Volume 293, number 1,2, 164–168 FEBS 10417 November 1991

Bacterioferritins and ferritins are distantly related in evolution

Conservation of ferroxidase-centre residues

S.C. Andrews, J.M.A. Smith, S.J. Yewdall, J.R. Guest and P.M. Harrison

Krebs Institute for Biomolecular Research and Department of Molecular Biology and Biotechnology, P.O. Box 594, Firth Court, Western Bank, The University of Sheffield, Sheffield S10 2UH, UK

Received 16 September 1991; revised version received 2 October 1991

Iron-storage proteins can be divided into two classes; the bacterioferritins and ferritins. In spite of many apparent structural and functional analogies, no significant amino acid sequence similarity has been detected previously. This report now reveals a distant evolutionary relationship between bacterioferritins and ferritins derived by 'Profile Analysis'. Optimum alignment of bacterioferritin and ferritin sequences suggests that key residues of the ferroxidase centres of ferritins are conserved in bacterioferritins.

Bacterioferritin; Ferritin; Iron metabolism; Profile analysis; Protein evolution; Ferroxidase centre

1. INTRODUCTION

The 'ferritins' are a family of iron-storage proteins (ISP) which are widespread in nature [1]. Ferritins contain 24 subunits forming a spherical shell enclosing a central cavity where up to about 4500 Fe(III) atoms may be stored. The ferritins of mammals have been extensively characterised and appear to fulfil an essential 'house-keeping' function [1]. Mammalian ferritins consist of 2 types of subunit (H- and L-chains) which are 55% identical and mutually interchangeable in the fully assembled tetracosameric (24-mer) molecule. Initiation of iron-core growth occurs by protein-catalysed oxidation of Fe(II) to Fe(III). Recently a 'ferroxidase-centre' was identified in mammalian ferritin H-chains by structural and functional studies [2,3]. Seven amino acid residues which serve directly or indirectly in metal binding at the ferroxidase centre are absolutely conserved in mammalian H-chains, but only 3 or 4 of these are conserved in L-chains. Ferritin subunits from other species can be defined as 'H-like' or 'L-like' on the basis of conservation of the 7 ferroxidase-site residues.

Bacteria possess structurally-related iron-storage proteins, 'bacterioferritins' (BFR), which differ from ferritin in being haemoproteins with about 12 haem-*b* moieties per 24-mer [4]. No significant amino acid sequence similarity was detected between ferritin and the BFR of *Escherichia coli* (the only fully sequenced BFR) raising the possibility that they have evolved by convergence from distinct ancestors [5]. Recently, a gene (*gen-165*) expressing a ferritin-like protein (25% sequence identity to human H-chain ferritin) was cloned from *E. coli* [6]. This suggests that *E. coli*, and other bacteria, possess both classes of ISP.

This report describes the use of 'Profile Analysis' [7] to detect a distant evolutionary relationship between BFR and ferritins. Furthermore, an analysis of an optimum ferritin-BFR amino acid sequence alignment has revealed that the ferroxidase-centre residues of H-like ferritin chains are highly conserved in BFR and are absolutely conserved in FTN (the *gen-165* product), suggesting that BFR and FTN both possess a ferroxidase centre. Other structural and functional implications raised by the Profile Analysis are discussed.

Correspondence address: S.C. Andrews, Department of Molecular Biology and Biotechnology, University of Sheffield, Western Bank, Sheffield S10 2UH, UK. Fax: (44) (742) 728697.

Abbreviations: BFR, bacterioferritin; FTN, *Escherichia coli* ferritin (*gen-165* product); ISP, iron-storage protein; EcBFR, *Escherichia coli* BFR.

