Amino-acid sequence and predicted three-dimensional structure of pea seed (*Pisum sativum*) ferritin§

Stéphane LOBREAUX,* Stephen J. YEWDALL,† Jean-François BRIAT*‡ and Pauline M. HARRISON†‡

*Laboratoire de Biologie Moléculaire Végétale, Centre National de la Recherche Scientifique (Unité de Recherche Associée 1178) and Université Joseph Fourier, B.P. 53X, F-38041 Grenoble Cedex, France, and †The Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, P.O. Box 594, Firth Court, Western Bank, Sheffield S10 2UH, U.K.

The iron storage protein, ferritin, is widely distributed in the living kingdom. Here the complete cDNA and derived amino-acid sequence of pea seed ferritin are described, together with its predicted secondary structure, namely a four-helix-bundle fold similar to those of mammalian ferritins, with a fifth short helix at the C-terminus. An N-terminal extension of 71 residues contains a transit peptide (first 47 residues) responsible for plastid targeting as in other plant ferritins, and this is cleaved before assembly. The second part of the extension (24 residues) belongs to the mature subunit; it is cleaved during germination. The amino-acid sequence of pea seed ferritin is aligned with those of other ferritins (49 % amino-acid identity with H-chains and 40 % with L-chains of human liver ferritin in the aligned region). A three-dimensional model has been constructed by fitting the aligned sequence to the coordinates of human H-chains, with appropriate modifications. A folded conformation with an 11-residue helix is predicted for the N-terminal extension. As in mammalian ferritins, 24 subunits assemble into a hollow shell. In pea seed ferritin, its N-terminal extension is exposed on the outside surface of the shell. Within each pea subunit is a ferroxidase centre resembling those of human ferritin H-chains except for a replacement of Glu-62 by His. The channel at the 4-fold-symmetry axes defined by E-helices, is predicted to be hydrophilic in plant ferritins, whereas it is hydrophobic in mammalian ferritins.

INTRODUCTION

Iron is an essential element for virtually all forms of life because of its role in fundamental processes such as respiration, photosynthesis, nitrogen fixation and cell division. However, the low solubility of iron and its ability to catalyse the formation of reactive hydroxyl radicals necessitate the control of iron uptake and storage by cells (Crichton & Charloteaux-Wauters, 1987; Theil, 1987). Storage is achieved by ferritins, a class of proteins widely distributed among animals, plants and bacteria. These proteins are essential to cellular iron homeostasis because of their ability to sequester inside their protein coats up to 4500 iron atoms, in a safe form, either as the crystalline mineral ferrihydrite or as amorphous hydrous ferric oxyphosphate (Theil, 1987; Harrison *et al.*, 1989, 1991).

Structure, function and synthesis of animal ferritins have been studied extensively (Theil, 1987; Harrison *et al.*, 1989; Klausner & Harford, 1989). Two different polypeptide chains, H and L, are assembled in various ratios in the ferritin isoforms of different mammalian tissues (Jain *et al.*, 1985; Theil, 1987). The following general conclusions have been gained from X-ray crystallographic analysis of horse spleen ferritin (Ford *et al.*, 1984; Harrison *et al.*, 1989), rat liver ferritin, recombinant rat L-chain ferritin (Harrison *et al.*, 1991), recombinant human H-chain ferritin (Lawson *et al.*, 1991) and the recombinant H-chain ferritin from *Schistosoma mansoni* (Hirzmann *et al.*, 1991). Ferritin subunits in these different species show a remarkable conservation of their three-dimensional structure, backbone atoms being superposable to within ± 0.1 nm (1.0 Å). The shape of the subunit is cylindrical, with a length of 5 nm (50 Å) and a width of 2.5 nm (25 Å). Each subunit is composed of a four- α -helix bundle containing two antiparallel helix pairs (A,B and

C,D). Helices in each pair are linked by a short turn while the B- and C-helices are connected by a long non-helical stretch of 18 residues (the BC-loop) spanning the length of the bundle. A fifth α -helix, or E-helix, lies at one end of the bundle, at about 60° to its axis. The helical regions of the human H-subunit encompass the following residues: A, 14–43; B, 50–76; C, 96–123, D, 129–136 and 137–158 with a break at 136 due to a wide turn; and E, 163–172. Non-helical regions are N, 1–13; the AB-turn, 44–49; the BC-loop, 77–95; the CD-turn, 124–128; and the DE-turn, 159–162; with residues 177–182 at the C-terminus not being visible in electron density. Ferritin subunits assemble as compact 24-mers with 432 symmetry. The subunits are arranged in antiparallel pairs making 12 flat rhombs, the assembly of which approximates to a rhombic dodecahedron. The resulting packed shell contains many intersubunit interactions except that at the 4-fold and 3-fold-symmetry axes, narrow channels of about 0.15 nm (1.5 Å) wide are formed. The 4-fold channels are flanked by four nearly parallel E-helices so that their length extends to about 1.3 nm (13 Å).

Bacterioferritins (BFR) share a low sequence identity with animal ferritins (Andrews *et al.*, 1991, 1992). However, a combined prediction of *Escherichia coli* BFR secondary structure gave a helical distribution quite similar to that observed by X-ray crystallography for mammalian ferritins, namely a four-helix-bundle conformation resembling that of ferritin (Andrews *et al.*, 1991).

