Biol Metals (1991) 4:95–99

BIOLOGY OF METALS

The location of exon boundaries in the multimeric iron-storage protein ferritin

Pauline M. Harrison, Geoffrey C. Ford, John M. A. Smith, and Janice L. White

Krebs Institute for Biomolecular Research and Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, England

Received November 23, 1990

Summary. The nature of the amino acids whose codons border introns in ferritin genes is novel; the disposition of these intron boundaries within the three-dimensional structure of the 24-subunit molecule differs significantly from that of other proteins. These observations are discussed in relation to the functions of isoferritins.

Key words: Exon boundaries – Ferritin – Multimeric protein

Introduction

Ferritin molecules are protein shells storing an iron-containing mineral in their cavities. Their 24 subunits are structurally equivalent but the polypeptide fold can accommodate two distinct chains with differing amino acid sequences, known as H and L chains; the chains can assemble into the 24-mer either singly or as co-polymers. The three-dimensional structures of horse spleen apoferritin (85% L, 15% H) (Rice et al. 1983; Ford et al. 1984), rat liver apoferritin (65% L, 35% H) (Lawson 1990), and recombinant rat-liver L chain (Lawson 1990; Thomas et al. 1988) and human H chain (Lawson 1990) homopolymers are isomorphous. Nevertheless the two classes of polypeptide show only about 47% identify for H versus L amino acid sequences as compared with about 85% identity within H sequences [human (Costanzo et al. 1986), rat (Murray et al. 1987), chicken (Stevens et al. 1987) and mouse (Miyazaki et al. 1988; Torti et al. 1988; Beaumont et al. 1989)] and about 79% identity within L sequences [human (Santoro et al. 1986), rat (Leibold and Munro 1987), horse (Heusterspreute and Crichton 1981), rabbit (Daniels-McQueen et al. 1988) and mouse (Beaumont et al. (1989)]. Nucleotide sequences are known for H genes of human (Constanzo et al. 1986), rat (Murray et al. 1987) and chicken (Stevens et al. 1987) and for L genes of human (Santoro et al. 1986) and rat (Leibold and Munro 1987). A striking feature of these sequences is the constant position of genomic intron splice junctions when examined with respect to the encoded protein's chain fold and quaternary structure (although the introns are of different lengths). Here these positions are displayed and discussed in relation to structural subdomains and to isoferritin function. In addition, cDNA sequences have been determined for mature mRNAs of rabbit L chains (Daniels-McQueen et al. 1988), mouse H chains (Miyazaki et al. 1988; Torti et al. 1988; Beaumont et al. 1989) and L chain (Beaumont et al. 1989) and for three different types of tadpole red cell (Dickey et al. 1987) ferritin [named H, M, and L, from their relative electrophoretic mobilities although all three more closely resemble the other known H chains (61–71% identity) than L chains (48–59%)]. Analysis of the nucleotides and amino acids in these ferritins suggests that the intron splice boundries are likely to occur in positions equivalent to those observed directly in gene sequences.

The conformation of a single horse spleen ferritin L subunit is shown in Fig. 1, together with its two-fold axis related neighbour and parts of adjacent chains within the protein shell. The positions of introns in the five known genomic H and L sequences are shown in Fig. 1. With respect to the horse L primary structure, the intron splice junctions fall between residues 33 and 34, 82 and 83, and 124 and 125 (and not within codons). The first delimiter is placed towards the C-terminus of helix A (which comprises residues 9–39), the second near the centre of the long loop L (residues 74–90) and third near the N-terminus of helix D (residues 122–154). Thus, broadly speaking, the exons correspond to four sub-domains: (1) the N-terminus and most of helix A; (2) the C-terminus of helix A, the AB turn, helix B (residues 44–73) and half of loop L; (3) the C-terminus of L and the helix C (residues 91–121); and (4) helix D (residues 122–154), the DE turn, helix E (residues 158–171) and a C-terminal extension. Analysis of other proteins (Gilbert 1978; Blake 1978; Artymiuk et al. 1981;