2. EXPERIMENTAL

2.1. *Computer Analysis*

Fourteen ferritin amino acid sequences were selected: Art (although not strictly a ferritin, artemin (Art) has been included in the alignment because its subunits are 29% identical in amino acid sequence to chicken H-chain ferritin [13]), ChiH, HumH and HumL, Sma1 and Sma2, TadH, TadL and TadM (see Fig. 1 for abbreviations and references), *Lymnaea stagnalis* 1 and *L. stagnalis* 2 (W. Bottke and M. von Darl, personal communication), Mouse H [8], Rat H [9] and *Xenopus laevis* [10]. These were optimally aligned using the computer programs SCORE, PREALIGN and ALIGN in the PAPA package of programs [11]. The optimum alignment was converted into Profile format using the program LineUp in the Profile Analysis suite of programs [7]. A 'profile' was calculated with PROFILEMAKE utilising a comparison table derived from the MDM78 mutation data

 Published by Elsevier Science Publishers B.V.

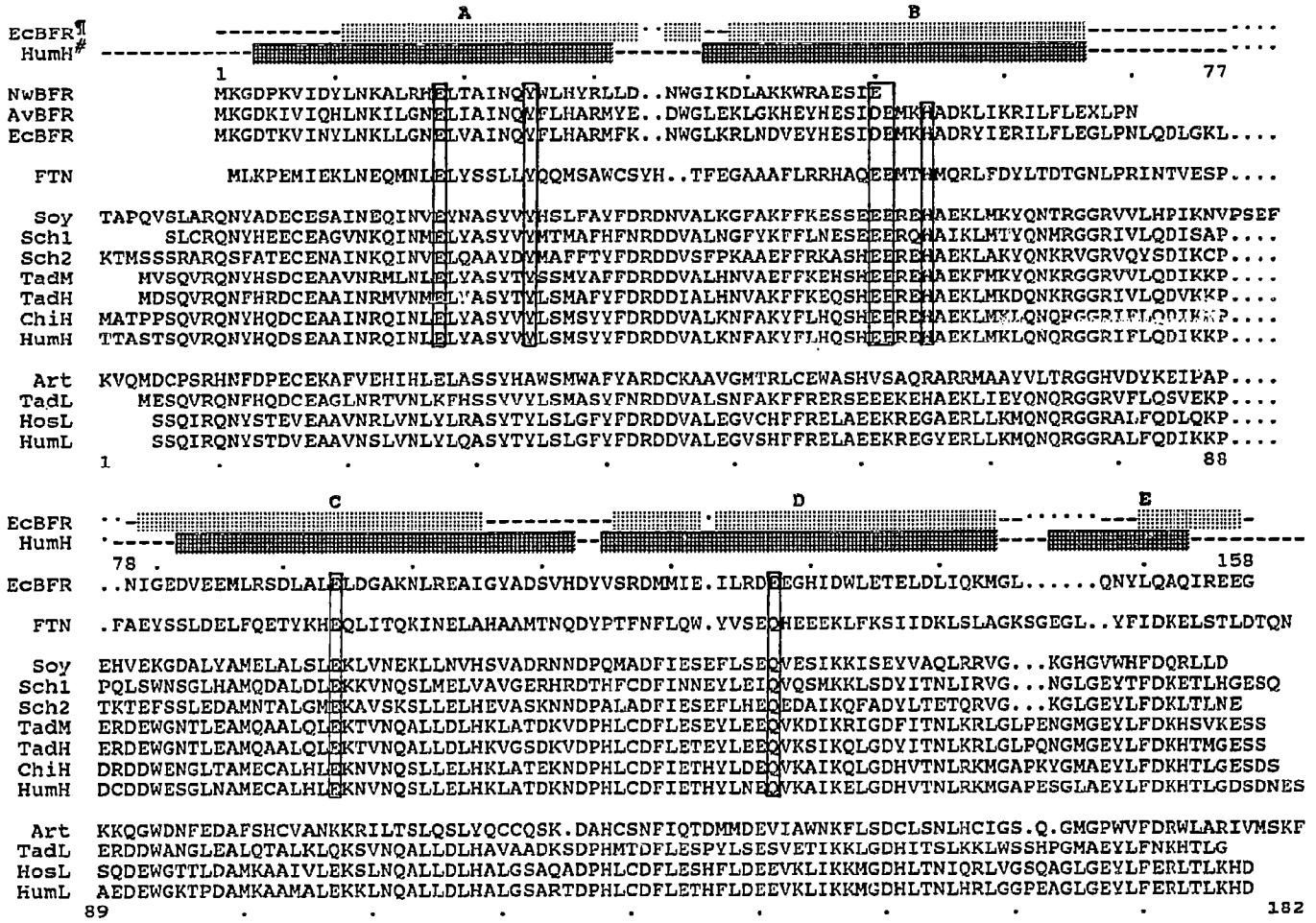

Fig. 1. Multiple sequence alignment of BFR and ferritins. The established and putative ferroxidase-centre residues are boxed. The α-helical regions of H-chain human ferritin (#) and the computer predicted α-helical regions of EcBFR (¶) are indicated by bars (dark and light shading, respectively) above the sequences. Non-helical regions are indicated with hyphens. Numbering above and below the multiple alignment corresponds to the EcBFR and HumH coordinates, respectively. Dots indicate positions of padding characters added to achieve optimum alignment. Abbreviations are as follows: Art, residues 16–194 of artemin from *Artemia* [13]; ChiH, residues 1–180 of chicken H-chain ferritin [17]; FTN, residues 1–165 of the *gen-165* product of *E. coli* [6]; HosL, residues 1–174 of horse L-chain ferritin [18]; HumH, residues 1–182 of human H-chain ferritin [19]; HumL, residues 1–174 of human L-chain ferritin [19]; Sch1, residues 1–172 of *S. mansoni* type 1 ferritin [20]; Sch2, residues 2–175 of *S. mansoni* type 