The complete primary structures of two plant ferritins have recently been determined by sequencing soybean (Lescure *et al.*, 1991) and French bean cDNAs (Spence *et al.*, 1991); about 50 % of the amino-acid sequence of the pea seed ferritin subunit has also been deciphered through microsequencing of its N-terminus and CNBr peptides (Ragland *et al.*, 1990). These plant ferritins

Abbreviations used: BFR, bacterioferritin; SSPE, 1.1 M-NaCl, 60 mM-sodium phosphate, 6 mM-EDTA, pH 7.7; UTR, untranslated regions.

‡ To whom all correspondence and reprint requests should be addressed.

§ The pea seed ferritin sequence reported in this paper has been deposited in the EMBL database under the accession no. X64417.

Fig. 1. Sequence analysis of pea seed ferritin cDNA

(a) Nucleotide sequence and derived amino acids of pea seed cDNA. The position of cleavage between the transit peptide and the mature subunit sequence is indicated by an arrow. Overlined amino acids correspond to the *N*-terminal sequence and to two sequences of internal CNBr peptides of the ferritin subunit purified from pea seeds (Ragland *et al.*, 1990). A presumptive polyadenylation signal is underlined in the 3'-UTR. (b) Comparison of the sequence of various transit peptides from plant ferritins. Abbreviations: PsaS, pea seed ferritin (this work); SoLc-, soybean ferritin from iron-induced cells (Lescure *et al.*, 1991); Pb-S, French bean seed ferritin (Spence *et al.*, 1991). (c) Comparison of primary structures of some plant and animal ferritins. Plant ferritin sequences correspond to the mature subunit coding region described above (without the transit peptide). The numbering used is the one which applies to the human

are 94% identical and share between 34% and 59% identity with animal ferritin amino-acid sequences. From the available data (Ragland *et al.*, 1990; Lescure *et al.*, 1991; Spence *et al.*, 1991; this work), plant ferritins seem to be built from only one type of subunit, although it cannot be ruled out that others will be identified in the future.

Despite an evidently common origin with animal ferritins, various specific features are found within plant ferritin sequences. First, an additional plant-specific sequence is found in the *N*-terminal part of the soybean and French bean ferritin subunit (Ragland *et al.*, 1990; Spence *et al.*, 1991). The first part of this extension encodes a transit peptide responsible for plastid targeting. The plant ferritin precursors are synthesized from nuclear genes. The mature ferritin subunits are obtained by cleavage of the transit peptides before assembly of apoferritin within plastids (van der Mark *et al.*, 1983; Proudhon *et al.*, 1989; Lescure *et al.*, 1991) into 24-mers which have been visualized by conventional (Seckbach, 1982) and immunogold-decoration (Lescure *et al.*, 1991) electron microscopy. The second-part of the *N*-terminal extension, retained during assembly of the mature plant ferritin subunits, has also never been observed in animal ferritins or BFRs. It is specifically cleaved *in vitro* during iron exchange (Laulhère *et al.*, 1989) or *in vivo* during germination (Lobreaux & Briat, 1991). In the C-terminal regions (about 20 residues) the amino-acid sequence of soybean (Lescure *et al.*, 1991) and French bean (Spence *et al.*, 1991) ferritins differ markedly from those of animal ferritins. This raises the question as to whether plant ferritins contain E-helices and hence the 4-fold channels described above. Finally, four residues of the pea seed sequence have to be classed as insertions in the BC-loop in order to align them with those of animal ferritins (Lescure *et al.*, 1991).

In this paper, we report the complete primary structure of the pea seed ferritin subunit determined by sequencing of its corresponding cDNA, and provide an alignment of the three plant sequences with those of human H- and horse L-chain ferritins. We also analyse its secondary structure and propose a model of the three-dimensional structure of the assembled pea seed ferritin. We have been particularly interested in the effects that the extension of the *N*-terminus of the subunit and the divergence of the E-helix region would have on the conformations of plant ferritin subunits and the structure of assembled molecules.

MATERIALS AND METHODS

Cloning and sequencing of pea seed (*Pisum sativum*) ferritin cDNA

Poly(A)⁺ RNA was extracted as previously described (Proudhon *et al.*, 1989) from young pea seeds (14 days after flowering) harvested from plants grown in hydroponic cultures containing 100 µM-Fe-EDTA in the culture medium (Lobreaux & Briat, 1991). cDNAs were synthesized from 4 µg of poly(A)⁺ RNA using the cDNA 'system plus' kit (Amersham International, Aylesbury, Bucks., U.K.) and cloned on λNM 1149 arms according to Martin *et al.* (1991). Recombinant phages

H-chain subunit (Lawson *et al.*, 1991). The first 24 residues are specific to plant ferritins and are therefore numbered from 'a' to 'x'. HoS-L = horse spleen L-subunit (Heusterspreute & Crichton, 1981); HuL-H = human liver L-subunit (Boyd *et al.*, 1985). Residues in the alignment which are either conserved or conservatively changed with respect to human H-chain ferritin are boxed. Hyphens indicate gaps in the sequences to allow the best alignment. The additional amino acids in the plant sequences, owing to gaps in the animal sequences, are numbered using lower-case letters.

