# AMINO ACID SEQUENCE OF HORSE SPLEEN APOFERRITIN

Michel HEUSTERSPREUTE and Robert R. CRICHTON
*Unité de Biochimie, Université Catholique de Louvain, Place L. Pasteur, 1, 1348 Louvain-la-Neuve, Belgium*

Received 8 May 1981

## 1. Introduction

Iron is essential for the metabolism and growth of virtually all living organisms. However, the solution chemistry of iron, which, in aqueous media at physiological pH-values, is dominated by the tendancy to oxidise and hydrolyse with formation of insoluble ferric hydroxides and oxyhydroxides has necessitated the development of specialised transport and storage systems for iron in biological systems. At the microbial level ferric iron chelators of high affinity (siderophores) are synthesized and excreted into the extracellular medium where they complex iron; the iron–siderophore complex is taken up by a receptor-mediated mechanism by the microbial cells, and once inside the cell the iron is released from the chelator, a process often accompanied by destruction of the ligand [1,2]. In mammalian cells iron is transported in the serum and in the extravascular space by transferrin, a $\beta_1$-globulin. Intracellular iron storage in many different cell types (including animals, plants, fungi and bacteria such as *Azotobacter vinelandii*) is assured by a specialised protein, ferritin which consists of a hydrolysed polymeric ferric iron core deposited in the interior of a hollow protein shell (apoferritin), which maintains the iron in a soluble, non-toxic and bio-available form, from which iron can be mobilised as and when required by the cell.

The best characterised mammalian iron storage protein is the ferritin of horse spleen. First isolated by Laufberger* [3] it has been the object of extensive studies since then, and remains the standard preparation for studies on iron storage and of storage iron deposition and mobilisation. The iron-free protein shell, apoferritin, has $M_r$ 440 000–480 000 [4–7] and is an oligomer composed of 24 polypeptide chains (subunits) for which 18 000–19 000 $M_r$ was determined by physico-chemical techniques [7,8]. The apoferritin molecule is roughly spherical, of diameter 12–13 nm, with a central cavity of diameter 6–7 nm, in which the iron core is deposited, mostly as FeO . OH, but with some associated phosphate [9,10]. X-ray crystallographic studies have resulted in an electron density map of the apoferritin subunit at a nominal resolution of 0.28 nm [11,12]. The subunit can best be described as a parallel bundle of four α-helices reminiscent of a number of other proteins (haemerythrin, cytochrome $c'$, tobacco mosaic virus, myohaemerythrin, cytochrome $b_{562}$ [12]) with two shorter helices situated perpendicular to them, interconnected by non-helical regions.

We have been engaged in the determination of the primary structure of horse spleen apoferritin for some time [13,14] and we present here the complete primary structure of the horse spleen apoferritin subunit. Some conclusions are presented in terms of the sequence and its implications for three-dimensional studies.

* This article is dedicated to Professor Vilem Laufberger of the Charles University, Prague, Czechoslovakia

## 2. Materials and methods

Ferritin was isolated from horse spleen by a modification of the procedure in [15] involving acidification of a tissue homogenate to pH 4.75 followed by $(NH_4)_2SO_4$ (35%, w/v) precipitations and several recrystallisations from $CdSO_4$ (0.3 M) containing 0.4 M NaCl [16]. The homogeneity of the preparations was controlled by polyacrylamide gel electrophoresis in the presence and absence of sodium dodecyl sulphate and by amino acid analysis after hydrolysis for 16 h in vacuo with 6 N HCl at 110°C, using a Locarte amino acid analyser (Locarte Co., London). Apoferritin was prepared as in [16].

Prior to cleavage with trypsin, staphylococcal protease and chymotrypsin, apoferritin was reduced and carboxymethylated [17] and the lysine residues citraconylated [18] in 7 M guanidine. These digests were fractionated by gel filtration on Sephadex G-50 in 0.5% (w/v) $NH_4HCO_3$ and the smaller peptides were further purified by ion-exchange chromatography and by preparative paper chromatography [16]. Since an Asp–Pro sequence was found in the staphylococcal peptide S12 we cleaved the carboxylmethylated protein in 70% formic acid (v/v) for 4 days at 37°C and isolated the two fragments P1 (residues 1–122) and P2 (residues 123–174) by gel filtration on Sephadex G-75 in 1% (v/v) $NH_3$ after citraconylation of their lysine residues [18]. The fragment P1 was digested with trypsin after blocking the arginine residues with cyclohexanedione [19] in order to obtain lysyl peptides from this N-terminal fragment. P1 and P2 were digested with staphylococcal protease. The products of these three digests were purified by gel filtration on Sephadex, followed by ion-exchange chromatography and where necessary preparative paper chromatography [16]. Carboxymethylated apoferritin was cleaved by CNBr in HCOOH 70% (v/v) after incubation with 3 mM dithiothreitol to reduce methionine sulphoxide to methionine [20]. The digest was fractionated by gel filtration on Sephadex G-75 in 1% $NH_3$ (v/v). The C-terminal CNBr peptide CN4 was digested with trypsin and the resulting peptides purified by ion-exchange chromatography.

