Primary structure of a Photoactive Yellow Protein from the phototrophic bacterium Ectothiorhodospira halophila with evidence for the mass and the binding-site of the chromophore

ARTICLE *in* PROTEIN SCIENCE · JULY 1993

Impact Factor: 2.85 · DOI: 10.1002/pro.5560020706

Protein Science (1993), *2*, 1114–1125. Cambridge University Press. Printed in the USA.

Primary structure of a photoactive yellow protein from the phototrophic bacterium *Ectothiorhodospira halophila*, with evidence for the mass and the binding site of the chromophore

JOZEF J. VAN BEEUMEN,[1] BART V. DEVREESE,[1] STEFAAN M. VAN BUN,[1,5] WOUTER D. HOFF,[2] KLAAS J. HELLINGWERF,[2] TERRANCE E. MEYER,[3] DUNCAN E. McREE,[4] AND MICHAEL A. CUSANOVICH[3]

[1] Laboratory of Microbiology, University of Ghent, 9000 Ghent, Belgium
[2] Department of Microbiology and Biotechnology Center, Universiteit van Amsterdam, The Netherlands
[3] Department of Biochemistry, University of Arizona, Tucson, Arizona 85721
[4] Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037

(RECEIVED February 16, 1993; REVISED MANUSCRIPT RECEIVED April 9, 1993)

Abstract

The complete amino acid sequence of the 125-residue photoactive yellow protein (PYP) from *Ectothiorhodospira halophila* has been determined to be

MEHVAFGSEDIENTLAKMDDGQLDGLAFGAIQLDGDGNILQYNAAEGDITGRDPKEVIGKNFFKDVAP
CTDSPEFYGKFKEGVASGNLNTMFEYTFDYQMTPTKVKVHMKKALSGDSYWVFVKRV.

This is the first sequence to be reported for this class of proteins. There is no obvious sequence homology to any other protein, although the crystal structure, known at 2.4 Å resolution (McRee, D.E., et al., 1989, *Proc. Natl. Acad. Sci. USA 86*, 6533–6537), indicates a relationship to the similarly sized fatty acid binding protein (FABP), a representative of a family of eukaryotic proteins that bind hydrophobic molecules. The amino acid sequence exhibits no greater similarity between PYP and FABP than for proteins chosen at random (8%).

The photoactive yellow protein contains an unidentified chromophore that is bleached by light but recovers within a second. Here we demonstrate that the chromophore is bound covalently to Cys 69 instead of Lys 111 as deduced from the crystal structure analysis. The partially exposed side chains of Tyr 76, 94, and 118, plus Trp 119 appear to be arranged in a cluster and probably become more exposed due to a conformational change of the protein resulting from light-induced chromophore bleaching. The charged residues are not uniformly distributed on the protein surface but are arranged in positive and negative clusters on opposite sides of the protein.

The exact chemical nature of the chromophore remains undetermined, but we here propose a possible structure based on precise mass analysis of a chromophore-binding peptide by electrospray ionization mass spectrometry and on the fact that the chromophore can be cleaved off the apoprotein upon reduction with a thiol reagent. The molecular mass of the chromophore, including an SH group, is 147.6 Da (±0.5 Da); the cysteine residue to which it is bound is at sequence position 69.

Keywords: amino acid sequence; chromophore structure; electrospray mass analysis; photoactive yellow protein

Photoactive yellow protein (PYP) is representative of a new class of proteins that combine the photoactive characteristics of bacteriorhodopsin and the structural properties of the proteins that bind hydrophobic molecules. PYP is a small soluble protein of 14 kDa that to date is found only in the extremely halophilic purple phototrophic bacterium *Ectothiorhodospira halophila* (Meyer, 1985), in the moderately halophilic purple bacterium *Rhodospirillum salexigens* (Meyer et al., 1990), and in *Chromatium salexigens* (Meyer, unpubl.). It contains an unidentified chromophore that undergoes photochemical transformations remarkably similar to those of the membrane-bound proteins bacteriorhodopsin, halorhodopsin, and the two