2 ferritin [20]; Soy, residues 4–180 of soybean ferritin [21]; TadH, residues 1–176 of bullfrog H-chain ferritin [22]; TadL, residues 1–173 of bullfrog L-chain ferritin [22]; TadM, residues 1–176 of bullfrog M-chain ferritin [22]; AvBFR, residues 1–70 of *Azotobacter vinelandii* BFR [23]; EcBFR, residues 1–158 of *E. coli* BFR [5]; NwBFR, residues 1–50 of *Nitrobacter winogradskyi* BFR [24].

matrix [12]. This generated a position-specific score matrix consisting of a mathematical description of the alignment of the 14 ferritin sequences. The PROFILEGAP program was used to provide an optimum alignment of 2 other ferritin amino acid sequences (Soy and FTN; Fig. 1) which could not be aligned properly using ALIGN. The latter sequences were added to the 14-sequence alignment using LineUp, according to the results of PROFILEGAP. A fresh profile of 16 ferritin sequences was generated from the new alignment with PROFILEMAKE. All programs were implemented on the SEQNET computing facility, SERC Daresbury Laboratory, Daresbury, Warrington.

3. RESULTS AND DISCUSSION

3.1. *Detection of an evolutionary relationship between BFR and ferritin*

A profile of 16 ferritin sequences was compared with the 25 321 and 18 741 protein sequences in the NBRF [14] and SWISSPROT [15] databases, respectively, using PROFILESEARCH. The ten highest scoring sequences (excluding ferritins) selected by the searches were compared with the ferritin-profile using PROFILESEGMENTS which generated optimum alignments. The highest quality score was obtained with *E. coli* BFR (EcBFR). The likelihood of this occurring by chance is extremely remote, and it is therefore highly likely that BFRs and ferritins are related in evolution, albeit distantly.

3.2. *Analysis of the BFR-ferritin alignment*

The Profile-Analysis-derived alignment of 3 BFR sequences (one complete and 2 partial) with 12 ferritin sequences is displayed in Fig. 1 and the pair-wise degrees of amino acid sequence identity and similarity for