The purity of peptides was established by N-terminal analysis using the dansyl chloride procedure [21] and by amino acid analysis. Large peptides (>20 residues) were sequenced by liquid phase Edman degradation in a Beckman 890 sequencer using the programme of [22]. After conversion of the thiazolinones to PTH amino acids identification of the latter was carried out by high-pressure liquid chromatography [22], by gas–liquid chromatography [23] and by thin-layer chromatography [24]. The smaller peptides were sequenced by the dansyl-Edman method [25] and the dansyl amino acids were identified by thin-layer chromatography [18].

## 3. Results

The complete primary structure of horse spleen apoferritin was deduced from the sequence of peptides obtained from the chemical and enzymatic cleavages in section 2. In addition to the peptides shown here (the chymotryptic peptides are not included in fig.1) a large number of peptic peptides were also useful in confirming the sequence given in fig.1. We present here the sequence together with its justification.

The N-terminal arginyl peptide T1 was already known [26]. We isolated a ninhydrin-negative, arginine-positive peptide of this composition as well as a staphylococcal protease peptide which has a composition corresponding to residues 1–11 and which on trypsin digestion gave the N-terminal blocked peptide and the sequence 6–11. A peptic peptide (4–11) confirms the overlap of T1 and T2. The arginine residues in positions 18 and 25 were only partially cleaved, while the Arg–Asp bond 39–40 was resistant to tryptic cleavage.

The peptides T2–T4 were isolated as an aggregated mixture and were sequenced as such in liquid phase. Their sequence as well as their order was confirmed by the peptides S3a and S3b (also obtained and sequenced as a mixture from staphylococcal protease digestion of P1, presumably resulting from incomplete cleavage of the Glu–Ala sequence 13–14) and by a number of peptic and chymotryptic peotides (not shown). The peptides S4 and S5 are important since they permit the alignment of T4–T6. A number of peptic and chymotryptic peptides were also isolated and sequenced in the region and confirm the sequence up to Leu 65. A peptic peptide was isolated and sequenced corresponding to the sequence from Glu 57–Leu 65. The juxtaposition of T6 and T7 thus depends solely on the Leu at position 65, but this is the only order of the peptides that we have isolated which is consistent with the complete sequence. The peptide TC1, obtained from tryptic digestion of P1 at lysine residues, after modification of the arginine residues by cyclohexanedione, allow us to align T7–T9. The sequence of T9, determined by liquid phase degradation to the position 104 is confirmed by the staphylococcal protease peptides S8–S10 and by the CNBr peptide CN3. The continuation of the sequence is assured by CN3 and by a tryptic peptide TC2 resulting from the same disgest of P1 as TC1. A tryptic peptide T9b, resulting from cleavage of Lys 104 (which is presumably not very well citraconylated) confirms the sequence of CN3 and the alignment of S11 and S12. The C-terminal sequence is established from Pro 123 by the peptide P2 up to Glu 163. This sequence itself is confirmed