Reprint requests to: Jozef Van Beeumen, Vakgroep Biochemie, Fysiologie en Microbiologie, Laboratorium voor Microbiologie, Universiteit Gent, Ledeganckstraat, 35, 9000 Gent, Belgium.
[5] Present address: Applied Biosystems B.V., 3606 AS Maarssen, The Netherlands.

sensory rhodopsins from the unrelated halophilic bacterium *Halobacterium halobium* (Meyer et al., 1987, 1989, 1991; Miller et al., 1992). Crystal structure analysis (Kinemage 1; McRee et al., 1989) indicates strong resemblance to a fatty acid binding protein (Scapin et al., 1992) and to P2 myelin protein (Jones et al., 1988). These are members of a group of homologous proteins, including cellular retinol binding protein and cellular retinoic acid binding protein, all of which share the capacity to bind hydrophobic molecules. This group of proteins shows more distant relationship to another group, which binds hydrophobic molecules, including serum retinol binding protein (Newcomer et al., 1984), insecticyanin (Holden et al., 1987), and β-lactoglobulin (Monaco et al., 1987), by some called the lipocalin protein family (Pervaiz & Brew, 1987; Åkerström & Lögdberg, 1990). We have now determined the complete amino acid sequence of PYP for comparison with the hydrophobic molecule binding proteins. Using accurate mass analysis methods such as plasma desorption (PDMS) and electrospray ionization mass spectrometry (ESMS), we also proved that the chromophore is covalently bound to a cysteine residue.

Results

Amino acid sequence

The complete amino acid sequence of PYP from *E. halophila* is given in Figure 1. It was obtained by automated sequence analysis of peptides from five different digests of the carboxymethylated protein in addition to N-terminal sequence analysis of the native and the carboxymethylated protein. All of the peptides from the *Staphylococcus aureus* protease digest (Fig. 2), from the partial acid hydrolysis (Fig. 3), and some major peptides of the other digests (separations given in the Supplementary material on the Diskette Appendix) were submitted to sequence analysis. Apart from the N-terminal sequencer run, which positioned the peptides S4 and S10, peptide H10 was very useful in providing an overlap between the *S. aureus* protease peptides S8A, S2, and S5. Mild acid cleavage of the Asp 53–Pro 54 bond, expected to be successful from the preliminary detection of this acid-labile sequence in the N-terminal protein sequencer run, indeed revealed a major peptide, D/P2, which confirmed the overlap between S8A and S2. It was the chymotryptic peptide C4 that provided the proof that peptides S10 and S8A were connected in the sequence. Finally, peptide Kc4 from the Lys-C proteinase digest showed the linkage of peptides S5 and S9. The chemical evidence showing that the C-terminal sequence of the yellow protein ends with -Lys-Arg-Val comes partly from the release of amino acids upon incubation of peptide H7 with carboxypeptidase P although the amounts released after 1 h could suggest the presence of four instead of three valines in this peptide. In the former case, one would expect the C-terminal sequence to be -Arg-Val-Val, with the last valine being difficult to detect as a phenylthiohydantoin (PTH) derivative upon Edman degradation. The final evidence for the C-terminal sequence ending with -Lys-Arg-Val was obtained by mass analysis of the peptides H7 and S9 (see further).

The amino acid composition, based on the sequence proposed in Figure 1, is in good agreement with the one experimentally determined from acid hydrolysis (Table 1). The slight aberrations from the calculated values are mainly due to minor contamination of the starting preparation of protein as revealed by the N-terminal sequence analysis. The yellow protein contains one cysteine residue, at position 69, and one tryptophan, six residues before the C-terminal Val 125. The cysteine residue was positively identified as the carboxymethyl derivative.