	NwBFR	AvBFR	EcBFR	FTN	Soy	Sch1	Sch2	TadM	TadH	ChiH	HumH	Art	TadL	HosL	HumL
NwBFR		54.0	60.0	20.0	22.0	20.0	22.0	22.0	20.0	22.0	22.0	12.0	16.0	16.0	16.0
AvBFR	70.0		74.3	18.6	18.6	24.3	21.4	20.0	22.9	22.9	22.9	10.0	20.0	15.7	14.3
EcBFR	72.0	82.9		12.7	16.5	17.7	19.6	19.0	19.0	21.5	20.9	9.5	17.7	17.1	15.8
FTN	30.0	34.3	31.7		21.8	24.2	23.6	24.2	22.4	21.2	22.4	17.0	24.2	18.8	18.8
Soy	42.0	34.3	32.9	40.0		49.4	51.2	51.7	50.0	51.4	50.9	24.3	45.1	42.5	42.0
Sch1	40.0	34.3	34.2	37.6	65.1		45.4	54.1	58.1	55.8	56.4	26.7	48.8	44.8	44.8
Sch2	42.0	35.7	37.3	39.4	61.4	57.0		48.9	50.0	50.6	51.2	25.3	39.9	44.3	44.3
TadM	44.0	37.1	38.6	40.6	68.2	68.0	64.4		83.5	68.2	66.5	27.8	61.9	57.5	59.2
TadH	44.0	44.3	38.6	38.8	66.5	70.9	64.4	92.6		71.0	68.2	27.3	65.3	58.6	59.8
ChiH	44.0	41.4	39.2	38.8	66.1	70.4	62.6	80.7	84.1		90.6	27.9	63.0	52.9	55.2
HumH	44.0	41.4	38.6	39.4	65.5	70.9	64.4	79.6	83.0	93.9		25.1	61.3	52.9	55.8
Art	30.0	25.7	29.8	33.3	41.8	43.6	43.1	46.0	47.2	44.1	43.6		24.3	22.4	24.1
TadL	38.0	34.3	32.9	42.4	68.2	65.1	59.0	75.7	78.0	76.9	75.7	44.5		51.5	49.1
HosL	36.0	35.7	36.1	38.8	59.8	65.1	58.6	74.1	75.3	71.3	72.4	41.4	72.3		87.4
HumL	34.0	34.3	36.1	38.2	58.6	63.4	59.2	74.1	76.4	72.4	73.6	43.7	68.8	92.0	

Fig. 2. Pairwise identities and similarities of the sequences aligned in Fig. 1. Similarities (identities and conserved substitutions scoring ≥0.6 in the mutation-data-matrix-78 [12] derived comparison table [7]) were calculated utilising the program DISTANCES. For each pairwise comparison the sum of similarities or identities were divided by the length of the shorter sequence to generate % identity or similarity (above and below the diagonal, respectively).

all sequences in the alignment are shown in Fig. 2. The highest amino acid sequence similarity between EcBFR and any of the ferritin sequences in Fig. 1 occurs with chicken H-chain ferritin (21.5% identity and 39.2% similarity). The average sequence identity and similarity between EcBFR and all 12 ferritin sequences is 17.2% and 33.9%, respectively.

The subunits of ferritins from mammals [3] are each comprised of a bundle of 4 long α-helices (A–D) and a short helix (E) which together account for 75% of the total secondary structure. Modelling studies suggest that other ferritins have very similar conformations. Although the detailed structure of BFR is unknown, secondary-structure predictions indicate that the BFR subunit also has a high (approx. 80%) helix content, consistent with the presence of a ferritin-like four-helix-bundle [5]. The predicted helical regions of EcBFR and the defined helical regions of ferritin are indicated in Fig. 1. Combined secondary-structure predictions from ferritin sequences give very accurate predictions [16] and, likewise, the predicted helical regions of BFR are highly congruent with the known helical regions of ferritins: 92.7% of the predicted 'helix residues' of EcBFR are aligned with ferritin helices and 85.7% of ferritin helix residues are aligned with those predicted for EcBFR (Fig. 1). This provides further evidence for the proposed evolutionary relationship supporting the notion that BFR possesses a very similar subunit conformation and quaternary structure to those of ferritin.