Fig. 2. Prediction of the secondary structure of the pea seed ferritin

The prediction has been obtained using a combination of eight secondary structure predictions (Eliopoulos, 1986). The horizontal bars represent the joint prediction. The numbering is that of human H-chain ferritin as used by Lawson *et al.* (1991). Helices found in human H-chain ferritin are indicated by cylinders.

were plated and screened with a partial soybean ferritin cDNA (Ragland *et al.*, 1990) labelled using [32 P]dCTP (18.5 Bq/mm Amersham) and a random priming kit (Pharmacia). Hybridizations were performed in $6\times$ SSPE, 0.1% SDS, 0.02% polyvinylpyrrolidone, 0.02% Ficoll, 50 μ g of salmon sperm DNA/ml at 58 $^{\circ}$ C. Filters were washed twice in $6\times$ SSPE, 0.1% SDS at 60 $^{\circ}$ C. Inserts from recombinant phages were subcloned in pUC 18 vector and their sequences were determined by double-strand sequencing using the T7 sequencing kit (Pharmacia).

Computer analysis

Plant and mammalian ferritin sequences have been aligned by using the profile-analysis method (Andrews *et al.*, 1991) with minor modifications imposed by structural constraints. Prediction of the secondary structure of pea seed ferritin was made by a combination of eight different programs according to Eliopoulos (1986).

The molecular modelling program FRODO (Jones, 1978) implemented on an Evans and Sutherland series 'V' workstation supporting 'crystal eyes' stereo graphics was used for all model-building studies on pea seed ferritin. The known X-ray structure coordinates of recombinant human H-ferritin (Lawson *et al.*, 1991) were used as the skeletal framework on which to model the pea seed ferritin subunit. By the use of 432 symmetry operators all 24 subunits of the pea ferritin molecule were generated in order to investigate the effects of the pea ferritin sequence on intersubunit contacts, with special reference to channel formation, sequence extension and residue interactions.

Replacement of the human H-chain sequence by the pre-aligned pea seed sequence proceeded systematically. Structural

identities were preserved wherever possible. Those portions of the structure away from symmetry axes were then modelled, first making those side-chain replacements which were deemed of little consequence (e.g. with respect to size or charge). Idealized geometries, torsion angles (Ramachandran & Sasisekharan, 1968) and hydrogen-bonding distances of 0.285 nm (2.85 $\text{\AA}$) were generated throughout. The modelling of the N-terminal region (comprising 24 plant-specific residues and 12 residues similar to those of animal H ferritins) utilized pre-generated pieces of secondary structure (limited by the presence of four prolines) as predicted. Side-chains were added and their orientations chosen to give good geometry and reasonable interactions. The extra four residues in the BC-loop were accommodated with minor rearrangements in that region. The final model was subjected to repeated cycles of geometrically restrained regularization using the method described by Hermans & McQueen (1974). Figs. 3–6 inclusive were prepared using the MOLSCRIPT drawing program (Kraulis, 1991).

RESULTS AND DISCUSSION

Primary structure of the pea seed ferritin subunit as derived from its cDNA sequence

Ferritin protein accumulates preferentially in pea seed, as a store of iron for the needs of seedlings during the early stages of germination (Lobreaux & Briat, 1991). Immunoprecipitation of ferritin from translation products obtained *in vitro*, using the rabbit reticulocyte cell-free extract and poly(A)⁺ mRNA purified from young pea seed (14 days after flowering) revealed that translatable ferritin mRNA is abundant at this stage of pea seed development (results not shown). Therefore, a cDNA library was constructed from pea seed poly(A)⁺ mRNAs to isolate pea seed ferritin cDNA by heterologous screening with 32 P-labelled partial soybean ferritin cDNA (Ragland *et al.*, 1990). This confirmed the abundance of ferritin mRNA among pea seed poly(A)⁺ mRNAs: 20 positives were isolated from 10^4 recombinant phages plated. Among the clones selected according to their length, two were sequenced. The longest (pPeSF1) is shown in Fig. 1(a). It starts with ATGGC, the canonical eukaryotic translation initiation sequence, and therefore it lacks a 5'-untranslated region (UTR). An open reading frame of 253 amino acids was observed, followed by a 3'-UTR of 266 bp, within which a consensus polyadenylation site was clearly defined (underlined in Fig. 1a). Within the open reading frame, a stretch of 48 amino acids has a sequence identical to the N-terminus of pea seed ferritin (Ragland *et al.*, 1990) (overlined in Fig. 1a), with the exception of Val-90 [Fig. 1(a) H-chain numbering] which is an Ile residue in the protein sequence. This minor difference could be due either to an error in the protein sequence or to a weak polymorphism of plant ferritins. The amino-acid sequences of two internal CNBr peptides (Ragland *et al.*, 1990) are also overlined in Fig. 1(a); they are 100% identical to that derived from the corresponding cDNA sequence. The complete amino-acid sequence of the mature pea seed ferritin subunit is therefore identified without ambiguity. It is preceded by a sequence of 47 amino acids showing the characteristics of a transit peptide for plastid targeting (Heijne *et al.*, 1989). This is consistent with the finding that plant ferritins are synthesized as precursors (Proudhon *et al.*, 1989; van der Mark *et al.*, 1983) and transported to plastids (Seckbach, 1982; Lescure *et al.*, 1991). The precursor polypeptide of pea seed ferritin has a molecular mass of 28 kDa and the mature ferritin subunit a value of 23.5 kDa, as calculated from the amino-acid sequence shown in Fig. 1(a).