Considering the specificity of the cleavage methods used, it should be mentioned that several Glu–X bonds have not been fully cleaved by the *S. aureus* protease, e.g., Glu–Asp in peptide S3, Glu–His in S4, and Glu–Gly in S10. The N-terminal sequence of the latter peptide was also found upon sequencing of one peptide eluted earlier (S7A) and two peptides eluted later than S10 (S11 and S12; Fig. 2). It is possible that the latter peptides terminate at Glu 74 originating from an uncleaved Glu 56–Val 57 bond and/or are due to different degrees of modification of Cys 69. We also noted the unusual cleavage of a Gly–

Table 1. *Amino acid composition of the Ectothiorhodospira halophila yellow protein*[a]

	Native protein	Oxidized protein		Sequence
		24 h	96 h	
Asp	17.49	17.41	17.50	12
Asn	–	–	–	6
Thr	6.44	6.07	5.47	7
Ser	4.38	4.72	3.91	5
Glu	11.61	12.21	12.39	8
Gln	–	–	–	4
Pro	3.95	3.73	3.98	4
Gly	12.44	13.08	13.68	13
Ala	8.99	8.95	8.66	9
Val	8.59	8.07	8.85	9
Met[b]	2.78	3.90	4.20	5
Ile	4.65	4.70	4.55	5
Leu	6.58	7.11	6.81	7
Tyr	4.05	–	–	5
Phe	8.72	8.28	7.50	9
His	1.96	1.85	1.87	2
Lys	10.98	10.93	10.77	11
Trp	–	–	–	1
Arg	2.05	2.00	2.33	2
Cys[c]	–	1.09	0.93	1

[a] Amounts used for the analyses were respectively 1.10, 0.41, and 0.43 nmol.
[b] As methionine sulfone.
[c] As cysteic acid.

