Crystal structure of chicken riboflavin-binding protein

Hugo L. Monaco

Department of Genetics, University of Pavia, Via Abbategrosso 207, 27100 Pavia, Italy

The crystal structure of chicken egg white riboflavin-binding protein, determined to a resolution of 2.5 Å, is the prototype of a family that includes other riboflavin- and folate-binding proteins. An unusual characteristic of these molecules is their high degree of cross-linking by disulfide bridges and, in the case of the avian proteins, the presence of stretches of highly phosphorylated polypeptide chain. The structure of chicken egg white riboflavin-binding protein is characterized by a ligand-binding domain and a phosphorylated motif. The ligand-binding domain has a fold that appears to be strongly conditioned by the presence of the disulfide bridges. The phosphorylated motif, essential for vitamin uptake, is made up of two helices found before and after the flexible phosphorylated region. The riboflavin molecule binds to the protein with the isoalloxazine ring stacked in between the rings of Tyr75 and Trp156. This geometry and the proximity of other tryptophans explain the fluorescent quenching observed when riboflavin binds to the protein.

Keywords: crystal structure/fluorescence quenching/folate-binding protein/phosphoglycoprotein/riboflavin-binding protein

Introduction

Transport of riboflavin in mammalian plasma appears to be non-specific under normal conditions. Both albumin (Jusko and Levy, 1975) and riboflavin-binding immunoglobulins (Innis *et al.*, 1985; McCormick *et al.*, 1987) are riboflavin carriers, but in recent years evidence has emerged indicating that normal transport mechanisms are no longer sufficient during pregnancy. Since adequate amounts of riboflavin are essential for normal fetal development, it is not surprising that under these more demanding conditions a specific carrier system has evolved with the special task of vitamin delivery to the developing embryo. Thus, pregnancy-specific riboflavin-binding proteins have been found in rat (Muniyappa and Adiga, 1980), mice (Natraj *et al.*, 1987), bovine (Merril *et al.*, 1979), simian (Visweswariah and Adiga, 1987) and human (Murthy and Adiga, 1982b) plasma. Although these carriers have not been characterized thoroughly, they all appear to share a common feature: their similarity to the well-known chicken riboflavin-binding proteins (RfBPs). Antibodies against chicken RfBP have been found to cause the end of pregnancy in rats (Murthy and Adiga, 1982a; Krishnamurthy *et al.*, 1984), mice (Natraj *et al.*, 1987) and the bonnet monkey (Visweswariah and Adiga,

1987), thus showing that there is structural similarity between the avian RfBP and the mammalian pregnancy-specific riboflavin-binding proteins and that these proteins play an essential role in the survival of the fetus.

The family of the chicken RfBPs includes three well-known proteins that are the product of the same gene but have undergone different post-translational modifications (White and Merrill, 1988). The proteins can be purified from egg white (Rhodes *et al.*, 1959), egg yolk (Ostrowski *et al.*, 1962) and from the plasma of laying hens (Miller *et al.*, 1982a). Egg white RfBP is synthesized by the oviduct cells (Mandeles and Ducay, 1962), plasma RfBP is produced in the liver under estrogen control and yolk RfBP is the result of a proteolytic cleavage of the last 11–13 amino acids of plasma RfBP when the molecule crosses the oocyte membrane (Norioka *et al.*, 1985). The three proteins have a single binding site for riboflavin with a dissociation constant of 1.3 nM in the pH range 6–9; removal of the vitamin from the holoprotein is accomplished at lower pH where the affinity for the specific ligand is substantially reduced (Müller and van Berkel, 1991). Plasma and egg white RfBP share a sequence which is 219 amino acids long (Hamazume *et al.*, 1984) but they have different carbohydrates attached in the same positions: Asn36 and Asn147. The primary structures of the carbohydrates attached to plasma (Rohrer and White, 1992) and yolk (Tarutani *et al.*, 1993) RfBP have been determined and shown to be identical. Although the carbohydrate composition of the egg white RfBP is known, the primary structure of its oligosaccharide chains has not yet been determined.

It was shown recently that all three RfBPs can associate with the lipid carrier vitellogenin and that the macromolecular complex of the two carriers recognizes a multifunctional oocyte-specific lipoprotein receptor (Mac Lachlan *et al.*, 1994). A peculiar characteristic of the three chicken RfBPs is the presence of a highly phosphorylated region that extends from amino acids 186 to 197, where there are up to eight phosphates bound to serines (Hamazume *et al.*, 1984). This region is involved in oocyte uptake of the plasma protein (Miller *et al.*, 1982b), possibly by playing an essential role in the interaction of plasma RfBP with vitellogenin (Mac Lachlan *et al.*, 1994). Another unique characteristic of the three chicken RfBPs is the presence of nine disulfide bridges, whose assignment was made chemically in the egg white protein (Hamazume *et al.*, 1987), that play an important structural role since refolding of the molecule is fast and efficient when denaturation is carried out, leaving them intact (McClelland *et al.*, 1995). The interactions of the chicken RfBPs with different flavin derivatives are probably the best studied of any flavoprotein (Lubas *et al.*, 1977; Walsh *et al.*, 1978; Nishina *et al.*, 1980; Becvar and Palmer, 1982; Matsui *et al.*, 1982; Wessiak *et al.*, 1984). Upon binding to the

Fig. 1. (A) Ribbon diagram of the riboflavin-binding protein molecule. The ligand-binding domain is colored in green and the phosphorylated motif in red. The vitamin is represented by the yellow stick drawing. The letters 'N' and 'C' identify the N- and C-termini of the protein, the labels CHO-36 and CHO-147 indicate the positions of the oligosaccharide chains. The gap between the two helices in the red motif is the disordered part of the molecule which corresponds to the phosphorylated region between Leu184 and Glu199. The diagram was produced with the program MOLSCRIPT (Kraulis, 1991). (B) Secondary structure topology of RfBP. α -Helices are shown as cylinders, β structure is shown as arrows and non-repetitive loops as solid lines.

chicken RfBPs, riboflavin loses its characteristic fluorescence, a fact interpreted as due to a stacking of the isoalloxazine ring of the vitamin with aromatic side chains of the protein molecule (Blankenhorn, 1978). A search for sequence homology reveals that the RfBPs are structurally related to the folate-binding proteins (FBPs) (Zheng *et al.*, 1988) but not to any protein of known three-dimensional structure.