The transit peptide of pea seed ferritin is respectively 42% and 49% identical to those of French bean (48 residues) and soybean

Fig. 3. Structure of the pea seed ferritin subunit

A stereoscopic superposition of the structure of the human H-chain ferritin subunit (bold) (Lawson *et al.*, 1991) and of the pea seed ferritin model (light). The N- and C-termini, helices A to E and the proposed P-helix are indicated. The C_α-positions of the four residues classed as insertions in the BC-loop region of the pea seed ferritin model are shown as small spheres. (b) Detail of the structure of the plant-specific extension peptide found in the N-terminal part of the pea seed ferritin subunit. Amino acids specific to the plant sequence are numbered from 'A' to 'X'. Ser-1, Val-2 and Pro-3 corresponds to the numbering of human H-chain ferritin subunit (Lawson *et al.*, 1991).

(49 residues) ferritins (Fig. 1b). Comparison of the remaining amino-acid sequence of pea seed ferritin with those of soybean, French bean, human liver H- (Boyd *et al.*, 1985) and horse spleen L-ferritin (Heusterspreute & Crichton, 1981) is presented in Fig.

1(c). Pea seed ferritin shares 90% and 84% amino-acid sequence identity with soybean and French bean ferritin subunits respectively, but only 49% identity with human liver H-ferritin and 39% with horse spleen L-ferritin in the common regions of

Fig. 4. A stereoscopic view of seven subunits showing the stacking of aromatic residues within the pea seed ferritin shell as predicted by modelling

this alignment. Four amino acids are classed as insertions in the plant sequences [positions 87a, 88a and 96a and 96b, in Fig. 1(c)], and a gap of three amino acids was introduced at positions 160–162 to obtain the best alignment with animal ferritins. It is notable that the 20 C-terminal amino acids are highly conserved within the three plant ferritins, but that they have diverged almost entirely from their animal counterparts.

Secondary structure prediction of pea seed ferritin

Secondary structure predictions of the mature pea seed ferritin subunit are shown in Fig. 2. The joint prediction indicates a very high level of similarity with the secondary structure of mammalian ferritin, especially the high helical content consistent with the adoption of a 4-helix-bundle conformation by residues 14–158 (numbers as in human H-chain). A stretch of 21 residues [positions 77–99 in Fig. 1(c)] gives a prediction of either a turn/coil or a β -strand and thus could link the B- and C-helices via a BC-loop like that characteristic of the animal-ferritin-subunit fold (Harrison *et al.*, 1989).

In the *N*-terminal extension the prediction indicates an additional helix (three turns) which we name the 'P-helix'. This predicted helix, of 11 amino acids in length (residues 'M' to 'W'), is within the plant-specific sequence of 24 amino acids (Ragland *et al.*, 1990) which is cleaved *in vivo* during the early stages of germination or *in vitro* during iron exchange.

A notable prediction is the conservation of the short C-terminal E-helix, despite the divergent primary structures of plant and animal ferritins in this region.

A model of the three-dimensional structure of pea seed ferritin subunit

A three-dimensional conformation for the pea seed ferritin subunit based on the structure of recombinant human H-chain ferritin was built with good geometry although it required the replacement of 97 of its 182 amino acids. The predicted secondary structure was used in modelling the *N*-terminal domain and along much of the length of the four long helices A, B, C and D residues have been conserved or conservatively replaced with substitutions mainly on outside or cavity surfaces. The complete model is shown, in Figs. 3(a) and 3(b), superimposed on the known subunit structure of recombinant human H ferritin. The

model suggests a remarkable conservation of many of the structurally important motifs, both within the subunit fold [highlighted in Fig. 3(a)] and across subunit interfaces in the assembled molecule (see below). However, some specific aspects of the pea-seed-model structure warrant discussion.

The *N*-terminal region and its interactions

The P-helix, flanked by proline residues, which accounts for nearly half of the 24 residue *N*-terminal extension, is folded back on to the surface of the subunit (Fig. 3b). The sequence LAVPS (residues 'U' to 1) within which free-radical cleavages occur during iron exchange (Laulhère *et al.*, 1989) is found at the C-terminal end of this helix, and this region is predicted to lie on the outer surface of the assembled ferritin molecule (Fig. 6a).