The B and C helices of ferritin are connected by a long non-helical loop (L) of 17 residues which spans the helix bundle from end to end. According to the alignment (Fig. 1), the equivalent region in BFR comprises just 15 residues, but the same connectivity as ferritin may be achieved by a minor structural rearrangement of the loop. Four gaps were required in the EcBFR sequence to achieve alignment with the ferritin sequences (Fig. 1). Three lie between, or at the beginning of, helices. The fourth is a single residue gap at the position of a kink in the D helix of ferritin caused by an extra residue in one turn of helix. These changes would probably not impose major structural discrepancies between ferritin and BFR. No gaps were required in the ferritin profile for optimum alignment with BFR.

3.3. *Conservation of ferroxidase-centre residues*

The alignment in Fig. 1 indicates that the ferroxidase-centre residues of H-like ferritins are either absolutely conserved or conservatively substituted in BFR. These findings suggest that BFR possesses a ferroxidase centre located, as for ferritin, in the centre of its helix bundle (Fig. 3).

3.4. *Putative haem pockets*

The coaxial haem-iron ligation of BFRs from 3 species, including *E. coli*, has been established as bis-methionine: a haem coordination that has not been reported for any other haemoprotein [25]. Preliminary modelling [26] highlights 2 potential haem sites: site I, an intra-subunit site near the outer molecular surface with haem-iron ligands Met^{31} and Met^{86} (1 haem per subunit); site II, an inter-subunit site near the cavity surface with ligands Met^{52} and $Met^{52'}$ from diad-related neighbours (1 haem per 2 subunits). No other pair of methionines is predicted to serve as haem-iron ligands. However, Met^{31} is not conserved in *Nitrobacter winogradskyi* BFR and Met^{52} is not conserved in the BFR of *Synechocystis* PCC 6803 [27], so the location of haem binding remains uncertain.

