked	blocked	

a Values obtained after 20 h of hydrolysis are listed for CN-fragments

b Cmc = S-carboxymethylated cystein; present in CN-fragments

c Values obtained after 72 h of hydrolysis

d Contains about 15% CN3

e All values are determined by isoelectric focusing

3.3. *Cold tryptic digestion*

Tryptic digestion of 12°C of apo-peroxidase with intact disulfide bridges was stopped after 30 min by addition of trypsin inhibitor. These milder conditions were aimed at isolation of three troublesome tryptic peptides T15 (125–149), T9a-ox2 (207–224) and T14 (265–283), but were furthermore useful by providing incompletely cleaved, overlapping fragments. The digest was first fractionated by gel filtration on Sephadex G-50, and peptides T14 and T15 were obtained from low molecular weight pools by cold plate paper electrophoresis at pH 2. The first broad peak of the gel filtration was performic acid oxidized and rechromatographed yielding three separate fragments AoxI, AoxII and AoxIII. AoxI contained residues 1–149 and was digested with elastase. The results confirmed the established sequence of the N-terminal half of horseradish peroxidase. Traditional tryptic digestion of AoxII (residues 160–224) yielded five peptides, among them T9a-ox2 which was isolated by cold plate paper electrophoresis. AoxII yields the overlap from CN2 to CN3. Edman-HI sequence analysis on AoxIII (residues 265–308) provided the overlap within T14 and from CN3 to CN4 to CN5.

3.4. *N- and C-terminal analyses*

Incubation of incompletely solubilized *S*-carboxymethylated peroxidase with pyrrolidonecarboxylyl peptidase released up to 14% of pyrrolidonecarboxylic acid. Incubation of native peroxidase gave no release, but as native peroxidase dissolved in 1% SDS reacted neither with phenylisothiocyanate in 50% pyridine, nor with dansyl chloride, the native peroxidase has, most likely, a pyrrolidonecarboxylyl N-terminus which is buried. It is thus unlikely that pyrrolidonecarboxylyl of *S*-carboxymethylated peroxidase is an artefact of heme extraction carried out at acid pH. Carboxypeptidase A digestion of peroxidase derivatives proved the presence of valine, asparagine and serine at the C-terminus of peroxidase.

4. Discussion

4.1. *Which peroxidase isoenzyme?*

The horseradish peroxidase isoenzyme of which the amino acid sequence is displayed in fig.1, was obtained from Mann Inc., and its homogeneity was discussed previously [4]. Completion of the sequence determination showed no signs of mixture of isoenzymes in the form of homologous sequences, except for the previously mentioned C-terminal heterogeneity that presumably is chemically rather than genetically determined. From amino acid composition and isoelectric point, as determined by column isoelectric focusing (see table 1) the present horseradish peroxidase isoenzyme is identical to Shannon's isoperoxidase C [2,13], and to Paul's isoperoxidase IIIb [3], both being cathodic, and constituting close to 50% of total peroxidase activity in horseradish roots. The sequenced peroxidase isoenzyme is furthermore indistinguishable from two commercial preparations as tested by tryptic peptide mapping, that is the electrophoretically purified peroxidase HPOFF from Worthington Corp. [4], and the major peroxidase isoenzyme obtained by ion exchange chromatography and gel filtration of peroxidase purum, Fluka AG [7].

4.2. *Difficulties in the sequence work*

Several tryptic peptides were very difficult to obtain in good yields. In retrospect the reasons for this may be summarized as one or more of the following: (i) carbohydrate heterogeneity, (ii) few charged groups at pH 2, raising extractive loss during liquid cooled electrophoresis, (iii) unspecific cleavage at phenylalanyl–phenylalanine bonds. Peptide T14 (residues 265–283) showed all three characteristics, isolation of peptide T15 (residues 125–149) was obstructed by (ii) and (iii), and isolation of peptide T9a-ox2 (residues 207–224) by (i) and (ii) in both its carboxymethylated and performic acid oxidized forms.