Residues 12–16 at the start of the A-helix and 121–124 at the C-terminal end of the C-helix are close together in the three-dimensional structure. Although residues within these two sequences make a number of important interactions, several substitutions are found in the plant sequences. A hydrogen-bonding pair Asp-15 and Lys-124 of human H-chain is replaced by Glu-15 and Asn-124 in pea seed ferritin such that two hydrogen bonds replace the single interaction O^{δ1} Asp-15 ... N^ε Lys-124. Substitution of His-13 by Ala provides a space which allows Lys ('R') from the P-helix to interact with O^{δ1} of Asn-124. A hydrogen bond from O^{δ1} Asp-14 (replacing Gln) to the main chain carbonyl O of residue 'X' (Pro) provides a further link to the *N*-terminal extension.

A largely conserved sequence in animal and plant ferritins comprises residues 9–11. In animal ferritins these residues participate in an extensive set of both intra- and inter-subunit bonds that are important for the maintenance of both tertiary and quaternary structure in the assembled molecules. Arg-9 and Gln-10 are both conserved in the plant ferritins. Arg-9 forms hydrogen bonds with Asn ('S'), Glu-17 (near the start of the A-helix) and the main chain O of Lys-108 of a neighbouring subunit. N^{ε2} Gln-10 interacts with N^{δ2} Asn-111 (again highly conserved) and with main chain O of Lys-108, both from another subunit. N^{δ2} Asn-11 bonds to the main chain O of residue 125 in the CD-turn and the main-chain NH group of Arg-79 bonds to O^{ε1} Glu-17. These interactions, which will be described in more detail elsewhere, are conserved in the pea seed model.

The C-terminus of the B-helix and the CD-turn

Conserved amino acids Gln-73, Arg-76, Gly-77, Gly-78, Arg-79 (in a short sequence at the C-terminus of the B-helix and the start of the BC-loop) and Asp-126 at the sharp CD-turn are involved in important interactions maintaining the subunit fold which are reproduced in the pea seed model. Gln-73 is hydrogen bonded to O^{δ1} Asn-21 and also to the main chain O of residue 17. Arg-76 interacts with Asp-126 and is hydrogen bonded to the main chain O of residue 128. The two glycines, residues 77 and 78, are involved in a tight turn. N^{η1} Arg-79 hydrogen bonds to the main chain O of Gly-78 and also to that of residue 4.

The A/A', B/B', BC-loop/B'C'-loop inter-subunit interface

The intersubunit interface which spans virtually the entire subunit length has been described previously for horse spleen ferritin (Rice *et al.*, 1983; Ford *et al.*, 1984). Examination of side-chain packing in this interface (at the centre of which lies a 2-fold rotational axis) suggests that it is structurally conserved in pea seed ferritin [as it is with some sequence substitutions in human H-ferritin (Lawson, 1990)]. The packing of residues Asn-21, Ile-24 and Met-70 of human H-chains is maintained (residues 24 and 70 are Val and Leu respectively in horse L-chains). In pea seed ferritin, however, Tyr-28 and His-35 form an intersubunit bond replacing the apolar pair Leu-28 and Leu-35 of the mammalian H- and L-chains. Isoleucines 85 interact across the 2-fold axis as in H-chain homopolymers. Near the centre of the 2-fold interface on the cavity side, Arg-63 interacts with Glu-67 of the same subunit and O^γ Ser-59 of its neighbour in the pea seed model as it does in human H-ferritin (but residue 59 is Ala in horse L-chains). Lys-56 of pea seed ferritin can also form a hydrogen bond to Glu-67 and this interaction is similar to that of horse L-chains (which have Arg-56 and Glu-67), whereas in human H-chain residue 56 is a leucine.

The BC-loop and N-terminal end of the C-helix

The start of the BC-loop, residues 78 and 79, are highly conserved as already described, and residues within the region 82–87 contributing to the A/A', BC-loop/B'C'-loop interface are for the most part conserved or conservatively replaced, except for Asp-84/Pro substitution, which causes the loop to pucker thus creating space for the sequence insertions of the pea seed. The remaining part of the BC-loop and start of the C-helix is a region containing four residue insertions according to the alignment of Fig. 1(c). Two of these, at positions 96a and 96b, have been placed so as to extend the C-helix at its N-terminus, which now starts at Asp-96a instead of Leu-97.

The DE-turn and the E-helix

The amino-acid sequence alignment of Fig. 1(c) suggests a deletion of three residues at the DE-turn, residues 160–162. A tight turn was easily introduced into the pea seed ferritin molecule by allowing flexibility at glycine residues 159, 164 and 166. Interestingly this flexible deletion region lies in juxtaposition to the two-residue insertion at the start of the C-helix, which can thus readily be accommodated. Trp-168 fits easily under the new DE-turn to pack against a symmetry-related subunit and enables the generation of an E-helix in the same place and orientation as those found in animal ferritins.