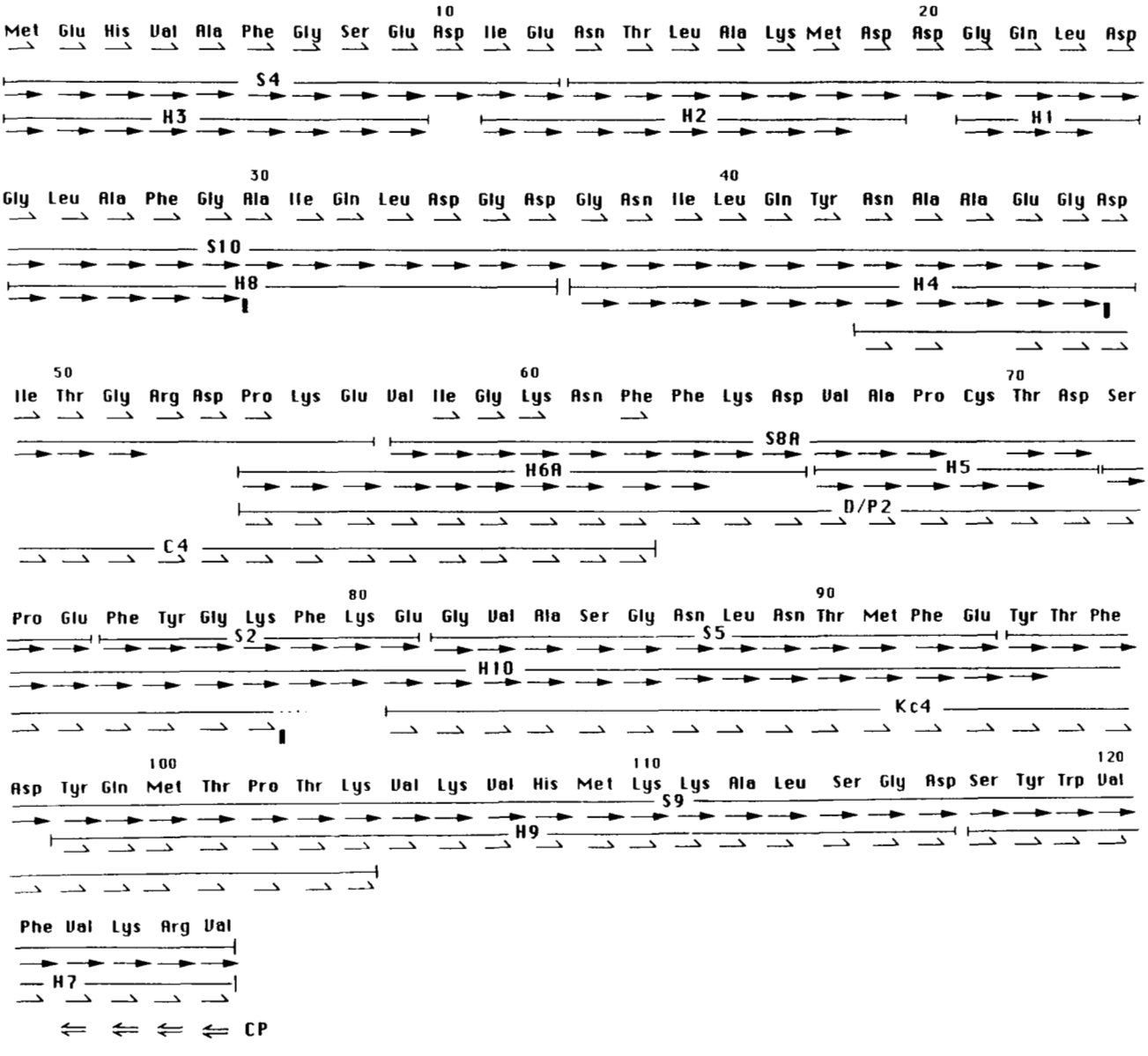

Fig. 1. Amino acid sequence of the yellow protein from *E. halophila*. The peptides, which are shown, were obtained from carboxymethylated protein digested with *S. aureus* protease (S), Lys-C endoproteinase (Kc), and chymotrypsin (C), as well as from partial acid hydrolysis (H) and specific Asp/Pro cleavage (D/P). Amino acids identified by off-line phenylthiohydantoin (PTH) analysis using an Applied Biosystems 470A gas-phase sequencer are indicated by single arrows. Residues that were detected on line using an Applied Biosystems 477A pulsed-liquid sequencer are shown as double arrows. The symbol ▮ indicates where the sequence runs were deliberately stopped.

Ala bond at position 29 of the protein sequence, resulting in peptide S6 and most likely S7A. Peptide S7C originated from a cleavage at the C-terminal side of Asp 116. Several of the small peaks eluted between S1 and S2 (Fig. 2) originated from minor cleavages of Asp–X bonds such as at the positions 65, 24, and 34 (results not shown). In the series of peptides obtained under conditions where Asp–Pro bonds were expected to be cleaved, we also noted that the minor fractions, eluted earlier than D/P1, originated from partial cleavage of Asp–X bonds with X = Gly (e.g., Gly 25 and Gly 37) or X = Ser (e.g., Ser 72 and Ser 117); also a limited cleavage of the Lys 34–Gly 35 bond has been observed.

Absorbance changes of the yellow protein induced by proteinase K

Upon incubations of the native yellow protein with proteinase K, the major absorbance with a maximum at 445 nm decreased gradually as a function of time, whereas absorbance at around 335 nm was concomitantly rising (Fig. 4A). When the A445 had nearly completely disap-

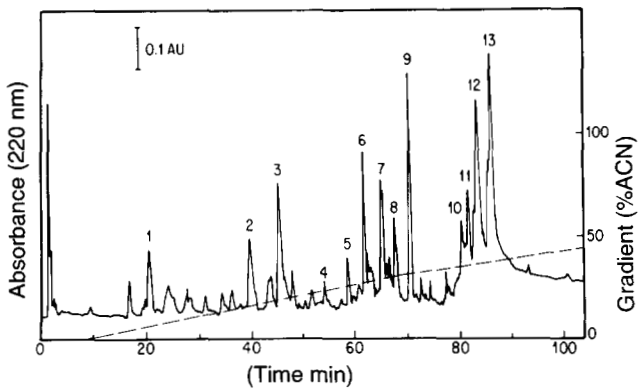

Fig. 2. Reversed-phase HPLC separation of the peptides from a digest of the carboxymethylated yellow protein with *S. aureus* V8 protease. Conditions are described under Methods (Diskette Appendix). The course of the gradient is shown by the broken line. AU, absorbance units.

peared after 250 min of incubation, addition of NaOH brought down the absorbance at 335 nm (Fig. 4B). Simultaneously, a shoulder at 307 nm became apparent, although the absorbance at this wavelength had already increased a small amount after the proteinase alone had been added. A summary of the changes at the three wavelengths mentioned as well as at the wavelength of 278 nm is given in Figure 5.