The crystal structure of hen egg white RfBP, the first member of this new structural family, is presented here. It is expected that the model described will serve as a starting point for structural studies of other RfBPs and also for the FBPs.

Results and discussion

Overall structure of chicken RfBP

Chicken RfBP is a globular monomeric protein of approximate dimensions $50 \times 40 \times 35$ Å (Figure 1). The most distinctive feature of this new fold is the presence of a ligand-binding domain which runs from the N-terminus up to about Cys169, and a phosphorylated motif which runs from there to the C-terminus. In the ligand-binding domain, there is a cleft ~ 20 Å wide and 15 Å deep that accommodates the bound vitamin. Figure 2 shows the electron density of the isoalloxazine ring of riboflavin stacked in between the side chains of Tyr75 and Trp156. About 30% of the residues in RfBP are found in α -helices and a little less than 15% in β structure. There are a total of six α -helices, designated A–F, and four series of discontinuous areas of β structure (a, b, c and d). Helices B and D each have a three residue interruption in a position where there is a pronounced

Fig. 2. Electron density of the riboflavin molecule and the rings of Tyr75 and Trp156 in the solvent-flattened MIR map. The map is contoured at a level of 1.0σ . The final model is colored by atom type: C, yellow; N, blue; O, red. The figure was produced with the program O (Jones *et al.*, 1991).

bend that completely changes their direction. The first three areas of β structure, a, b and c, present one, two and three gaps filled by coil. The longest continuous portion of β structure in the molecule is about five residues long. The ligand-binding domain includes helices A–D and the four areas of β structure; the phosphorylated

Table I. Disulfide bridges of chicken riboflavin-binding protein

[coil]	Cys5–Cys32	[coil]
[coil]	Cys24–Cys73	[helix]
[coil]	Cys33–Cys77	[helix]
[coil]	Cys57–Cys138	[coil]
[helix]	Cys64–Cys110	[helix]
[strand]	Cys99–Cys169	[coil]
[helix]	Cys103–Cys152	[helix]
[coil]	Cys116–Cys134	[strand]
[coil]	Cys167–Cys202	[helix]

Assignment of the cysteine residues of chicken RfBP to the categories of helix, strand and coil, defined according to the program PROCHECK (Laskowski *et al.*, 1993).

motif is made up of helices E–F and, in between, the phosphorylated region which is not well ordered in the maps. Other regions of the molecule that are not ordered in the crystals are the first two and the last seven residues. The structure of the ligand-binding domain appears to be strongly influenced by the presence of eight out of a total of nine disulfide bridges. The ninth bridge, which links Cys167 to Cys202, anchors helix F to the ligand-binding domain and is probably the reason why this portion of the map is ordered, although it follows the flexible phosphorylated area in the sequence. Characteristic of the topology of the ligand-binding domain is an alternation of the α -helices and the areas of discontinuous β structure (see Figure 1B). In the topological diagram, the arrows do not represent two pairs of antiparallel β -strands but areas of β structure that have twists, bends and gaps and which are found in between the helices represented in the figure. The areas of β structure are very complex, as can be appreciated by inspection of Figure 1A.

Table I lists the secondary structure assignments of the 18 cysteines present in the molecule, all of them participating in disulfide bridges. Nine of these residues are assigned to coils, seven to helices and two to the areas of β structure. Helix C is disulfide-bridged to both helices B and D. The first 35 amino acids in the sequence, the longest portion of the chain that is virtually without elements of repetitive secondary structure, include Cys5, Cys24, Cys32 and Cys33 participating in three disulfide bridges.

The position of the two oligosaccharide chains of the molecule is indicated in Figure 1A. The first chain, bound to Asn36, is found at the beginning of helix A, and the second, linked to Asn147, is in the middle of the coil present in the bend of helix D. Although clear electron density is observed in both cases for the proximal sugar residues, no attempt was made to include them in the model since their primary structure is unknown.

It has been pointed out (Hamazume *et al.*, 1984) that there exists genetic variability at position 14 that can be either a lysine or an asparagine. Since this position is on the surface of the molecule and the electron density of its side chain is not well defined, it has not been possible to decide which of the two residues is present in these crystals.

A search of the EMBL database using the program DALI (Holm and Sander, 1993) confirmed that the fold of chicken RfBP is not similar to that of any of the 680 proteins that were included in that file

Table II. Inter-atom distances at the ligand-binding site

Ser50	CO–	2.9 Å	–O ₄ *	Rfl
	OG–	4.2 Å	–O ₄ *	Rfl
Ser52	OG–	2.4 Å	–O ₄ *	Rfl
Ile71	CG2–	3.4 Å	–C7M	Rfl
	CG2–	3.4 Å	–C8M	Rfl
Glu72	OE2–	2.6 Å	–O ₃ *	Rfl
Tyr75	PLANE–	3.7 Å	–PLANE	Rfl
	OH–	4.0 Å	–N3	Rfl
	OH–	4.2 Å	–O2	Rfl
	OH–	4.8 Å	–O4	Rfl
Arg76	NH1–	4.0 Å	–O ₃ *	Rfl
	NH1–	3.6 Å	–OE1	Glu72
Ala91	CO–	3.6 Å	–N3	Rfl
	CB–	4.3 Å	–C4	Rfl
Ala92	CB–	4.3 Å	–C4	Rfl
Leu121	CD2–	3.8 Å	–C4	Rfl
Arg126	NH2–	3.6 Å	–O ₅ *	Rfl
Trp156	PLANE–	3.7 Å	–PLANE	Rfl
	CZ3–	3.7 Å	–C9	Rfl
Ser159	OG–	4.0 Å	–C6	Rfl
Phe160	CZ–	3.6 Å	–C7M	Rfl
Rfl	N10–	2.9 Å	–O ₂ *	Rfl
Rfl	O ₂ *–	3.6 Å	–O ₃ *	Rfl
Rfl	O ₃ *–	2.9 Å	–O ₄ *	Rfl
Rfl	O ₄ *–	2.6 Å	–O ₅ *	Rfl

Main inter-atom distances between the riboflavin molecule and the RfBP amino acids in closest contact with it.