Offprint requests to: P. M. Harrison

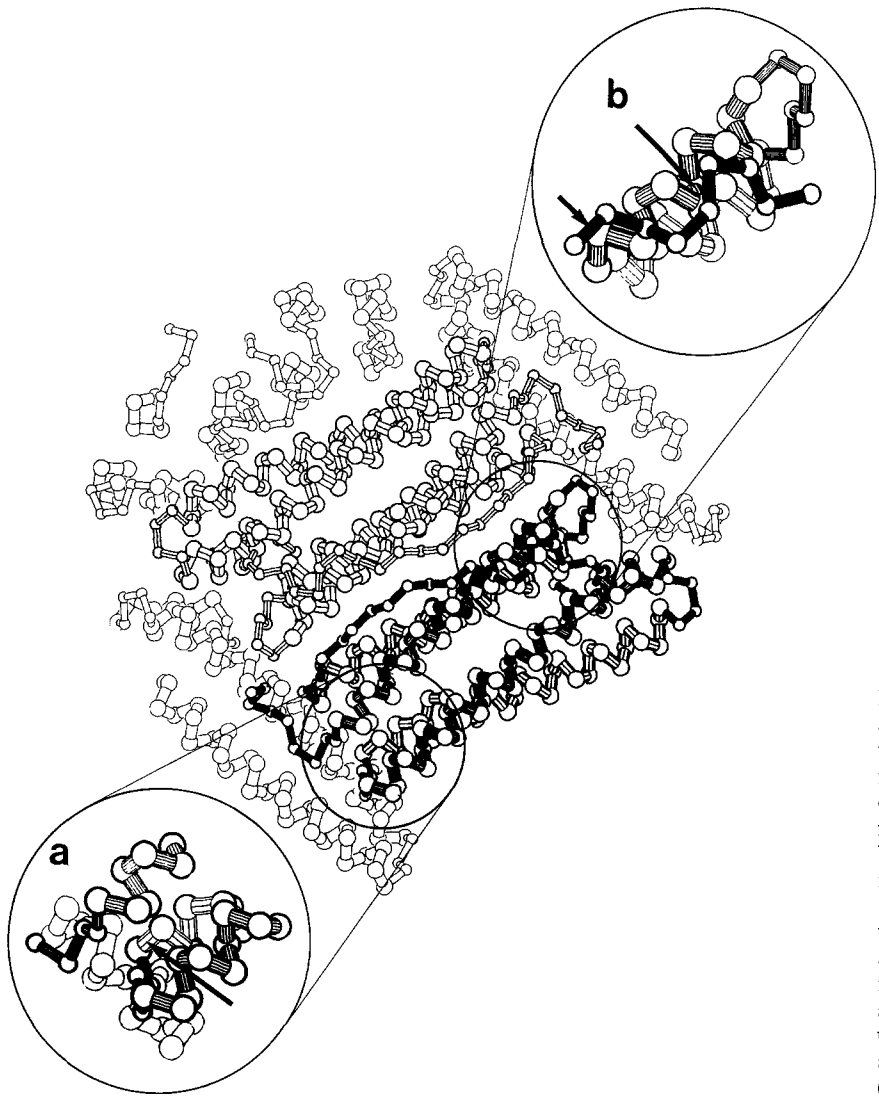

Fig. 1. Part of the protein shell of horse spleen apoferritin (Ford et al. 1984) around one of its twofold axes. Regions of the shell containing intron splice junction positions (arrows) are circled and are reproduced in the enlarged insets. Intron splice junction 1 occurs in a buried intrasubunit position and junction 2 near the centre of a long external loop (inset **b**). Splice junction 3 (inset **a**) is buried within the shell (residues on either side of the junction are either on a subunit interface making several interactions with residues from neighbouring subunits or buried within the subunit). The three-dimensional structures of human H chain and rat liver L chain ferritins are very similar (Lawson 1990)

Gilbert et al. 1986; Go and Nosaka 1987) suggests that introns may mark structural modules or domains. Ferritin seems to fall within this pattern. Domains or modules are frequently separated by coil or loop regions and these are often found on the surface of the folded protein. Hence such inter-domain exon/intron splice junctions may be expected to occur between surface residues (Craik et al. 1982). In the large dimeric enzyme glycogen phosphorylase, although the 19 splice junctions do not necessarily mark structural or functional subdomains, 17 of them do occur on surface positions of the folded dimeric molecule (Burke et al. 1987). This could be explained by a bias in the distribution of nucleotides around the splice junction. Craik et al. (1982) analysed 31 splice junctions in 10 protein chains of known gene sequence and three-dimensional structure; they found that there was no restraint on the amino acid following the 3′ splice junction, but on the 5′ side of the 5′ junction, hydrophilic residues were over-represented and hydrophobic residues under-represented. A consensus sequence [C or A] A G / G T based on the analysis of 400 exon/intron boundaries in vertebrate genes (Padgett et al. 1986) leads to the same conclusion. Most of the proteins analysed (Craik et al. 1982; Burke et al. 1987) were either monomers or oligomers of small numbers of subunits.