Mass analyses

After the complete sequence analysis was finished, we were able to determine the precise molecular weight of a newly prepared sample of the pure protein by ESMS, a method known to have a mass accuracy of 0.01% (Fenn et al., 1990). We obtained a value of 14,021.4 (Fig. 6), which is 146.9 higher than the molecular weight deduced

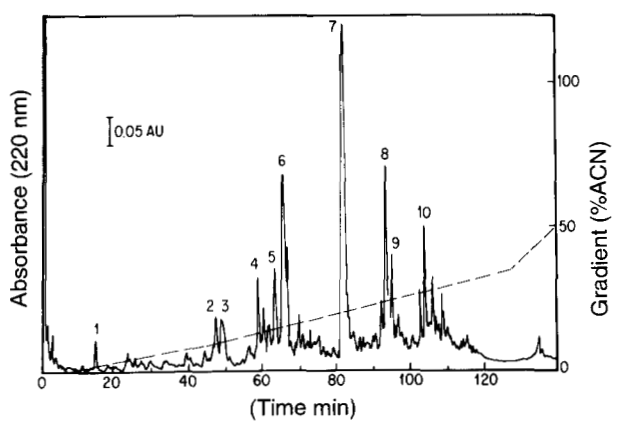

Fig. 3. HPLC separation of peptides obtained after partial acid hydrolysis of the carboxymethylated yellow protein. Conditions are as described under Methods (Diskette Appendix).

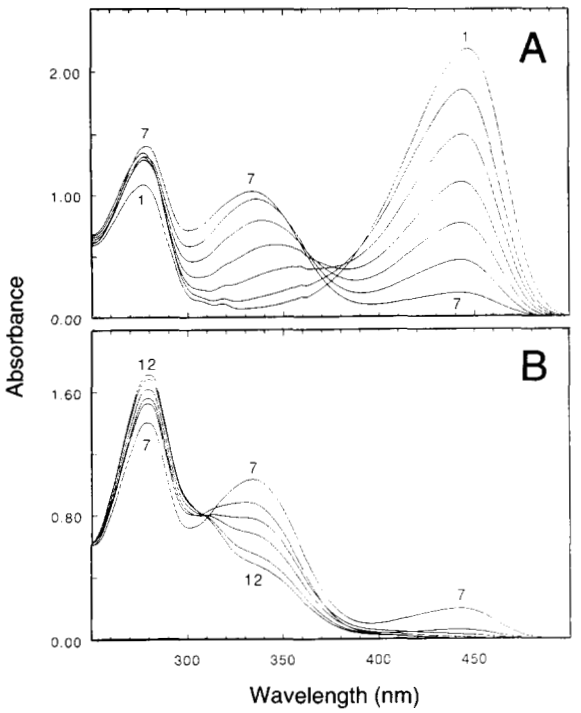

Fig. 4. Changes in the UV/VIS spectrum of the yellow protein upon incubation with proteinase K (**A**), and with 0.1 N NaOH added after 227 min of incubation with proteinase K (**B**). Spectra were recorded at the following time intervals: 1, 0 min; 2, 8 min; 3, 13 min; 4, 23 min; 5, 39 min; 6, 66 min; 7, 122 min. The dip in the spectra at 360 nm is due to the change of lamp source in the spectrophotometer. Spectrum 7 in B is the same as spectrum 7 in A. The time intervals were: 8, 230 min; 9, 235 min; 10, 250 min; 11, 300 min; and 12, 389 min.