The ligand-binding site

The electron density for riboflavin and the amino acid side chains that are in contact with it is very clear in the DM multiple isomorphous replacement (MIR) map (see Figure 2). Binding of riboflavin occurs in a cleft with the vitamin isoalloxazine ring stacked between the parallel planes of Tyr75 and Trp156. A question which was less simple to resolve was whether binding takes place with the pyrimidine or the xylene moieties buried most deeply in the protein. The isoalloxazine ring of flavins is amphipathic since the xylene portion is hydrophobic and the pyrimidine moiety hydrophilic. When flavins bind to a protein with the hydrophilic end buried, as for example in the case of flavodoxins (Watenpaugh *et al.*, 1973; Rao *et al.*, 1992), a series of strong hydrogen bonds to the hydrophilic portion of the molecule characterizes the protein–ligand interactions. The electron density for the vitamin and the position of the protein side chains indicated quite clearly that in RfBP this is not the case, i.e. that in RfBP it is the xylene moiety of the triple ring that is buried most deeply in the protein. This mode of binding is in agreement with the results of many studies of flavin derivative binding to chicken RfBP (Merrill and McCormick, 1978; Choi and McCormick, 1980; Mifflin and Langerman, 1983; Ghisla and Massey, 1986). Table II lists the main inter-atom distances measured between the protein side chains in the ligand-binding site and the vitamin groups near them and in between atoms of the ribityl moiety. Figure 3A is a stereo diagram of the main side chains present in the ligand-binding site, and Figure 3B is a schematic representation of their interactions. As expected, the major interactions of the isoalloxazine ring with the protein are hydrophobic.

A characteristic feature of ligand binding to RfBP is the almost complete fluorescence quenching observed upon binding of not only the natural ligand, riboflavin,

Fig. 3. (A) Stereo diagram showing the amino acids in the ligand-binding site that are in closest contact with the bound vitamin. The distances between the atoms are given in Table II. (B) Schematic representation of the interactions shown in (A).

Fig. 4. Stereo diagram showing the relative positions of the isoalloxazine ring of riboflavin and of the side chains of Tyr75 and of all the tryptophans present in the molecule. Some pertinent distances are: Trp156 (CZ2)–4.0 Å–Trp120 (CE3); Trp156 (CH2)–4.0 Å–Trp120 (CE3); Trp156 (CD1)–4.5 Å–Trp106 (CH2); Trp120 (CZ3)–3.9 Å–Trp124 (CZ2); Trp120 (CH2)–3.9 Å–Trp124 (CZ2); Trp120 (CZ3)–4.5 Å–Trp54 (CD1).

but of flavin analogs as well. Figure 4 is a stereo diagram that shows Tyr75 and all the tryptophan side chains present in the protein molecule. With the exception of Trp84, all the other tryptophans are close to the active site. Trp120 is ~4 Å distant from the crucial Trp156 stacked onto the vitamin plane and roughly at the same distance from Trp124. Trp106 is only slightly further from Trp156, ~4.5 Å. The same distance separates Trp120 from Trp54.

Thus, five out of the six tryptophans present in the protein molecule cluster in the vicinity of the ligand-binding site, and the plane of one of them (Trp156) is in direct contact with the plane of the ligand.

Bound riboflavin was found to protect an essential carboxyl group in the ligand-binding site from inactivation by carbodiimide (Kozik, 1982a). The position of the residue in the sequence has not been determined, but

Fig. 5. Position of the nine disulfide bridges relative to the riboflavin molecule. The two bridges that are closest to the vitamin are Cys24–Cys73 (SG 73–9.5 Å–C8M) and Cys103–Cys152 (SG 152–9.8 Å–C7M).

inspection of Table II reveals that the most likely candidate is Glu72, which is hydrogen-bonded to the O_3^* group of the ribityl moiety of the vitamin.

Cleavage of a single disulfide bond results in a loss of the binding capacity of RfBP, but all the disulfide bridges in the holoprotein show the same reactivity as in the apoprotein, indicating that the critical bridge is not located in the ligand-binding site (Kozik, 1982b). Figure 5 shows the position of the bridges relative to the bound riboflavin. None of them is at the binding site. The two that are closest are Cys24–Cys73 and Cys103–Cys152, which are at a distance of ~9–10 Å from the methyl groups of the xylene moiety of the isoalloxazine ring.

The phosphorylated motif

The phosphorylated motif is found after Cys169, a member of the last bridge which is located exclusively in the ligand-binding domain, and it continues to the C-terminus of the protein molecule (see Figure 1A). A remarkable feature is the highly anionic region that runs from residue 186 to 199 (E-PS-PS-E-E-PS-PS-PS-M-PS-PS-PS-E-E, PS = phosphoserine). If the phosphates are removed with acid phosphatase, plasma protein uptake by the oocyte decreases dramatically (Miller *et al.*, 1982b). After it was shown that RfBP binds to an oocyte-specific receptor in association with vitellogenin, it was suggested that this interaction involves bridges with calcium ions between the phosphoserines of the two proteins (Mac Lachlan *et al.*, 1994). Unlike most of the rest of the protein molecule, this anionic region is not ordered in the electron density maps, but the polypeptide chain immediately before and after it folds clearly into the last two helices of the molecule, E and F. An attempt to render it more ordered was made by soaking a crystal in $CaCl_2$, but the two electron density maps, in the presence and in the absence of calcium, did not differ significantly. The phosphorylated motif can thus be described as made up of a flexible anionic region which is inserted in between

Fig. 6. Succinate binding in the cleft that separates the ligand-binding domain (green) and the phosphorylated motif (yellow). The succinate and riboflavin molecules are represented in red. The main effect of succinate binding appears to be a rigid body rotation about an axis not far from that of the first portion of helix D.

two antiparallel helices. It has been noticed that the β -caseins have phosphorylated regions preceded and followed by chains that are similar in sequence to that of chicken RfBP (Holt and Sawyer, 1988). Therefore, structurally analogous motifs may be present in the similarly phosphorylated caseins. The second helix of the motif, F, is anchored to the ligand-binding domain through the Cys167–Cys202 bridge. The two helices are found on the surface of the molecule and the anionic region protrudes into the solvent, which is compatible with the function which has been assigned to the motif (see Figure 1).