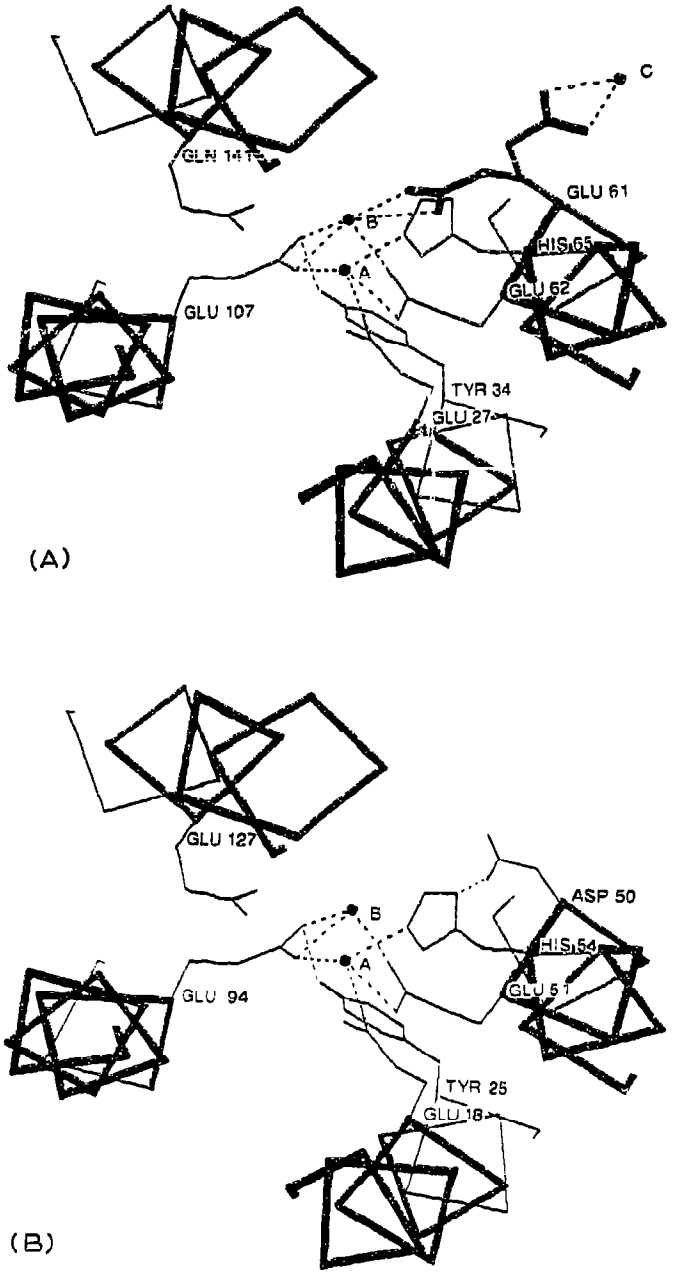

Fig. 3. (A) Schematic diagram of the ferroxidase centre of human H-chain ferritin [3] showing the 7 highly conserved residues which either participate directly in metal ligation or indirectly through water molecules. Metal sites A, B and C, observed in crystals of human H-ferritin containing $TbCl_3$ [3], are postulated as iron sites. The side chain of Glu^{61} occupies 2 alternative positions and may assist the movement of Fe(III) from site B to site C on the cavity surface for iron-core nucleation. (B) Schematic representation of the putative ferroxidase-centre region of the BFR of *E. coli*, modelled on the predicted structural similarity with human H-chain ferritin. Note that 2 of the 7 residues shown in (A) are replaced: Glu^{127} has replaced Gln^{141}; Asp^{50}, substitutes for Glu^{61}.

3.5. *The ferritin-like protein (FTN) of E.* coli

The alignment (Fig. 1) shows that all 7 ferroxidase-centre residues are absolutely conserved in FTN. Only 4 gaps were required to align the FTN sequence with the other ferritin sequences, and these are all at similar positions to those of EcBFR. FTN has an average identity and similarity with all other ferritin sequences in Fig. 1 of 21.4% and 38.3%, respectively. These observations provide further support for the view that FTN is a functional ferritin.

3.6. *Evolution*

The sequence similarity scores (Fig. 2) suggest that ferritins (class I ISPs) and BFRs (class II ISPs) evolved from a common ancestral iron-storage protein within bacteria via a gene duplication event followed by divergence. The 2 classes of ISP were apparently retained by bacteria, but eukaryotes seem to have lost the class II ISP and, in many cases, possess multiple versions of class I. Many questions regarding the physiological roles of BFR and FTN remain to be answered, such as: can FTN subunits assemble to form a spherical shell and a functional ferritin, can FTN and BFR co-assemble, and are there other ISPs (possibly novel classes) yet to be found in bacteria or other species?

The discovery of an evolutionary relationship between BFRs and ferritins and the implied conformational similarities allow the BFR structure to be modelled upon that of ferritin (now in progress in Sheffield). Comparison of the structures of the 2 classes of ISP will highlight those common features essential for their role in cellular homeostasis as iron-sequestering agents.

Acknowledgements: This work was supported by the Wellcome Trust and the Science and Engineering Research Council. We thank the SERC for providing the computational facilities at Daresbury, the Woolfson Foundation and Wellcome Trust for financial support of the graphics facility, and W. Bottke and J.-F. Briat for communicating results prior to publication.

REFERENCES

[1] Theil, E.C. (1987) Annu. Rev. Biochem. 56, 289–315.
[2] Lawson, D.M., Treffry, A., Artymuik, P.J., Harrison, P.M., Yewdall, S.J., Luzzago, A., Cesareni, G., Levi, S. and Arosio, P. (1989) FEBS Lett. 254, 207–210.
[3] 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. and Harrison, P.M. (1991) Nature 349, 541–544.
[4] 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. and Yewdall, S.J. (1991) Adv. Inorg. Chem. 36, 449–486.
[5] Andrews, S.C., Smith, J.M.A., Guest, J.R. and Harrison, P.M. (1989) Biochem. Biophys. Res. Commun. 158, 489–496.
[6] Izuhara, M., Takamune, K. and Takata, R. (1991) Mol. Gen. Genet. 225, 510–513.
[7] Gribskov, M., Luthy, R. and Eisenberg, D. (1990) Methods Enzymol. 183, 146–159.
[8] Beaumont, C., Dugast, I., Remaudie, F., Souroujon, M. and Grandchamp, B. (1989) J. Biol. Chem. 264, 7498–7504.
[9] Murray, M.T., White, K. and Munro, H.N. (1987) Proc. Natl. Acad. Sci. USA 84, 7438–7442.
[10] Moskaitis, J.E., Pastori, R.L. and Schoenberg, D.R. (1990) Nucleic Acids Res. 18, p. 2184.