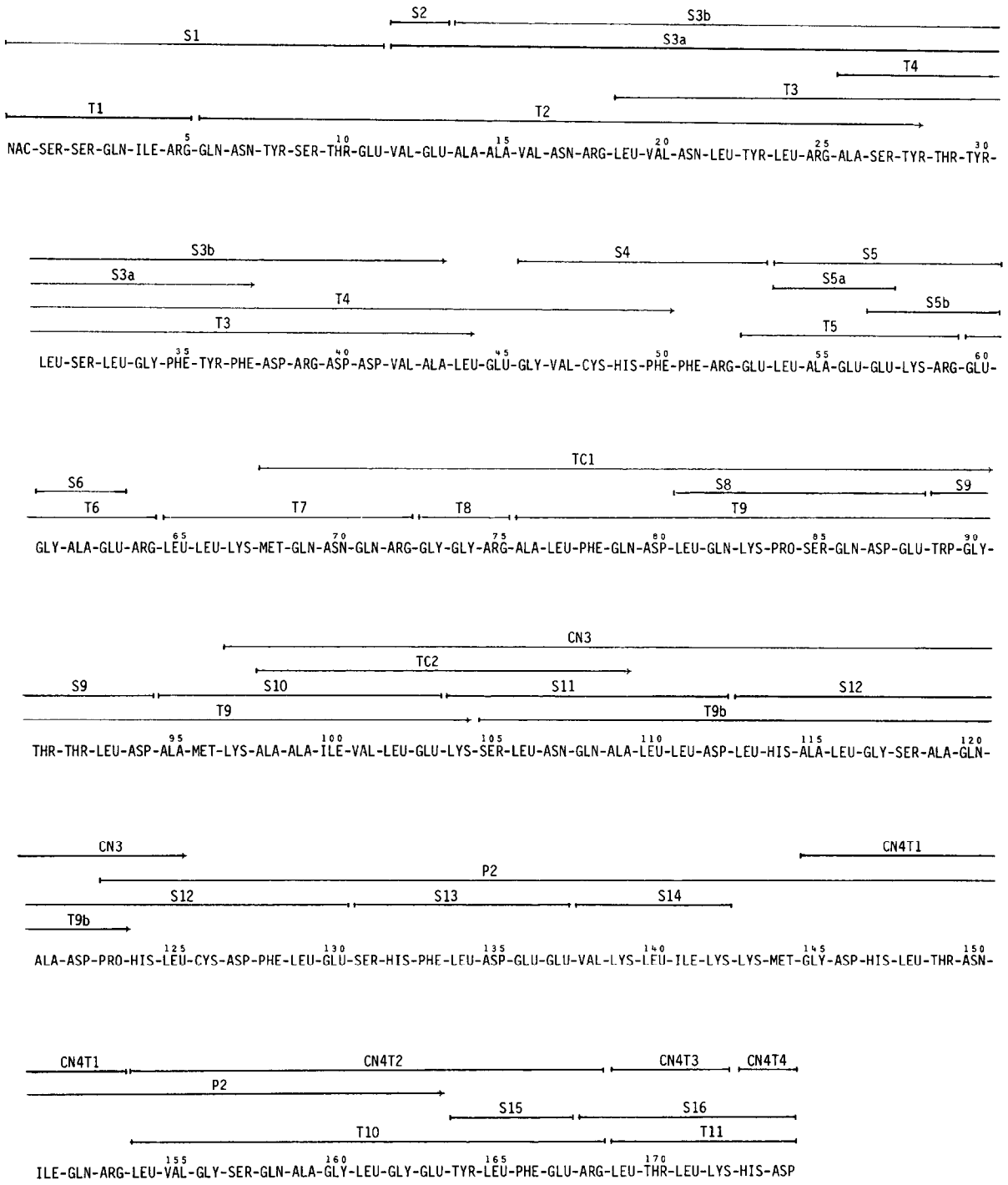


Fig.1. Amino acid sequence of horse spleen apoferritin. The figure includes the localisation and sequences of tryptic (T), staphylococcal protease (S) peptides and of the CNBr peptide (CN3) and the C-terminal peptide resulting from acid cleavage at residues 122–123 (P2). The sequences determined on 2 peptides resulting from tryptic cleavage of the fragment P1 after modification of the arginine residues, TC1 and TC2 are also indicated as are the sequences of the 4 tryptic peptides obtained from cleavage of the CNBr peptide CN4 (CN4T1, etc.). A large number of peptic and chymotryptic peptides (not shown) confirm the sequence.

Table 1
Amino acid composition of horse spleen apoferritin

| Amino acid | No. residues/subunit | |
|---|---|---|
| | (1) | (2) |
| Cys | 2 | 3.0 |
| Asx | 18 | 18.5 |
| Thr | 6 | 5.9 |
| Ser | 10 | 9.6 |
| Glx | 26 | 26.0 |
| Pro | 2 | 3.0 |
| Gly | 11 | 10.7 |
| Ala | 15 | 15.0 |
| Val | 8 | 7.4 |
| Met | 3 | 3.0 |
| Ile | 4 | 3.7 |
| Leu | 28 | 26.4 |
| Tyr | 6 | 5.3 |
| Phe | 8 | 7.8 |
| Trp | 1 | 2.1 |
| His | 6 | 6.2 |
| Lys | 9 | 9.3 |
| Arg | 11 | 10.2 |

The amino acid composition is given as residues of each amino acid per subunit of molecular weight 19 824; Column (1) calculated from the amino acid sequence, column (2) from ref. (8)

by the staphylococcal protease peptides S12–S14 and by peptic peptides covering the sequence 128–140. From Gly 145 the tryptic peptides CN4T1–CN4T4 as well as the arginyl peptides T10 and T11 and the staphylococcal protease peptides S15 and S16, as well as a number of peptic and chymotryptic peptides confirm the C-terminal sequence of the protein.