Succinate binding

The chicken egg white RfBP crystals were grown using ammonium sulfate as precipitant (Zanette *et al.*, 1984). When crystals grown in this mother liquor were transferred to solutions that contained 3.8 M ammonium succinate instead of sulfate, it was found that the crystals not only retained their integrity but they even diffracted to slightly better resolution. The unit cell parameters changed by <1% but the mean fractional isomorphous difference was found to be >40% (see below).

Inspection of the model of the protein in this second crystal form showed that the main effect of succinate binding was a rigid body rotation about an axis that is close to that of the first portion of helix D. An $F_{obs}-F_c$ map of the succinate form, calculated with the refined model, revealed that there was residual density that could be associated with four or five bound succinate molecules. The most interesting was present in a cleft found in between the ligand-binding domain and the phosphorylated motif (see Figure 6). The $F_{obs}-F_c$ map of the equivalent sulfate form showed that this density was not present there. A succinate molecule that explained the density found was therefore added to the protein model of the first crystal form. Further refinement with the added

C-RFBP	Q	Q	Y	G	C	L	E	G	A	D	T	H	H	K	A	K	P	S	P	E	P	N	M	H	E	-	Q	C	T	L	P	W	R	E	S	N	A	C	C	C	Y	S	S	T	N	N	F	T	T	E	S	Q	I	Q		40	
B-FBP	L	L	N	V	C	M	D	A	K	D	H	H	K	A	E	K	P	S	G	P	E	D	S	L	H	E	-	Q	C	R	P	P	W	R	K	N	A	C	C	C	S	S	A	V	T	N	N	F	T	T	E	S	Q	I	Q		52
H-FBP	L	L	N	V	C	M	D	A	K	D	H	H	K	A	E	K	P	S	G	P	E	D	S	L	H	E	-	Q	C	R	P	P	W	R	K	N	A	C	C	C	S	S	A	V	T	N	N	F	T	T	E	S	Q	I	Q		72
C-RFBP	L	A	H	S	P	I	S	K	V	S	N	S	Y	W	N	R	C	G	Q	L	S	K	A	C	E	D	F	T	K	K	I	E	C	F	Y	R	C	S	P	H										80							
B-FBP	E	A	K	K	D	I	S	Y	L	Y	R	F	N	W	D	H	C	G	Q	L	S	K	A	C	E	D	F	T	K	K	I	E	C	F	Y	R	C	S	P	N									92								
H-FBP	E	A	K	K	D	I	S	Y	L	Y	R	F	N	W	D	H	C	G	Q	L	S	K	A	C	E	D	F	T	K	K	I	E	C	F	Y	R	C	S	P	N									112								
C-RFBP	A	A	R	W	I	-	-	-	D	P	R	Y	T	A	A	I	Q	S	V	P	L	C	Q	S	E	D	F	C	D	S	W	Y	E	A	C	C	K	D	D	S	I	T	A	H	N	W		120									
B-FBP	L	G	P	W	I	R	E	V	N	Q	S	W	R	K	E	R	V	L	G	V	P	L	C	Q	S	E	D	F	C	D	S	W	Y	E	A	C	C	K	D	D	S	I	T	A	H	N	W		136								
H-FBP	L	G	P	W	I	R	E	V	N	Q	S	W	R	K	E	R	V	L	G	V	P	L	C	Q	S	E	D	F	C	D	S	W	Y	E	A	C	C	K	D	D	S	I	T	A	H	N	W		156								
C-RFBP	L	T	D	W	E	R	D	E	S	G	E	N	H	C	K	S	K	-	-	C	V	P	P	Y	S	E	M	Y	P	P	Y	G	T	D	M	C	Q	N	E	M	W	E	S	S	F		160										
B-FBP	H	K	G	W	N	W	-	T	S	G	Y	N	K	C	P	V	K	A	A	C	H	P	P	F	H	F	Y	S	E	M	Y	P	P	T	A	V	L	C	N	E	M	W	E	S	S	Y		177									
H-FBP	H	K	G	W	N	W	-	T	S	G	Y	N	K	C	P	V	K	A	A	C	H	P	P	F	H	F	Y	S	E	M	Y	P	P	T	A	V	L	C	N	E	M	W	E	S	S	Y		197									
C-RFBP	K	V	S	E	-	-	-	S	S	C	L	R	C	L	Q	M	N	K	K	D	M	V	A	I	K	H	L	L	S	E	S	S	E	E	S	S	M	S	S	E	E	H					200										
B-FBP	K	V	S	N	Y	S	R	G	S	C	G	R	C	C	I	Q	M	N	-	-	-	-	A	F	D	P	F	Q	Q	N	P	N	E	E	V	A	R	F	-	Y	A	E	N	P		213											
H-FBP	K	V	S	N	Y	S	R	G	S	C	G	R	C	C	I	Q	M	N	-	-	-	-	A	F	D	P	F	Q	Q	N	P	N	E	E	V	A	R	F	-	Y	A	E	N	P		233											
C-RFBP	A	C	Q	K	K	L	L	K	F	E	A	L	Q	Q	E	E	G	E	R	R																												221									
B-FBP	T	S	G	S	T	P	Q	G	I			L	L	S	L	A	L	M	L	L	W	L	L	S																							222										
H-FBP	S	G	A	G	P	W	A	A	W	P	F	L	L	S	L	A	L	M	L	L	W	L	L	S																							257										

Fig. 7. Comparison of the amino acid sequences of chicken RfBP (Hamazume *et al.*, 1984), bovine milk folate-binding protein (Svendsen *et al.*, 1984) and human folate-binding protein (Elwood, 1989). The boxed amino acids are identical in chicken RfBP and either one of the folate-binding proteins.

succinate showed that the new moiety behaved quite reasonably, and in the end its temperature factors converged to values that were ~50 for all its atoms (the temperature factors of the other regions of density present in the succinate and absent in the sulfate maps refined all to higher values).