In ferritin, however, only two of the intron delimiters occur on or near the subunit surface, 82/83 and 124/125, and only one, 82/83, on the surface of the molecule. None lies on the surface of the internal cavity. Residues 33 and 34 are buried within the subunit and within the shell. Residue 125 is buried within the subunit and residue 124, although on the subunit surface, makes a number of interactions with neighbouring subunits. Table 1 lists the amino acid residues on either side of the splice junction of known genomic sequences and the residues that occur in equivalent positions of other ferritin H and L polypeptides. Table 2 shows the exon lengths. Table 3, together with Fig. 1, locates the residues in the protein shell and gives the side-chain accessible surface in a folded but isolated subunit and in a complete 24-mer. The high degree of conservation of residues on either side of the splice junction (which fall between residue codons in all cases) is notable for 124/125 (residues involved in inter-subunit interactions) and to a lesser extent for the other two junctions, al-

Table 1. Location of intron splice boundaries in ferritins

Source	Amino acid sequences			Corresponding codons		
	33/34	82/83	124/125	33/34	82/83	124/125
HuL-L	SL/GF	IK/KP	PH/LC	CTG/GGC	AAG/AAG	CAT/CTC
RaL-L	SL/GF	VQ/KP	PH/LC	CTG/GGC	CAG/AAG	CAC/CTC
MoE-L	SL GF	VQ KP	PH LC	CTG GGC+	CAG AAG+	CAT CTC+
HoS-L	SL GF	LQ KP	PH LC			
RbL-L	SL GF	VQ KP	PH LC	CTG GGC+	CAG AAG+	CAC CTC+
HuL-H	SM/SY	IK/KP	PH/LC	ATG/TCT	AAG/AAA	CAT/TTG
RaL-H	SM/SC	IK/KP	PH/LC	ATG/TCT	AAG/AAA	CAC/TTA
MoM-H	SM SC	IK KP	PH LC	ATG TCT+	AAG AAA+	CAC TTA+
ChR-H	SM/SY	IK/KP	PH/LC	ATG/TTC	AAG/AAA	CAC/TTG
TaR-L	SM AS	VE KP	PH MT	ATG GCC+	GAG AAA+	CAC ATG+
TaR-M	SM YA	IK KP	PH LC	ATG TAT+	AAG AA+	CAT CTG+
TaR-H	SM AF	VK KP	PH LC	ATG GCT+	AAG AAA+	CAT CTG+

Splice junctions are indicated by / and, in all cases, are located between amino acid codons in the mature mRNA. Possible junctions that are suggested by homology arguments are indicated by a space. Codon sequences that were determined for the cDNA copy of mature mRNA are flagged by +. Ferritin sequences are as follows: HuL-L, human liver L chain (Santoro et al. 1986); RaL-L, rat liver L chain (Leibold and Munro 1987); MoE-L, mouse erythroid cell L chain (Beaumont et al. 1989); HoS-L, horse spleen L chain (Heusterspreute and Crichton 1981); RbL-L, rabbit liver L chain (Daniels-McQueen et al. 1988); Hul-H, human liver H chain (Costanzo et al. 1986); RaL-H, rat liver H chain (Murray et al. 1987); MoM-H, mouse macrophage H chain (Miyazaki et al. 1988); ChR-H, chicken reticulocyte H chain (Stevens et al. 1987); TaR-L, -M, -H, tadpole reticulocyte L, M, H chain (Dickey et al. 1987)

Table 2. Exon lengths in ferritin sequences

Source	Length of			
	exon 1	exon 2	exon 3	exon 4
HuL-L	33	49	42	50
RaL-L	33	49	42	58
MoE-L	33+	49+	42+	58+
HoS-L	33+	49+	42+	50+
RbL-L	33+	49+	42+	60+
HuL-H	37	49	42	54
RaL-H	37	49	42	53
MoM-H	37+	49+	42+	53+
ChR-H	37	49	42	52
TaR-L	34+	49+	42+	48+
TaR-M	34+	49+	42+	51+
TaR-H	34+	49+	42+	51+