[11] Feng, D.-F. and Doolittle, R.F. (1990) Methods Enzymol. 183, 375–387.
[12] Schwartz, R.M. and Dayhoff, M.O. 1979, in: Atlas of protein sequence and structure (Dayhoff, M.O. ed.) pp. 353–358, National Biomedical Research Foundation, Washington DC.
[13] de Graaf, J., Amons, R. and Moeller, W. (1990) Eur. J. Biochem. 193, 737–750.
[14] Barker, W.C., George, D.G. and Hunt, L.T. (1990) Methods Enzymol. 183, 31–49.
[15] Bairoch, A. and Boeckmann, B. (1991) Nucleic Acids Res. 19, 2247–2249.
[16] Harrison, P.M., Andrews, S.C., Ford, G.C., Smith, J.M.A., Treffry, A. and White, J.L. (1987) in: Iron Transport in Microbes, Plants and Animals (Winkelmann, G., van der Helm, D. and Neilands, J.B. eds) pp. 445–476, VCH Publishers, D-6940 Weinheim.
[17] Wilkins-Stevens, P., Dodgson, J.B. and Engel, J.D. (1987) Mol. Cell. Biol. 7, 1751–1758.
[18] Heuterspreute, M. and Crichton, R.R. (1981) FEBS Lett. 129, 322–327.
[19] Boyd, D., Vecoli, C., Bellcher, D.M., Jain, S.K. and Drysdale, J.W. (1985) J. Biol. Chem. 260, 11755–11761.
[20] Dietzel, J., Hirzmann, J., Preis, D., Symmons, P. and Kunz, W. (1991), Mol. Biochem. Parasitol. (in press).
[21] Lescure, A.M., Proudhon, D., Pesey, H., Ragland, M., Theil, E.C. and Briat, J.-F. (1991) Proc. Natl. Acad. Sci. USA 88 (in press).
[22] Dickey, L.F., Sreedharan, S., Theil, E.C., Didsbury, J.R., Wang, Y.H. and Kaufmann, R.E. (1987) J. Biol. Chem. 262, 7901–7907.
[23] Andrews, S.C., Keen, J.N., Findlay, J.B.C., Guest, J.R., Harrison, P.M. and Smith, J.M.A. (1991) Biochim. Biophys. Acta, 1078, 111–116.
[24] Kurokawa, T., Fukomori, Y. and Yamanata, T. (1989) Biochim. Biophys. Acta 976, 135–139.
[25] Cheesman, M.R., Thomson, A.J., Greenwood, C., Moore, G.R. and Kadir, F. (1990) Nature 346, 771–773.
[26] le Brun, N.E., Cheesman, M.R., Kadir, F.H., Thomson, A.J., Moore, G.R., Andrews, S.C., Smith, J.M.A., Guest, J.R. and Harrison, P.M. (1991) 10th International Conference on Iron and Iron Proteins, P10, Oxford, July 27–31, 1991.
[27] Laulhere, J.P., Laboure, A.M., van Wuytswinkel, O., Gagnon, J. and Briat, J.-F. (1991) Biochem. J. (in press).
[28] Grossman, M.J., Hinton, S.M., Minak-Bernero, V. and Stiefel, E.I. (1991) 10th International Conference on Iron and Iron Proteins, P10, Oxford, July 27–31, 1991.

NOTE ADDED IN PROOF

The complete amino acid sequence of the BFR of *A. vinelandii* has recently been determined and a possible evolutionary relationship with ferritin has been derived independently [28].