## 4. Discussion

We have established the complete amino acid sequence of the protein subunit of horse spleen ferritin and in the light of this result we consider here some of its implications. In the first place, we can establish that, compared with the amino acid composition (table 1) published some 10 years ago [8] there is excellent agreement for all amino acids except Cys, Pro and Trp, where in all cases the amino acid analysis indicated one residue more than we find in the sequence. This is perhaps not surprising in the case of Cys and Pro, but the result for Trp is disturbing, since it is generally accepted that the molar extinction coefficient of apoferritin at 280 nm (on the basis of the molecular mass calculated from the sequence) is 19 500 which is in flagrant disaccord with the value calculated from the content Trp, Tyr and Cys in the sequence (13 610). It is possible that this difference can be attributed to the presence of a UV-absorbing cofactor (perhaps involved in iron reduction) with apoferritin, which is carried through the isolation procedure [27,28].

The apoferritin subunit based on physico-chemical determinations [7,8] was estimated at $M_r$ 18 500. This is 7% less than the value established on the basis of the sequence, namely $M_r$ 19 824; however the apoferritin oligomer (24 subunits) calculated from this data is $M_r$ 476 000 which is in good agreement with X-ray crystallographic studies [4].

We can also draw some conclusions on the reasons which have led to such a long delay in establishing the amino acid sequence of this protein. Since 9 of the 11 arginine residues are located in the first 75 residues whereas the last 2 are in positions 153 and 168, cleavage of the protein by trypsin after citraconylation of the lysine residues (which proved to be essential in order to get reasonable yields of peptides, as for staphylococcal protease and chymotryptic digestion) leaves us with a large region of the molecule for which no overlaps with other digests were possible except in the N-terminal region (residues 76–104). Further it proved extremely difficult to separate the large arginyl peptide T9 from the highly aggregating T2–T4 mixture. The use of CNBr cleavage was also handicapped by the fact that the largest CNBR peptide CN1 (residues 1–68) is N-terminal blocked by an acetyl group, CN2 (residues 69–96) is of almost the same size as CN4 (residues 144–174), and the N-terminal Gln tends to cyclise rapidly, and CN3 (residues 97–143) aggregates as do all of the CNBr peptides and is difficult to separate from CN1. We were able to obtain all of the CNBr peptides in pure form towards the end of the sequence determination and the N-terminal sequence of CN3 [14] and CN4 [14] as well as the tryptic peptides of CN4 were helpful in confirming the C-terminal sequence of the subunit.

The distribution of aromatic and hydrophobic amino acid residues is quite interesting. Thus 9 of the 15 aromatic residues are present in the N-terminal sequence from Tyr 8–Phe 51 and 15 of the Ala, Val, Leu and Ile residues of a total of 55 are also present in this sequence. There is also a tendency for aromatic

residues to be present in clusters either together or separated by one other residue. We have already commented on the somewhat unusual distribution of Arg in the sequence. However, we may also note a rather pronounced tendency for charged residues to occur in clusters, the most notable being the sequence Glu–Glu–Lys–Arg–Glu (residues 54–60) and Asp–Glu–Glu (residues 135–137). It should be pointed out that if the sequence 53–63 were to be in an α-helical conformation then Glu 53 and Glu 56 as well as Glu 57, Glu 60 and Glu 63 would be disposed on the same face of the helix, forming 2 potential metal binding sites. The same would be true of Asp 127 and Glu 130. Since carboxyl groups have been implicated from chemical modification studies in iron binding and in the catalysis of iron oxidation and deposition by the protein [29] and since a part of the putative catalytic (iron oxidation) site has been isolated, identified as the peptide 53–59 [30] it will be interesting to see whether the X-ray analysis of the structure confirms that the site of initial iron oxidation is in this region of the molecule.

The total number of amides is 17 (6 Asn and 11 Gln) giving a total of 27 β and γ carboxyls (12 Asp and 15 Glu). This is in good agreement with the number found by measurement of $NH_3$ liberated on acid hydrolysis and with the results of chemical modification [29,31] although the results in [31] are in much better accord than our own data [18,29].

On the basis of the primary structure we have attempted to predict the secondary structure of the protein using the methods in [32]. The results indicate α-helix from residues 9–18, 39–48, 50–72, 74–83, 93–119, 128–147 and 163–172 with β-turns at positions 7–11, 38–43, 70–76, 84–90 and 123–128 together with some evidence for β-pleated sheets in positions 2–10, 15–39, 46–52 and 147–159.