Ferritin sequences are referenced in Table 1. Exons postulated by homology are flagged by +

Table 3. Location of splice junction residues in apoferritin subunit

Residue	Radius vector (nm)	Accessible surface area (nm^2) in			
		monomer	dimer	24-mer	extended peptide
Leu 33	5.2	0.01	0.01	0.01	1.70
Gly 34	4.0	0	0	0	0.75
Gln 82	6.1	1.27	1.02	1.02	1.80
Lys 83	6.0	1.30	0.66	0.66	2.00
His 124	5.2	0.90	0.90	0.09	1.95
Leu 125	5.3	0	0	0	1.70

Accessible surface areas of residues in an isolated extended peptide are taken from Chothia (1976). Accessible surface areas of residues in apoferritin sub-structures, an isolated monomer, a twofold related dimer, and the complete 24-subunit protein shell, were calculated by program DSSP of Kabsch and Sander (1983). The radius vector for each residue is the distance from the centre of the complete ferritin molecule to the α-carbon atom of that residue

though there is considerable conservation within either H or L chains. Next-but-one residues are also generally conserved or conservatively replaced. This suggests that splice junctions occur in equivalent positions in all of the ferritins listed in Table 1. At the 5′ side of splice junctions Craik et al. (1982) noted the frequent occurrence of lysine, arginine, glutamate, glutamine and aspartate residues and the rare occurrence of hydrophobic residues. In ferritin, we find leucine or methionine at position 33, in keeping with its buried position within the subunit; lysine, glutamine or glutamate at position 82 and histidine at position 124, a residue not found in any of the 31 junctions listed in Craik et al. (1982), the 19 junctions in Burke et al. (1987) or in the consensus sequences of Padgett et al. (1986). At the 3′ side of the junction at position 34, as in other proteins, we find a wide variability of residue type. At position 83, the high degree of conservation of Lys83 is unexpected as the side chain points into solution, whereas at position 125 the conservation of Leu125 reflects its involvement in many hydrophobic interactions. The codons used are also listed in Table 1 for both genomic and cDNA sequences. It is notable that these codons tend to be of the same type for residues in the same positions in different seqences whereas at other positions different codons may be used. For example, the L chain has codon CTG for leucine at position 33 for human, rat and rabbit whereas CTC codes for leucine at position 125 in all three species. On the other hand, TTA or TTG is used for Leu125 in H-chain sequences.

The codon for Lys83 is AAA for all H chains but AAG for L chains. Examination of the intron sequences shows that although their lengths are quite variable, the usual pattern of bases (Breathnach et al. 1978), namely GT at the 5′ side and AG at the 3′ side of the junction, is maintained. The junction 82/83 cuts the molecule into two pieces of similar length, but there is no evidence for gene duplication in the amino acid sequence. It is notable that exons 2 and 3 are of identical lengths in all ferritin sequences (Table 2); this may reflect their importance in the conservation of three-dimensional structure. The number of residues in exon 1 is variable, but only in the N-terminal region (9–13 residues) before helix A and this lies on the outside of the protein shell. Similarly exon 4 varies only in the number of C-terminal residues that extend into the cavity, except for the L chains of rat and mouse in which an eight-residue insertion is found in the DE turn on the outside of the molecule, making this inter-helical region eleven residues in length insteads of the three residues of the other ferritins.