The carboxyl end of the liganded succinate which is buried deepest in the cleft appears to be hydrogen-bonded to the side chain of His80. The other end is hydrogen-bonded to Lys174 and the peptide carbonyl of Leu26. The residues which are in closest contact with the succinate moiety are Leu26 and Tyr27, Pro79 and His80 in the ligand-binding domain and Lys174, Asp176, Met177 and Leu180 in the phosphorylated motif. The residues of the ligand-binding domain are found in a helical turn of the coiled region that precedes helix A and in the bend of helix B. The amino acids of the phosphorylated motif are part of helix E.

The fact that this succinate-binding cleft is found at the interface between domain and motif suggests that it may not be a mere artifact but it may have some as yet unidentified function. Both egg white (Mano *et al.*, 1992) and yolk (Massolini *et al.*, 1995) RfBPs have been used as stationary phases for the separation of chiral compounds that in some cases have structures that are quite different from that of riboflavin; some of those compounds may bind in this cleft.

Structural homology with folate-binding proteins

Figure 7 compares the amino acid sequence of chicken RfBP (Zheng *et al.*, 1988) with those of bovine milk FBP (Svendsen *et al.*, 1984) and FBP from human malignant tissue culture (KB) and placenta cells (Elwood, 1989). The boxed residues are those that are identical in RfBP and at least one of the FBPs. The sequence similarity of the residues that in RfBP fold into the ligand-binding domain to equivalent portions in the two FBPs justifies the proposal that a structurally similar binding domain is present in the FBPs. The FBPs play an essential role in the distribution and assimilation of folic acid (for a review, see Henderson, 1990). They can be either water soluble or membrane bound; bovine milk FBP is a prototype of

the first class while human FBP has a C-terminal portion which is hydrophobic enough to make it compatible with a membrane-anchoring motif (Elwood, 1989).

It is worth noticing that the 16 cysteines that in RfBP form the eight bridges present in the ligand-binding domain are conserved in all the three proteins. The C-terminal portion of the molecules, on the other hand, shows very little similarity, which is what would be anticipated for motifs with different functions.

The structure established here is thus related more closely to the FBPs than to the known structures of other flavoproteins which bind ligands that contain the flavin moiety.

Materials and methods

Protein purification and crystallization

The purification and crystallization of chicken egg white RfBP have been described previously (Zanette *et al.*, 1984). The crystals were stable for a very long time if kept in the dialysis microcells in 43% (w/v) ammonium sulfate, 0.05 M Tris pH 8.5. When the crystals were transferred to a solution of 3.8 M ammonium succinate, 0.05 M Tris pH 8.5, the unit cell parameters changed by <1% but the mean fractional isomorphous difference between the two native data sets was 43.2%. The structure was solved using crystals transferred to the succinate mother liquor because resolution and data quality were slightly better in this medium.

Data collection and reduction

All data were collected at room temperature from crystals with typical dimensions 1.5×0.4×0.4 mm (Table III). The native data sets in succinate and sulfate and the Pt and 2 (acetoxy-mercuri)-4-nitrophenol (MNP) derivatives were collected on a Rigaku R-axis II imaging plate area detector mounted on a Rigaku RU-200 rotating anode X-ray generator. The source was operated at 50 kV and 160 mA using a focal spot size of 0.3×3 mm. Monochromatic copper K α radiation was obtained using a graphite crystal monochromator. These data were processed using the R-axis program package. The mercury derivative was collected on a Mar Research imaging plate area detector at the X11 beamline of the EMBL outstation of the DESY synchrotron (Hamburg, Germany). The wavelength was 0.9 Å and the data were processed using the DENZO package (Otwinowski, 1993). Friedel pairs were kept separated to be used in the phasing.

Phase determination

The Patterson difference functions of the Pt and the Hg derivatives collected in the laboratory were interpreted using programs of the CCP4

Table III. Data collection and MIR statistics

	Dataset	No. of reflections		R_{merge}^a (%)	MFID ^b (%)	Sites	FH /E ^c	R_c^d
		Total	Unique (% complete)					
Native 1 (in succinate)	overall ^e	72 829	17 056 (94.2)	6.85				
	10.0–2.8 Å		13 256 (97.3)	4.60				
	2.8–2.5 Å		3800 (84.9)	10.00				
Native 2 (in sulfate)	10.0–2.5 Å	39 601	14 936 (79.6)	8.47	43.2			
Hg(Ac) ₂ ^f	10.0–2.5 Å	116 254	17 833 (98.5)	6.20	18.4	1	2.20 (2.5 Å)	0.51
K ₂ Cl ₄ Pt ^g	10.0–2.5 Å	38 701	14 226 (74.9)	10.00	19.7	3	1.30 (3.5 Å)	0.85
MNP ^h	10.0–2.5 Å	58 680	14 874 (79.0)	11.10	16.1	5	1.10 (3.5 Å)	0.86

^a $R_{\text{merge}} = \sum |I_i - \langle I_i \rangle| / \sum \langle I_i \rangle$, where I_i is the intensity of an observation of reflection i and $\langle I_i \rangle$ is the average intensity for reflection i .

^bMFID = mean fractional isomorphous difference, $\sum |F_{\text{PH}}| - |F_P| / \sum |F_{\text{PH}}|$, where $|F_P|$ and $|F_{\text{PH}}|$ are the structure factor amplitudes of the native protein and the derivative.

^c $|F_H|/E$ is the phasing power and E is the residual lack of closure error.