Residues within the four-helix bundle

Stacking of aromatic residues. The pea-seed-subunit model has an internal structure resembling that of other ferritins: it contains a central hydrophilic region flanked by two hydrophobic cores. Plant ferritins are particularly remarkable, however, in the

Fig. 5. Detail of the structure of the ferroxidase centre of pea seed ferritin

Note that residue 62 is His in pea seed ferritin instead of Glu as seen in the other ferritins shown in Fig. 1(c).

predominantly aromatic character of their hydrophobic regions. Indeed aromatic side-chains Phe-12, Tyr-72, Phe-132, Tyr-137, His-65, His-62, Tyr-34, Phe-55, Phe-41, and His-176 are stacked along the entire bundle length as shown in Fig. 4. Roughly perpendicular to this aromatic chain are two adjacent, shorter stacks at the C-terminal end of the bundle. These comprise Phe residues 38, 55 and 54, and Tyr-151 in the first, and Phe-90, Tyr-40, Phe-41, Phe-51, Phe-170 and Trp-168 (from a neighbouring subunit) in the second. Phe-55 and Phe-41 are common to the axial chain. Pea seed ferritin also contains His-165 and His-169 protruding from the E-helices into the 4-fold channels, these, together with the other surface residues His-83, Tyr-28, His-35 and Phe-38 (juxtaposition Tyr-28), suggest, on examination of the quaternary structure, a potential network for electron transfer through the protein shell.

The ferroxidase centre. The central hydrophilic region of the subunit fold resembles those of H-chain ferritins. In H-chains this region contains seven conserved residues, namely Glu-27, Tyr-34, Glu-61, Glu-62, His-65, Glu-107 and Gln-141, and it has been shown to be associated with metal-binding and ferroxidase activity (Lawson *et al.*, 1989, 1991; Bauminger *et al.*, 1991). The equivalent hydrophilic region of L-chains, having only three of the required residues, lacks this activity. In soybean and French bean sequences all seven ferroxidase-centre residues are present (Andrews *et al.*, 1991), but in the pea seed sequence (both cDNA and CNBr peptide results) residue 62 is a histidine (Fig. 1c). The hydrogen-bond network involving Tyr-34 ... Glu-107 ... Gln-141 is preserved and extended by an additional interaction Thr-110 ... N^{ε2} Gln-141 due to the substitution of a valine residue (in human H-chains) by a threonine residue. The modelled ferroxidase centre of pea seed ferritin is shown in Fig. 5. Although altered in one residue this centre does seem to be associated with fast Fe(II) oxidation (Wade *et al.*, 1992).

Quaternary structure: 3-fold and 4-fold channels and cavity surface

The predicted structure of assembled pea seed ferritin molecules is shown in Fig. 6(a). Twenty-four subunits assemble in 432 symmetry to give a compact hollow shell similar to those of other ferritins (Ford *et al.*, 1984; Lawson *et al.*, 1991) with

Fig. 6. Computer graphics of the assembled molecule

(a) A stereoscopic view of a half molecule of the pea seed apoferritin model viewed from outside down the 4-fold symmetry axis. The location of the plant-specific extension on the outer surface of the molecule is shown in bold. (b) A detailed stereoscopic view of the structure showing that hydrophilic amino acids from four different subunits are involved in building the 4-fold channels. His-165 (hatched), His-169 and Arg-173 (bold) can be seen lining this channel. The tight DE-turn and the packing of Trp-168 between subunits are also demonstrated.

numerous intersubunit interactions. The plant-specific *N*-terminal extension is modelled as a compact, localized domain on the outside surface of the 24-mer: it does not interfere with subunit assembly or the formation of intersubunit channels at 3-fold and 4-fold-symmetry axes. It is also predicted not to interfere with the crystallization of pea seed ferritin in the face-centred cubic close-packed array given by horse spleen ferritin (Harrison, 1959) and other ferritins, although this would require the engineering of metal bridge ligands as in human H-chain ferritin (Lawson *et al.*, 1991): in this case Pro-84 to Asp and Lys-86 to Gln.

Narrow, hydrophilic channels are predicted to occur at 3-fold symmetry axes as in mammalian ferritins (Ford *et al.*, 1984; Harrison *et al.*, 1989) except that, at the cavity end, Asp-131 is substituted by His. When this substitution was engineered in human H-chain ferritin the molecule assembled and took up iron

at an only slightly impaired rate (Treffry *et al.*, 1989). The other plant sequences contain Asp-131.

The narrow 4-fold channels of L-ferritins are hydrophobic: three leucines residues (165, 169 and 173, Fig. 1c) from each of the four contributing subunits face the channels along one side of each E-helix (Ford *et al.*, 1984). In human H-chains the first two leucines are retained but the innermost residues are histidines (Fig. 1c). However, in the predicted pea seed structure the 4-fold channels become hydrophilic, Fig. 6(b), with His-165 and His-169 in place of the first two leucines and Arg-173 at the cavity end (note this residue is Ser in the French bean sequence, Fig. 1c). The four Arg-173 side-chains may be orientated so as to separate their positive charges. Apart from residue 173 the E-helix residues are entirely conserved in the three plant sequences. The functional consequences, if any, of the presence of two hydrophilic channels in plant ferritins are unknown.

The inside surface of the pea seed ferritin shell is, as expected for the lining of the iron storage cavity, largely hydrophilic, with most residues conserved or conservatively replaced compared with other ferritins. Residues 61, 64 and 67 are glutamates, as in all mammalian ferritins, but there is one respect in which the cavity surface of all three plant ferritins resemble those of L-ferritins more closely than H-ferritins. This is the presence of the additional glutamates (instead of histidines) at positions 57 and 60. The higher concentration of carboxylic acid groups on the cavity surface of the L-chains may make them more effective in iron-core nucleation (Andrews *et al.*, 1992; Levi *et al.*, 1992; Wade *et al.*, 1992).