Assignment of function among the four exons cannot be done with certainty at present. Ferritin molecules accelerate Fe(II) oxidation (Neiderer 1970; Bakker and Boyer 1986; Macara et al. 1973; Bryce and Crichton 1973) and promote nucleation of the mineral ferrihydrite in their central cavity (Macara et al. 1972). Recent work with recombinant ferritins shows that rapid oxidative activity is confined to the H chain although ferritins containing only L chains are able to sequester iron and form iron cores slowly at pH 7 (Lawson 1990; Thomas et al. 1988; Levi et al. 1988; Levi et al. 1989; Lawson et al. 1989). The site of Fe(II) oxidation is at a metal centre recently discovered in the crystal structure of recombinant human H-chain ferritin and shown to account for ferroxidase activity by following the effects of specific sequence changes (Lawson 1990; Lawson et al. 1989). This metal site has ligands from both exon 1 and exon 2 (residues 23, 58, and 61 based on the numbering system used here). The second and third of these lie on the B helix of exon 2. Glu57, which may be involved in moving iron from the oxidation centre to the central cavity, also lies within exon 2. Both H and L chains probably contain nucleation sites on the inside surface of the cavity. These may also be located on the B helices of exon 2 (Glu57, Glu61 and Glu64 have been suggested as ligands) although this is uncertain. However, two other residues near the ferroxidase centre of H chains which may also be essential for activity are Glu103 and Glu137 placed, respectively, on the C and D helices (exons 3 and 4), so that residues essential for rapid ferrihydrite production in H ferritins are not confined to a single exon. The D helix, lying almost entirely within exon 4, is also situated on the inside of the shell structure. The absence of any intron junctions on the internal surface of the cavity may be related to its functional importance in iron storage. Of the two types of inter-subunit channel that pierce the shell and may allow access for iron or small molecules, the residues lining the largely hydrophobic fourfold channel lie entirely in exon 4. The narrow part of the hydrophilic threefold channels are bounded by residues Asp127 and Glu130, lying within exon 4. However, the outer, wider part of these channels is edged mainly by residues near the C-terminus of exon 3. The intron splice junction 124/125 also lies close to this region. Exons 1 and 2, in addition to supplying ferroxidase centre ligands in H chains, are largely responsible for the stable inter-subunit interface around twofold axes shown in Fig. 1 that plays an essential role in subunit assembly (Rice et al. 1983; Ford et al. 1984).

Sequence conservation among the various exons in H and L chains may reflect their divergence to different functions. Thus in their common regions exons 1 and 2 show, respectively, 81% and 84% identical residues in the five L-chain sequences and 84% and 98% identities in the four sequenced H chains, as compared with 50% and 45% when all nine sequences are included in the comparison (tadpole H, L and M were excluded). The values for exons 3 and 4 are 71% and 76% respectively for L chains, 83% and 83% for H chains and 45% and 40% for equivalenced residues in both L and H chains. The very high degree of conservation in exon 2 of H chains is notable. It has been calculated that the 45% divergence of H and L chains overall would require some 300 million years to be accomplished (Munro et al. 1985).

Acknowledgements. We thank the Wellcome Trust and the Medical and Science and Engineering Research Councils for support. G. C. F. is a Wellcome Trust Reader.

References

Artymiuk PJ, Blake CCF, Sippel AE (1981) Genes pieced together – exons delineate homologous structures of diverged lysozymes. Nature 290:287–288

Bakker GR, Boyer RF (1986) Iron incorporation into apoferritins. The role of apoferritin as a ferroxidase. J Biol Chem 261:13182–13185

Beaumont C, Dugast I, Renaudie F, Souroujon M, Grandchamp B (1989) Transcriptional regulation of ferritin H and L subunit in adult erythroid and liver cells from the mouse. J Biol Chem 264:7498–7504

Blake CCF (1978) Do genes-in-pieces imply proteins-in-pieces? Nature 273:267

Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc Natl Acad Sci USA 75:4853–4857

Bryce CFA, Crichton RR (1973) The catalytic activity of horse spleen apoferritin. Preliminary kinetic studies and the effect of chemical modification. Biochem J 133:301–309

Burke J, Hwang P, Anderson L, Lebo R, Gorin F, Fletterick R (1987) Intron-exon structure of the human gene for the muscle isozyme of glycogen phosphorylase. Proteins 2:177–187

Chothia C (1976) The nature of the accessible and buried surfaces in proteins. J Mol Biol 105:1–14

Costanzo F, Colombo M, Staempfli S, Santoro C, Marone M, Frank R, Delius H, Cortese R (1986) Structure of gene and pseudogenes of human apoferritin H. Nucleic Acids Res 14:721–736

Craik CS, Sprang S, Fletterick R, Rutter WJ (1982) Intron-exon splice junctions map at protein surfaces. Nature 299:180–182

Daniels-McQueen S, Ray A, Walden WE, Ray BK, Brown PH, Thach RE (1988) Nucleotide sequence of cDNA encoding rabbit ferritin L chain. Nucleic Acids Res 16:7741

Dickey LF, Sreedharan S, Theil EC, Didsbury JR, Wang Y-H, Kaufman RE (1987) Differences in the regulation of messenger RNA for housekeeping and specialized-cell ferritin. J Biol Chem 262:7901–7907