^d $R_c = \sum ||F_{\text{PH}}| \pm |F_P| - |F_H|| / \sum ||F_{\text{PH}}| \pm |F_P||$.

^eAll the data sets were collected at room temperature. The two native sets, the Pt and MNP derivatives were collected on a Rigaku R-axis II imaging plate area detector mounted on a Rigaku rotating anode X-ray generator. The Hg derivative intensity data were collected using a wavelength of 0.9 Å at the X11 beamline of the EMBL outstation in Hamburg (Germany). The data were recorded on a Mar Research imaging plate and processed keeping the Bijvoet pairs separated to be used in phasing.

^fThe Hg derivative was prepared by first treating, for 10 days, crystals transferred to the succinate mother liquor with the same solution made 10^{-2} M in β -mercaptoethanol at room temperature. The crystals subsequently were transferred to mother liquor containing instead 10^{-2} M mercuric acetate and left in this solution for 10 days at room temperature before mounting.

^gThe platinum derivative was prepared by soaking crystals of the native protein for 2 days in succinate mother liquor made 10^{-2} M in K₂PtCl₄. The crystals were backwashed for 2 h before mounting.

^hThe MNP derivative was prepared by soaking native crystals for 2 days in a solution made 10^{-2} M in the mercurial.

Table IV. Model refinement statistics

	Model in succinate	Model in sulfate
No. of non-hydrogen protein atoms	1586	1604
Residues	1–184, 199–212	1–184, 196–212
No. of riboflavin atoms	27	27
No. of solvent molecules	1 (succinate)	0
Resolution limits	10.0–2.5 Å	10.0–2.5 Å
R -factor ^a	20.7%	21.7%
No. of reflections	16 635	14 750
Free R -factor (830 reflections)	25.5%	—
R.m.s. deviation in bond lengths ^b	0.006 Å	0.006 Å
R.m.s. deviation in bond angles ^b	1.8°	1.8°
R.m.s. deviation in torsion angles ^b	17.3°	18.3°
Outliers in the Ramachandran plot ^c	0	0

^a R -factor = $\sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o is the observed structure factor amplitude and F_c the structure factor amplitude calculated from the model.

^bThe r.m.s. deviations from the ideal values were calculated with the program TNT (Tronrud *et al.*, 1987).

^cResidues located in the generously allowed and non-allowed regions of the Ramachandran plot according to the program PROCHECK (Laskowski *et al.*, 1993).

(Collaborative Computational Project Number 4, 1994) and SHELX-90 (Sheldrick, 1991) packages. From the preliminary phasing, it became apparent that the second derivative potentially could be exploited for anomalous dispersion phasing, and it was decided to collect Bijvoet pairs under more favorable conditions. The mercury derivative data collected in Hamburg, which included 81% of the pairs up to a resolution of 2.5 Å, were used initially to resolve the ambiguity concerning the crystal space group. The crystals of chicken egg white RfBP are trigonal, with $a = b = 112.5$ Å and $c = 72.0$ Å. The diffraction pattern is compatible with the two enantiomeric space groups P3₁21 and P3₂21. The ambiguity was resolved using the classical method (Blundell and Johnson, 1976). The single major site of the Hg derivative was used to calculate single isomorphous replacement (SIR) plus anomalous dispersion phases which were then used to calculate the intensity of the Pt peaks in difference Fourier syntheses calculated for the two possible space groups. The intensity of the Pt peak corresponding to space group P3₂21 was approximately three times as high as that of the space group P3₁21. The heavy atom positions of the three derivatives were refined and MIR phases were calculated using the program MLPHARE (Otwinowski,

1991). The overall figure of merit for the reflections in the resolution interval from 10 to 2.8 Å was 0.69. The calculated map was then subjected to solvent flattening and histogram matching using the program DM (Cowtan, 1994). The map thus produced (see Figure 2) was of very good quality and readily allowed tracing of the polypeptide chain.

Model building and refinement

The RfBP model was built using the program O (Jones *et al.*, 1991). The presence of good density for the disulfide bridges facilitated the assignment of the primary sequence. Since the primary structure of the oligosaccharide chains is unknown, no attempt was made to include them in the model. In the region between Leu184 and Glu199 no clear density was present in the maps and, therefore, those amino acids are missing from the model. The R -factor for the initial model was 43.8% for the data in the resolution interval between 10 and 2.8 Å. After conventional conjugate gradient least squares refinement using the program XPLOR (Brünger, 1992a), the R -factor decreased to 31.5 and the free R -factor (Brünger, 1992b), calculated for 5% of the reflections, to 34.2%. Subsequent cycles of model building alternated with refinement

brought the *R*-factor to 26.2% and the free *R*-factor to 31.3%. During the process of refinement and model building, the quality of the model was controlled using the program PROCHECK (Laskowski *et al.*, 1993). Additional refinement was done using the program TNT (Tronrud *et al.*, 1987). When the model had an *R*-factor of ~21%, its coordinates were used to calculate an *R*-factor using the data collected from the native crystals in the sulfate mother liquor. This initial *R*-factor was 43.2% for data in the resolution interval between 10 and 2.8 Å. After rigid body refinement, this *R*-factor decreased to 31.7% and, after some minor corrections and further cycles of refinement, to 23.0%. Table IV lists the final refinement statistics of the models of RfBP in the succinate and sulfate mother liquors. All the residues of the two models are in the energetically favored regions of the Ramachandran plot (Laskowski *et al.*, 1993)

Acknowledgements

I am grateful to Alessandro Coda, Menico Rizzi, Andrea Carfi, Sandro Ghisla, Andrea Mattevi, Elena Giulotto and Silvia Onesti for many fruitful discussions. This work was supported by grants from the Italian Ministry of the Universities, the Italian National Research Council (P.F. Structural Biology) and the Italian Space Agency (ASI). I thank the European Union for support of the work done at the EMBL outstation in Hamburg through the HCMF to Large Installations Project, contract No. CHGE-CT93-0040.