Conclusions

Despite major differences in their cytological location and in the control of their synthesis (Ragland *et al.*, 1990; Lescure *et al.*, 1991), plant and animal ferritins are predicted to share a highly conserved three-dimensional structure. Two plant-specific additional features are proposed: an additional *N*-terminal α -helix, or 'P-helix' extension, and hydrophilic channels at 4-fold-symmetry axes instead of the hydrophobic channels of animals. The model structure is predicted to be a functioning ferritin with ferroxidase activity and ferrihydrite nucleation capacity.

With the exception of the *N*-terminal extension, which is absent from the ferritins of known three-dimensional structure [namely horse spleen L-chain apoferritin (Banyard *et al.*, 1978; Ford *et al.*, 1984); human recombinant H-chain apoferritin (Lawson *et al.*, 1991), rat liver L-chain and recombinant L-chain ferritin (Lawson, 1990; Lawson *et al.*, 1991) and type 1 ferritin (S.ma-1) from the human parasite *Schistosoma mansoni* (Hirzmann *et al.*, 1991; S. J. Yewdall, J. Hirzmann, P. J. Artymiuk & P. M. Harrison, unpublished work)] and which must therefore be considered at this stage to be hypothetical, the essential correctness of the proposed pea seed structure can be predicted with a high degree of confidence, based on the following observations: (1) although showing percentage identities of amino-acid sequence similar to those of pea seed and human H-chain (e.g. horse-L/human-H 53%; horse-L/S.ma-1 45%; rat-L/human-H 55%), main chain atoms of human-H, horse-L, rat-L and S.ma-1 apoferritins superimpose to within about 0.1 nm (1 Å) (Yewdall *et al.*, 1990; Andrews *et al.*, 1992), and (2) for ferritins the correlation between secondary structure prediction and known three-dimensional structure is rather good [e.g. rat L-chain ferritin (Andrews *et al.*, 1987; Lawson, 1990) and horse spleen L-chain ferritin (Harrison *et al.*, 1987)]. The relatively fast rate of iron incorporation into pea seed apoferritin compared with L-chain-rich horse spleen apoferritin (Wade *et al.*, 1992) also suggests the presence of an active ferroxidase centre as predicted by the model. The predicted structure should be tested by X-ray analysis and other physical methods when the recombinant protein becomes available.

Note added in proof (received 22 September 1992)

Very recently we have isolated and sequenced a new ferritin cDNA from the same pea seed cDNA library described here. The following changes from the sequence reported here have been observed, namely Ala-30 to Val, Leu-37 to Met, His-65 to Glu and Tyr-102 to His. Most important is the His-65 to Glu change, which means that this pea seed ferritin chain contains a 'consensus' ferroxidase centre like those of other plant ferritins and mammalian H-chains.

This work has been supported by grants from CNRS (URA 1178) and INRA-GEVES (Programme qualité des semences) to J.F.B. S.L. was the recipient of a grant from the Réseau Européen de Biologie Moléculaire Végétale. We also gratefully acknowledge support from the

Science and Engineering Research Council (S.J.Y. and P.M.H.) and from the Wellcome Trust (P.M.H.).