The correlation of the amino acid sequence with the three-dimensional structure [11,12] will be discussed in greater detail in [33]. We note that the primary structure presented here seems to be most compatible with one of the two connectivities proposed by the Sheffield group in which the helices are joined in the order A–B–(loop)–C–D–P–E [12]. This order is also in good agreement with the secondary structure predictions since, with the exception of the prediction of only a small part of helix A, and the lack of helix P, helices B–E are predicted at about the positions expected from the X-ray data. The fitting of the sequence to the electron density map calculated at a resolution of 0.28 nm [12] is well advanced and will be reported in [33].

**Acknowledgements**

We thank Professor Pauline Harrison and her colleagues in Sheffield for many useful discussions, Dr Paul Falmagne and his colleagues of the Université d'Etat à Mons for their help in the liquid phase sequence determinations and Francine Roland, Karin Schanck, Willy Vancaster and Annette Halloy for their technical and secretarial assistance. The support of the 'Institut pour l'Encouragement de la Recherche dans l'Industrie et l'Agriculture' (IRSIA) to M. H. and of the FRFC are gratefully acknowledged.

**References**

[1] Nielands, J. B. (1977) CIBA Found. Symp. 51, 107–124.
[2] Raymond, K. N. and Carrano, C. J. (1979) Acc. Chem. Res. 12, 183–190.
[3] Laufberger, V. (1937) Bull. Soc. Chim. Biol. 19, 1575–1582.
[4] Rothen, A. (1944) J. Biol. Chem. 152, 679–693.
[5] Harrison, P. M. (1959) J. Mol. Biol. 1, 69–80.
[6] Crichton, R. R., Eason, R., Barclay, A. and Bryce, C. F. A. (1973) Biochem. J. 131, 855–857.
[7] Björk, I. and Fish, W. W. (1971) Biochemistry 10, 2844–2848.
[8] Bryce, C. F. A. and Crichton, R. R. (1971) J. Biol. Chem. 246, 4198–4205.
[9] Harrison, P. M. (1963) J. Mol. Biol. 6, 404–422.
[10] Michaelis, L., Coryell, C. D. and Granick, S. (1943) J. Biol. Chem. 148, 463–481.
[11] Banyard, S. H., Stammers, D. K. and Harrison, P. M (1978) Nature 271, 282–284.
[12] Clegg, G. A., Stansfield, D. K., Bourne, P. E. and Harrison, P. M. (1980) Nature 288, 298–300.
[13] Crichton, R. R. and Barbirolli, V. (1970) FEBS Lett. 6, 134–136.
[14] Collet-Cassart, D. (1977) Doctoral thesis, Université Catholique de Louvain.
[15] Granick, S. (1942) J. Biol. Chem. 146, 451–461.
[16] Heusterspreute, M. and Crichton, R. R. (1981) in preparation.
[17] Rüegg, W. T. and Rudinger, J. (1977) Methods Enzymol 47, 111–116.
[18] Heusterspreute, M. (1981) Doctoral thesis, Université Catholique de Louvain.
[19] Smith, E. L. (1977) Methods Enzymol. 47, 156–161.
[20] Drapeau, G. R. and Yanofsky, C. (1967) J. Biol. Chem. 242, 5434–5441.
[21] Gray, W. R. (1972) Methods Enzymol. 25, 121–138.

[22] Hunkapillar, M. W. and Hood, L. E. (1978) Biochemistry 17, 2124–2133.
[23] Pisano, J. J. and Bronzet, T. J. (1969) J. Biol. Chem. 244, 5597–5607.
[24] Kulbe, K. D. (1974) Anal. Biochem. 59, 564–573.
[25] Gray, W. R. (1967) Methods Enzymol. 11, 139–151.
[26] Suran, A. A. (1966) Arch. Biochem. Biophys. 113, 1–4.
[27] Mareschal, J. C., Mingeot, M. P. and Crichton, R. R. (1980) 13th FEBS Meeting, Jerusalem 1980, abstr.
[28] Crichton, R. R., Rolin, D. and Mareschal, J. C. (1981) in preparation.
[29] Wetz, K. and Crichton, R. R. (1976) Eur. J. Biochem. 61, 545–550.
[30] Vandamme, E. (1980) Doctoral thesis, Université Catholique de Louvain.
[31] Silk, S. T. and Breslow, E. (1976) J. Biol. Chem. 251, 6963–6973.
[32] Chou, P. Y. and Fasman, G. D. (1974) Biochemistry 13, 211–245.
[33] Harrison, P. M., Crichton, R. R. et al. (1981) in preparation.