References

- Blankenhorn, G. (1978) Riboflavin binding in egg white flavoprotein: the role of tryptophan and tyrosine. *Eur. J. Biochem.*, **82**, 155–160.
- Becvar, J. and Palmer, G. (1982) The binding of flavin derivatives to the riboflavin-binding protein of egg white. *J. Biol. Chem.*, **257**, 5607–5617.
- Blundell, T.L. and Johnson, L.N. (1976) *Protein Crystallography*. Academic Press, London.
- Brünger, A.T. (1992a) XPLOR. Version 3.1. A System for X-ray Crystallography and NMR. Yale University Press, New Haven, CT.
- Brünger, A.T. (1992b) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
- Choi, J.-D. and McCormick, D.B. (1980) The interaction of flavins with egg white riboflavin-binding protein. *Arch. Biochem. Biophys.*, **204**, 41–51.
- Collaborative Computational Project Number 4 (1994) *Acta Crystallogr.*, **D50**, 760–767.
- Cowtan, K. (1994) DM: an automated procedure for phase improvement by density modification. In Bayley, S. and Wilson, K. (eds), *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*. pp. 34–38.
- Elwood, P.C. (1989) Molecular cloning and characterization of the human folate-binding protein cDNA from placenta and malignant tissue culture (KB) cells. *J. Biol. Chem.*, **264**, 14893–14901.
- Ghisla, S. and Massey, V. (1986) New flavins for old: artificial flavins as active site probes of flavoproteins. *Biochem. J.*, **239**, 1–12.
- Hamazume, Y., Mega, T. and Ikenaka, T. (1984) Characterization of hen egg white and yolk riboflavin-binding proteins and amino acid sequence of egg white riboflavin-binding protein. *J. Biochem.*, **95**, 1633–1644.
- Hamazume, Y., Mega, T. and Ikenaka, T. (1987) Position of disulfide bonds in riboflavin-binding protein of hen egg white. *J. Biochem.*, **101**, 217–223.
- Henderson, G.B. (1990) Folate-binding proteins. *Annu. Rev. Nutr.*, **10**, 319–335.
- Holt, C. and Sawyer, L. (1988) Primary and predicted secondary structures of the caseins in relation to their biological functions. *Protein Engng.*, **2**, 251–259.
- Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.
- Innis, W.S.A., McCormick, D.B. and Merrill, A.H., Jr (1985) Variations in riboflavin binding by human plasma: identification of immunoglobulins as the major proteins responsible. *Biochem. Med.*, **34**, 151–165.
- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
- Jusko, W.J. and Levy, G. (1975) Absorption, protein binding and elimination of riboflavin. In Rivlin, R.S. (ed.), *Riboflavin*. Plenum Press, New York, pp. 99–152.
- Kozik, A. (1982a) Carbodiimide modification of carboxyl groups in egg white riboflavin-binding protein. *Biochim. Biophys. Acta*, **704**, 542–545.
- Kozik, A. (1982b) Disulfide bonds in egg-white riboflavin-binding protein. *Eur. J. Biochem.*, **121**, 395–400.
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Krishnamurthy, K., Suroli, N. and Adiga, P.R. (1984) Mechanism of fetal wastage following immunoneutralization of riboflavin carrier protein in the pregnant rat: disturbances in flavin coenzyme levels. *FEBS Lett.*, **178**, 87–91.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
- Lubas, B., Soltysik, M., Steczko, J. and Ostrowski, W. (1977) Proton NMR study of the interaction of riboflavin with the egg-yolk apoprotein. *FEBS Lett.*, **79**, 179–182.
- Mac Lachlan, I., Nimpf, J. and Schneider, W.J. (1994) Avian riboflavin binding protein binds to lipoprotein receptors in association with vitellogenin. *J. Biol. Chem.*, **269**, 24127–24132.
- Mandel, S. and Ducay, E.D. (1962) Site of egg white protein formation. *J. Biol. Chem.*, **237**, 3169–3199.
- Mano, N., Oda, Y., Asakawa, N., Yoshida, Y. and Sato, T. (1992) Development of a flavoprotein column for chiral separation by high-performance liquid chromatography. *J. Chromatogr.*, **623**, 221–228.
- Massolini, G., De Lorenzi, E., Ponci, M.C., Gandini, C., Caccialanza, G. and Monaco, H.L. (1995) Egg yolk riboflavin binding protein as a new chiral stationary phase in high-performance liquid chromatography. *J. Chromatogr.*, **A704**, 55–65.
- Matsui, K., Sugimoto, K. and Kasai, S. (1982) Thermodynamics of association of egg yolk riboflavin binding protein with 8-substituted riboflavins. Comparison with the egg white protein. *J. Biochem.*, **91**, 1357–1362.
- McClelland, D.A., McLaughlin, S.H., Freedman, R.B. and Price, N.C. (1995) The refolding of hen egg white riboflavin-binding protein: effect of protein disulphide isomerase on the reoxidation of the reduced protein. *Biochem. J.*, **311**, 133–137.
- McCormick, D.B., Innis, W.S.A., Merrill, A.H., Jr, Bowers-Komro, D., Oka, M. and Chastain, J.L. (1987) An update on flavin metabolism in rats and humans. In Edmondson, D.E. and McCormick, D.B. (eds), *Flavins and Flavoproteins*. de Gruyter, Berlin, pp. 459–471.
- Merrill, A.H., Jr and McCormick, D.B. (1978) Flavin affinity chromatography: general methods for purification of proteins that bind riboflavin. *Anal. Biochem.*, **89**, 87–102.
- Merrill, A.H., Jr, Froelich, J.A. and McCormick, D.B. (1979) Purification of riboflavin-binding proteins from bovine plasma and discovery of a pregnancy-specific riboflavin-binding protein. *J. Biol. Chem.*, **254**, 9362–9364.
- Mifflin, T.E. and Langerman, N. (1983) Calorimetric studies of flavin binding protein: flavin analog binding. *Arch. Biochem. Biophys.*, **224**, 319–325.
- Miller, M.S., Bruch, R.C. and White, H.B., III (1982a) Carbohydrate compositional effects on tissue distribution of chicken riboflavin-binding protein. *Biochim. Biophys. Acta*, **715**, 126–136.
- Miller, M.S., Benore-Parsons, M. and White, H.B., III (1982b) Dephosphorylation of chicken riboflavin-binding protein and phosphovitin decreases their uptake by oocytes. *J. Biol. Chem.*, **257**, 6818–6824.
- Müller, F. and van Berkel, W.J.H. (1991) Methods used to reversibly resolve flavoproteins into the constituents apoflavoprotein and prosthetic group. In Müller, F. (ed.), *Chemistry and Biochemistry of Flavoenzymes*. CRC Press, Boca Raton, FL, Vol. I, pp. 261–274.
- Muniyappa, K. and Adiga, P.R. (1980) Occurrence and functional importance of a riboflavin-carrier protein in the pregnant rat. *FEBS Lett.*, **110**, 209–212.
- Murthy, C.V.R. and Adiga, P.R. (1982a) Pregnancy suppression by active immunization against gestation-specific riboflavin carrier protein. *Science*, **216**, 191–193.
- Murthy, C.V.R. and Adiga, P.R. (1982b) Isolation and characterization of a riboflavin-carrier protein from human pregnancy serum. *Biochem. Int.*, **5**, 289–296.
- Natraj, U., Kumar, A.R. and Kadam, P. (1987) Termination of pregnancy in mice with antiserum to chicken riboflavin-carrier protein. *Biol. Reprod.*, **36**, 677–685.
- Nishina, Y., Shiga, K., Horiike, K., Tojo, H., Kasai, S., Matsui, K., Watari, H. and Yamano, T. (1980) Resonance Raman spectra of semiquinone