from the sequence assuming the C-terminal sequence to be Lys-Arg-Val and not Lys-Arg-Val-Val. The uncertainty about the latter part of the protein was taken away by measuring the precise mass of peptide H7, the C-terminal peptide of the protein obtained after the partial acid hydrolysis cleavage. For this measurement we used the technique of PDMS, known to have a mass accuracy of 0.1% (Cotter, 1988; Loo et al., 1990). We found a mass of 1,185.1 Da, which is only 1.7 Da different from the calculated value of 1,183.4 Da, assuming that the peptide contains only three valine residues. Another confirmation of the presence of only 125 amino acids in the polypeptide chain of the yellow protein came from the mass of peptide S9 (Figs. 1, 2), which was measured by PDMS to be 3,857.9 Da, only 2.4 Da higher than the calculated value of 3,855.5 Da (Table 2).

On the basis of these facts and of the solid evidence for the identification of the amino acids in the complete sequence of the protein, it became clear that the mass difference of 146.9 Da between the calculated and the measured mass of the native protein was due to the presence of the unidentified chromophore, which had to be covalently linked to the polypeptide chain. In order to determine which of the peptides might be involved in the

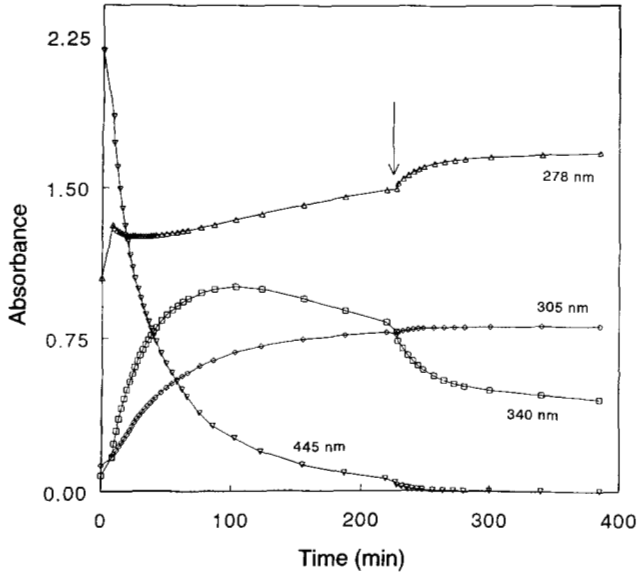

Fig. 5. Absorbance changes at four different wavelengths upon incubation of the yellow protein with proteinase K (0–227 min) and after further addition (arrow) of 0.1 N NaOH (227–400 min).

Table 2. *Calculated and experimentally determined masses of Lys-C endoprotease peptides of the native yellow protein and of some Staphylococcus aureus protease and acid hydrolysis peptides of the carboxymethylated yellow protein*

Peptide name	Position	Theoretical mass[a] (Da)	Experimental mass[b] (Da)	
			PDMS	ESMS
Kc1	Met 1–Lys 17	1,891.1		1,890.9 ± 0.8
Kc4a	Met 18–Lys 55	3,965.3		3,965.0 ± 0.8
Kca	Glu 56–Lys 60	544.6	545.2	
Kcb	Asn 61–Lys 64	554.6	555.1	
Kc2	Asp 65–Lys 78	1,528.5[c]	1,675.5	
Kc4b	Glu 81–Lys 104	2,745.0		2,744.2 ± 0.6
S9	Tyr 94–Val 125	3,855.5	3,857.9	
Kc3	Ala 112–Lys 123	1,371.6		1,371.0
Kc5	Met 18–Val 125	12,001.4[c]		12,146.9 ± 4.6
H7	Ser 117–Val 125	1,183.4	1,185.1	

[a] Calculated from the amino acid average masses.
[b] PDMS, plasma desorption mass spectrometry; ESMS, electrospray ionization mass spectrometry.
[c] Mass without chromophore.