Ford GC, Harrison PM, Rice DW, Smith JMA, Treffry A, White JL, Yariv J (1984) Ferritin: design and formation of an iron-storage molecule. Phil Trans R Soc Lond B304:551–565

Gilbert W (1978) Why genes in pieces? Nature 271:501

Gilbert W, Marchioni M, McKnight G (1986) On the antiquity of introns. Cell 46:151–154

Go M, Nosaka M (1987) Protein architecture and the origins of introns. Cold Spring Harbor Symp Quant Biol 52:915–924

Heusterspreute M, Crichton RR (1981) Amino acid sequence of horse spleen apoferritin. FEBS Lett 129:322–327

Kabsch W, Sander C (1983) Dictionary of protein secondary structure. Pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637

Lawson DM (1990) X-ray structure determination of recombinant ferritins. Ph.D thesis, University of Sheffield

Lawson DM, Treffry A, Artymiuk PJ, Harrison PM, Yewdall SJ, Luzzago A, Cesareni G, Levi S, Arosio P (1989) Identification of the ferroxidase centre in ferritin. FEBS Lett 254:207–210

Leibold EA, Munro HN (1987) Characterization and evolution of the expressed rat ferritin light subunit gene and its pseudogene family. J Biol Chem 262:7335–7341

Levi S, Luzzago A, Cesareni G, Cozzi A, Fransceschinelli F, Albertini A, Arosio P (1988) Mechanism of ferritin uptake: activity of the H chain and deletion mapping of the ferrooxidase site. J Biol Chem 263:18086–18092

Levi S, Salfeld J, Fransceschinelli F, Cozzi A, Dorner MH, Arosio P (1989) Expression and structural and functional properties of human ferritin L chains from *Escherichia coli*. Biochemistry 28:5179–5184

Macara IG, Hoy TG, Harrison PM (1972) The formation of ferritin from apoferritin. Kinetics and mechanism of iron uptake. Biochem J 126:151–162

Macara IG, Hoy TG, Harrison PM (1973) The formation of ferritin from apoferritin. Catalytic action of apoferritin. Biochem J 135:343–348

Miyazaki Y, Setoguchi M, Higuchi Y, Yoshida S, Akizuki S, Yamamoto S (1988) Nucleotide sequence of cDNA encoding the heavy subunit of mouse macrophage ferritin. Nucleic Acids Res 16:10373

Munro HN, Leibold EA, Vass JK, Aziz N, Rogers J, Murray M, White K (1985) Ferritin genes structure and expression. In: Spik G, Montreul J, Crichton RR, Mazurier J (eds) Proteins of iron storage and transport. Elsevier Science, Amsterdam, pp 331–339

Murray MT, White K, Munro HN (1987) Conservation of ferritin heavy subunit gene structure: implications for the regulation of ferritin gene expression. Proc Natl Acad Sci USA 84:7438–7442

Niederer W (1970) Ferritin: iron incorporation and iron release. Experientia 26:218–220

Padgett RA, Grabowski PJ, Konarska MM, Seiler S, Sharp PA (1986) Splicing of messenger RNA precursors. Annu Rev Biochem 55:1119–1150

Rice DW, Ford GC, White JL, Smith JMA, Harrison PM (1983) The spatial structure of horse spleen apoferritin. In: Thiel EC, Eichhorn GL, Marzilli LG (eds) Advances in inorganic biochemistry. vol 5. Iron binding proteins without cofactors or sulfur clusters. Elsevier Science, New York, pp 39–50

Santoro C, Maron M, Ferrone M, Costanzo F, Colombo M, Minganti C, Cortese R, Silengo L (1986) Cloning of the gene coding for human L apoferritin. Nucleic Acids Res 14:2863–2876

Stevens PW, Dodgson JB, Engel JD (1987) Structure and expression of the chicken ferritin H-subunit gene. Mol Cell Biol 7:1751–1758

Thomas CD, Shaw WV, Lawson DM, Treffry A, Artymiuk PJ, Harrison PM (1988) Expression and purification of the rat liver ferritin light chain after cloning in *Escherichia coli*. Biochem Soc Trans 16:838–839

Torti SV, Kwak EL, Miller SC, Miller LL, Ringold GM, Myambo KB, Young AP, Torti FM (1988) The molecular cloning and characterization of murine ferritin heavy chain, a tumour-necrosis-factor-inducible gene. J Biol Chem 263:12638–12644