- forms of flavins bound to riboflavin-binding protein. *J. Biochem.*, **88**, 411–416.
- Norioka,N., Okada,T., Hamazume,Y., Mega,T. and Ikenaka,T. (1985) Comparison of the amino acid sequences of hen plasma, yolk and white riboflavin-binding proteins. *J. Biochem.*, **97**, 19–28.
- Ostrowski,W., Skarzynski,B. and Zak,Z. (1962) Isolation and properties of flavoprotein from the egg yolk. *Biochim. Biophys. Acta*, **59**, 515–519.
- Otwinowski,Z. (1991) In Wolf,W., Evans,P.R. and Leslie,A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering: Proceedings of the CCP4 Study Weekend 25–26 January 1991*. SERC Daresbury Laboratory, Warrington, pp. 80–86.
- Otwinowski,Z. (1993) In Sawyer,L. and Bayley,S. (eds), *Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, pp. 56–62.
- Rao,S.T., Shaffie,F., Yu,C., Satyshur,K.A., Stockman,B.J., Markley,J.L. and Sundaralingam,M. (1992) Structure of the oxidized long-chain flavodoxin from *Anabaena* 7120 at 2 Å resolution. *Protein Sci.*, **1**, 1413–1427.
- Rhodes,M.B., Bennett,N. and Fenney,R.E. (1959) The flavoprotein–apoprotein system of egg white. *J. Biol. Chem.*, **234**, 2054–2060.
- Rohrer,J.S. and White,H.B.,III (1992) Separation and characterization of the two Asn-linked glycosylation sites of chicken serum riboflavin-binding protein. *Biochem. J.*, **285**, 275–280.
- Sheldrick,G.M. (1991) In Wolf,W., Evans,P.R. and Leslie,A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering: Proceedings of the CCP4 Study Weekend 25–26 January 1991*. SERC Daresbury Laboratory, Warrington, pp. 23–28.
- Svendsen,I., Hansen,S.I., Holm,J. and Lyngbye,J. (1984) The complete amino acid sequence of the folate-binding protein from cow's milk. *Carlsberg. Res. Commun.*, **49**, 123–131.
- Tarutani,M., Norioka,N., Mega,T., Hase,S. and Ikenaka,T. (1993) Structures of sugar chains of hen egg yolk riboflavin-binding protein. *J. Biochem.*, **113**, 677–682.
- Tronrud,D.E., Ten Eyck,L.F. and Matthews,B.W. (1987) An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr.*, **A43**, 489–501.
- Visweswariah,S.S. and Adiga,P.R. (1987) Purification of a circulatory riboflavin carrier protein from pregnant bonnet monkey (*M.radiata*): comparison with chicken egg vitamin carrier. *Biochim. Biophys. Acta*, **915**, 141–148.
- Walsh,C., Fisher,J., Spencer,R., Graham,D.W., Ashton,W.T., Brown,J.E., Brown,R.D. and Rogers,E.F. (1978) Chemical and enzymatic properties of riboflavin analogues. *Biochemistry*, **17**, 1942–1951.
- Watenpaugh,K.D., Sieker,L.C. and Jensen,L.H. (1973) The binding of riboflavin-5'-phosphate in a flavoprotein: flavodoxin at 2.0 Å resolution. *Proc. Natl Acad. Sci. USA*, **70**, 3857–3860.
- Wessiak,A., Schopfer,L.M., Yuan,L.-C., Bruice,T.C. and Massey,V. (1984) Use of riboflavin-binding protein to investigate steric and electronic relationships in flavin analogs and models. *Proc. Natl Acad. Sci. USA*, **81**, 4246–4249.
- White,H.B.,III and Merrill,A.H.,Jr (1988) Riboflavin-binding proteins. *Annu. Rev. Nutr.*, **8**, 279–299.
- Zanette,D., Monaco,H.L., Zanotti,G. and Spadon,P. (1984) Crystallization of hen eggwhite riboflavin-binding protein. *J. Mol. Biol.*, **180**, 1185–1187.
- Zheng,D.B., Lim,H.M., Pene,J.J. and White,H.B.,III (1988) Chicken riboflavin binding protein: cDNA sequence and homology with milk folate binding protein. *J. Biol. Chem.*, **263**, 11126–11129.

Received on October 29, 1996; revised on December 13, 1996