binding of the chromophore, we carried out a cleavage of the native protein with Lys-C endoproteinase, separated the peptides by HPLC (Fig. 7), and determined the mass of each of the peptides by either PDMS or ESMS. Table 2 reports the masses of all these peptides. They fit the calculated values within the accuracy of the method used except, firstly for peptide Kc2, which differed from the calculated mass by 147.0 Da (Fig. 8). This difference, together with the fact that peptide Kc2 showed UV absorbance at 307 nm, suggested that the chromophore was covalently bound to an amino acid in the region between Asp 65 and Lys 78 of the apoprotein. The other main peak showing 307 nm absorbance (Kc5) appeared to contain the chromophore as well because its molecular mass was 145.5 Da higher than the calculated mass of the sequence Met 18–Val 125. This peptide must have arisen from incomplete cleavage of the Lys-X peptide bonds at positions 55, 60, 74, and 104. Sequence analysis did show the N-terminal sequence to be Met-Asp-Asp-Gly. The smaller peak Kc2′ appeared to have the same molecular

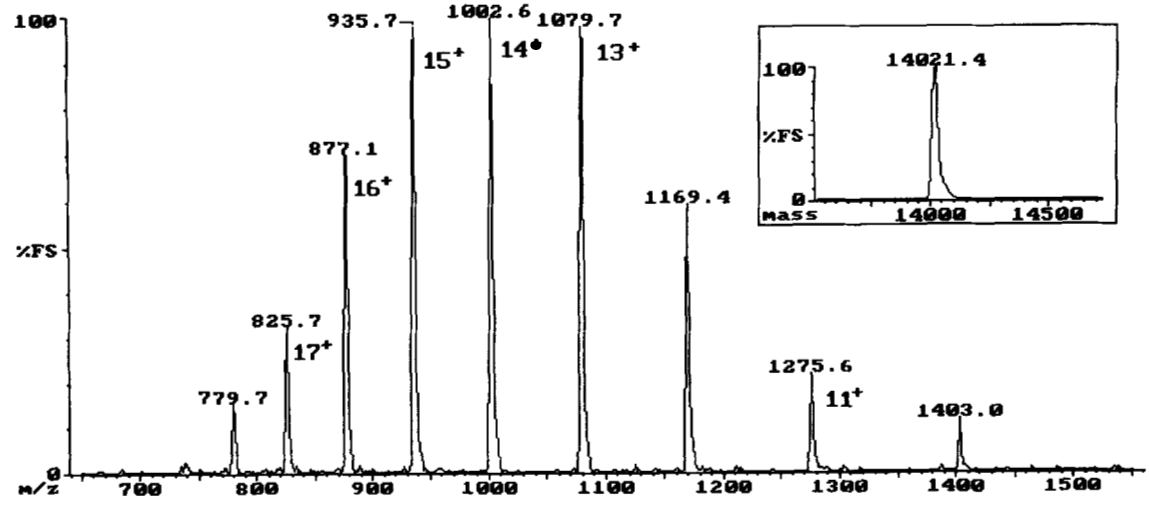

Fig. 6. Electrospray mass spectrum of the native yellow protein of *E. halophila*. The insert shows the mass peak of the protein calculated by the "Transform" program. The values preceding the + signs refer to the different number of protons for each of the *m/z* values detected during a single scan.