REFERENCES

- Andrews, S. C., Treffry, A. & Harrison, P. M. (1987) *Biochem. J.* **245**, 447–453.
- Andrews, S. C., Smith, J. M. A., Yewdall, S. J., Guest, J. R. & Harrison, P. M. (1991) *FEBS Lett.* **293**, 164–168.
- Andrews, S. C., Arosio, P., Bottke, W. W., Briat, J.-F., von Darl, M., Harrison, P. M., Lahlère, J.-P., Levi, S., Lobreaux, S. & Yewdall, S. J. (1992) *J. Inorg. Biochem.* **47**, 161–174.
- Banyard, S. H., Stammers, D. K. & Harrison, P. M. (1978) *Nature (London)* **271**, 282–284.
- Bauminger, E. R., Harrison, P. M., Hechel, D., Nowik, I. & Treffry, A. (1991) *Biochim. Biophys. Acta* **1118**, 48–58.
- Boyd, D., Vecoli, C., Belcher, D. M., Jain, S. K. & Drysdale, J. W. (1985) *J. Biol. Chem.* **260**, 11755–11761.
- Crichton, R. & Charlotiaux-Wauters, M. (1987) *Eur. J. Biochem.* **164**, 485–506.
- Eliopoulos, E. E. (1986) Ph.D. Thesis, University of Leeds, Leeds, U.K.
- Ford, G. C., Harrison, P. M., Rice, D. W., Smith, J. M. A., Treffry, A., White, J. L. & Jariv, J. (1984) *Phil. Trans. R. Soc. Lond.* **B304**, 551–565.
- Harrison, P. M. (1959) *J. Mol. Biol.* **1**, 69–80.
- Harrison, P. M., Andrews, S. C., Ford, G. C., Smith, J. M. A., Treffry, A. & White, J. L. (1987) in *Iron Transport in Microbes, Plants and Animals: Ferritin and Bacterioferritin—Iron Sequestering Molecules from Man to Microbe* (Winkelman, G., van der Helm, D. & Neilands, J. B., eds.), pp. 445–475, VCH Verlagsgesellschaft, Weinheim.
- Harrison, P. M., Artymiuk, P. J., Ford, G. C., Lawson, D. M., Smith, J. M. A., Treffry, A. & White, J. L. (1989) in *Bio-mineralization—Chemical and Biomedical Perspectives: Ferritin—Function and Structural Design of an Iron-Storage Protein* (Mann, S., Webb, J. & Williams, R. J. P., eds.), pp. 257–294, VCH Verlagsgesellschaft, Weinheim.
- Harrison, P. M., Andrews, S. C., Artymiuk, P. J., Ford, G. C., Guest, J. R., Hirzmann, J., Lawson, D. M., Livingstone, J. C., Smith, J. M. A., Treffry, A. & Yewdall, S. J. (1991) *Adv. Inorg. Chem.* **36**, 449–486.
- Heijne, G. V., Steppuhn, J. & Herrman, R. G. (1989) *Eur. J. Biochem.* **180**, 535–545.
- Hermans, J. & McQueen, J. E. (1974) *Acta Crystallogr.* **A30**, 730–739.
- Heusterspreute, M. & Crichton, R. R. (1981) *FEBS Lett.* **129**, 322–327.
- Hirzmann, J., Dietzel, J., Treffry, A., Yewdall, S. J., Symmons, P., Artymiuk, P. J., Harrison, P. M. & Kunz, W. (1991) *Int. Conf. on Iron and Iron Proteins*, 10th, 1991, Oxford, abs. 16.
- Jain, S. K., Barrett, K. J., Boyd, D., Favreau, M. F., Campton, J. & Drysdale, J. W. (1985) *J. Biol. Chem.* **260**, 11762–11768.
- Jones, T. A. (1978) *J. Appl. Crystallogr.* **11**, 268–272.
- Klausner, R. D. & Harford J. B. (1989) *Science* **246**, 870–872.
- Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.
- Lahlère, J.-P., Labouré, A.-M. & Briat, J.-F. (1989) *J. Biol. Chem.* **264**, 3629–3635.
- Lawson, D. M. (1990) Ph.D. Thesis, University of Sheffield, Sheffield, U.K.
- Lawson, D. M., Treffry, A., Artymiuk, P. J., Harrison, P. M., Yewdall, S. J., Luzzago, A., Cesareni, G., Levi, S. & Arosio, P. (1989) *FEBS Lett.* **254**, 207–210.
- Lawson, D. M., Artymiuk, P. J., Yewdall, S. J., Smith, J. M. A., Livingstone, J. C., Treffry, A., Luzzago, A., Levi, S., Arosio, P., Cesareni, G., Thomas, C. D., Shaw, W. V. & Harrison, P. (1991) *Nature (London)* **349**, 541–544.
- Lescure, A. M., Proudhon, D., Pesey, H., Ragland, M., Theil, E. C. & Briat, J.-F. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8222–8226.
- Levi, S., Yewdall, S. J., Harrison, P. M., Santambrogio, P., Cozzi, A., Roviata, E., Albertini, A. & Arosio, P. (1992) *Biochem. J.* **288**, 591–596.
- Lobreaux, S. & Briat, J.-F. (1991) *Biochem. J.* **274**, 601–606.
- Martin, W., Lagrange, T., Li, Y. F., Bizance-Seyer, C. & Mache, R. (1991) *Curr. Genet.* **18**, 553–556.
- Proudhon, D., Briat, J.-F. & Lescure, A. M. (1989) *Plant Physiol.* **90**, 586–590.
- Ragland, M., Briat, J.-F., Gagnon, J., Lahlère, J.-P., Massenet, O. & Theil, E. C. (1990) *J. Biol. Chem.* **265**, 18339–18344.
- Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–437.

- Rice, D. W., Ford, G. C., White, J. L., Smith, J. M. A. & Harrison, P. M. (1983) in *Adv. Inorg. Biochem.* **5**, 39–50
- Seckbach, S. (1982) *J. Plant Nutr.* **5**, 369–394
- Spence, M. J., Henzl, M. T. & Lammers, P. J. (1991) *Plant Mol. Biol.* **17**, 499–504
- Theil, E. C. (1987) *Annu. Rev. Biochem.* **56**, 289–315
- Treffry, A., Harrison, P. M., Luzzago, A. & Cesareni, G. (1989) *FEBS Lett.* **247**, 268–272
- van der Mark, F., van der Briel, W. & Huisman, H. G. (1983) *Biochem. J.* **214**, 943–950
- Wade, V. J., Treffry, A., Laulhère, J.-P., Bauminger, E. R., Cleton, M. I., Mann, S., Briat, J.-F. & Harrison, P. M. (1992) *Biochim. Biophys. Acta*, in the press
- Yewdall, S. J., Lawson, D. M., Artymiuk, P. J., Treffry, A., Harrison, P. M., Luzzago, A., Cesareni, G., Levi, S. & Arosio, P. (1990) *Biochem. Soc. Trans.* **18**, 1028–1029

Received 22 May 1992/7 July 1992; accepted 13 July 1992