Sequence work was not seriously impeded by carbohydrate side chains. Asparagine linked to neutral carbohydrate couples well with phenylisothiocyanate, but less than 10% of the thiazolinone derivative is removed upon extraction with *n*-butyl acetate. Sequencing procedures identifying these derivatives show a 'hole' at such a residue, whereas dansyl-Edman produces dansyl-aspartic acid correctly, but carries this residue as background through the following steps.

4.3. *Carbohydrate moiety*

Carbohydrate carrying sequences Asn (carbohydrate)–X–Thr/Ser are exposed on the surface of the

molecule [14]. Just one such sequence, residues 286–288, of the present horseradish peroxidase isoenzyme has apparently not been accessible to the glycosyl transferase initiating carbohydrate biosynthesis. In the present study only hexosamine analyses were carried out, but recently the carbohydrate composition of glycopeptides, obtained on trypsin and pronase digestion of isoperoxidase C, was published by Clarke and Shannon [13]. The compositions differ somewhat among the glycopeptides, but also suggest heterogeneity within each carbohydrate side chain. All compositions, however, allow for the generalized scheme suggested by Montreull [15,16] in which the residues closest to asparagine are 2 residues of *N*-acetylated glucosamine and 3 residues of mannose.

Clarke and Shannon [13] claimed some discrepancies on comparison of their glycopeptides to our sequences of carbohydrate attachment [6]. These discrepancies, I think, are largely explained by their incomplete sequence analyses. From the reported amino acid compositions, their glycopeptide 1 corresponds with residues 253–260, glycopeptide 2 with residues 186–191, glycopeptides 3A and 5 with residues 55–58 and 55–57, glycopeptide 4 with the disulfide linked residues 8–13 and 89–93, and glycopeptide 6 with residues 154–159. Glycopeptide 3B is admittedly heterogeneous and fits nowhere. The one residue difference in glycopeptide 1 at threonine residue 259 will need to be supported further before the identity of the sequenced isoperoxidase shown in fig.1 and isoperoxidase C of Shannon [2, 13] can be dismissed.

Acknowledgements

Amino acid sequence determination of horseradish peroxidase was initiated in collaboration with Drs L. B. Smillie and G. R. Schonbaum, The University of Alberta, and supported by The Medical Research Council of Canada in 1968–1971. The work has been completed at The University of Copenhagen in 1971–1974 and supported by grants from the Danish 'Statens Naturvidenskabelige Forskningsråd' (511-1141, -1651, -2845, -2949 and -3583).

References

[1] Maehly, A. C. (1955) Methods in Enzymology 2, 801–813.
[2] Shannon, L. M., Kay, E. and Lew, J. Y. (1966) J. Biol. Chem. 241, 2166–2172.
[3] Paul, K.-G. and Stigbrand, T. (1970) Acta Chem. Scand. 24, 3607–3617.
[4] Welinder, K. G., Smillie, L. B. and Schonbaum, G. R. (1972) Can. J. Biochem. 50, 44–62.
[5] Welinder, K. G. (1973) FEBS Lett. 30, 243–245.
[6] Welinder, K. G. and Smillie, L. B. (1972) Can. J. Biochem. 50, 63–90.
[7] Welinder, K. G. and Mazza, G. (1975) Eur. J. Biochem. 57, 415–424.
[8] Doolittle, R. F. (1970) Methods in Enzymology 19, 555–569.
[9] Hayashi, R., Moore, S. and Stein, W. H. (1973) J. Biol. Chem. 248, 2296–2302.
[10] Drapeau, G. R., Boily, Y. and Houmard, J. (1972) J. Biol. Chem. 247, 6720–6726.
[11] Houmard, J. and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[12] Welinder, K. G. and Mazza, G. (1976) Eur. J. Biochem. in press.
[13] Clarke, J. and Shannon, L. M. (1976) Biochim. Biophys. Acta 427, 428–442.
[14] Beintema, J. J., Gaastra, W., Scheffer, A. J. and Welling, G. W. (1976) Eur. J. Biochem. 63, 441–448.
[15] Montreuil, J. (1975) Xth Meet. Fed. Eur. Biochem. Soc., Abstract 956, Paris.
[16] Montreuil, J. (1975) Pure Appl Chem. 42, 431–477.