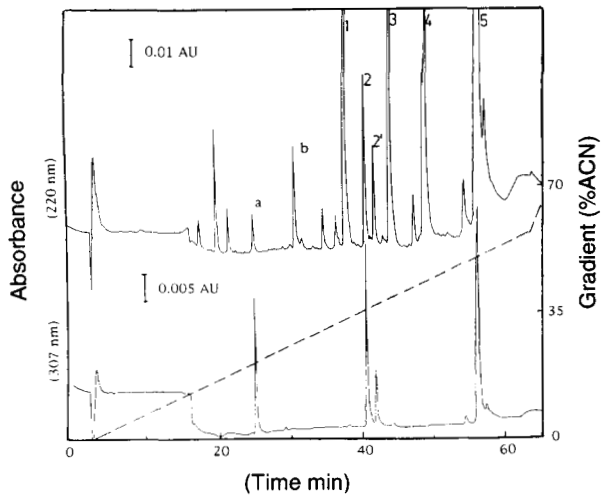

Fig. 7. HPLC separation of the peptides obtained by a digest of the native yellow protein with Lys-C endoproteinase. The chromatogram is comparable to the one obtained for the carboxymethylated protein although the separation was carried out on a different column (see Methods, Diskette Appendix). Peptides are numbered as in Figure S3 (Supplementary material, Diskette Appendix). Peaks a and b are both pentapeptides covering the sequence regions Ala 56–Lys 60 and Ala 61–Lys 64.

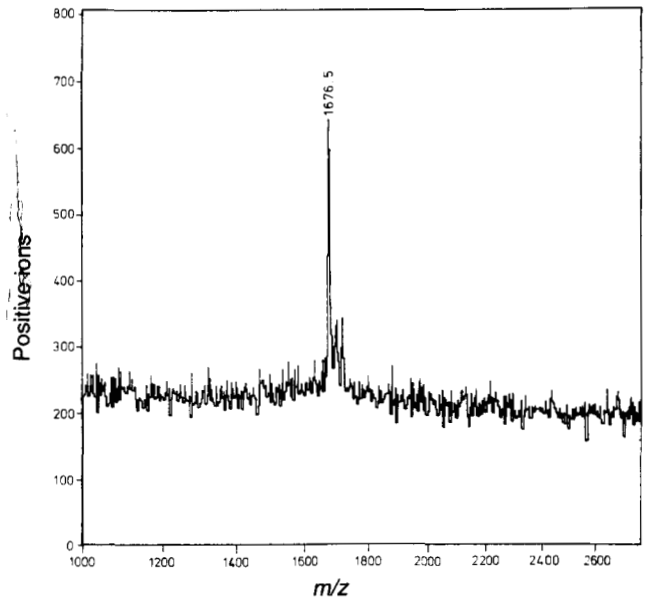

Fig. 8. Plasma desorption mass spectrum of peptide Kc2 obtained by Lys-C endoproteinase cleavage of the native yellow protein. The spectrum was acquired for 10^6 fission events, corresponding to 12 min of data collection.

mass as peptide Kc2. It is not clear why this peptide eluted at a different solvent concentration from that for peptide Kc2.

Proteinase K digest

In order to narrow down the size of the peptide linking the presumed chromophore, we repeated the digest of the native protein with the nonspecific enzyme proteinase K. The resulting peptides were first subjected to gel filtration on a Sephadex G25 column. Both fraction 9 and fraction 10, showing the highest absorbance at 307 nm and at 333 nm, respectively (Fig. 9), were collected and further purified by reverse-phase HPLC. They essentially gave the same separation pattern (Fig. 10), but peaks 28 and 30 were relatively higher in fraction 10. Their absorption spectra as well as that of P18, the major peak at 220 nm, were recorded on line with the separation tracked by a photodiode array detector (Fig. 11). Both P28 and P30 have a maximum at 332 nm. For P20 it is at 307 nm, and for P18 it is at 278 nm. Fractions 20, 28, and 30 were submitted to sequence, amino acid, and mass analysis. P20 did not give results with any of these methods, whereas for P28 and P30, using 1 nmol of material, the sequence Val-Ala-Pro was found with no information from the fourth Edman cycle onward. Their amino acid composition determined on ±400 pmol of material and after 24 h of acid hydrolysis at 110 °C gave the results Val 1.0, Ala 1.0, Pro 1.0 for P28, and Val 0.9, Ala 1.1, Pro 1.0 for P30. The PDMS spectrum, given in Figure 12, revealed two main peaks, one at 1,071.4 Da and the other at 536.2 Da. The latter value is, within the limits of accuracy, exactly the molecular mass of the peptide Val-Ala-Pro-Cys plus the chromophore of mass 146.9 Da, which is the same mass difference as found for the holoprotein and for peptides Kc2 and Kc5. The combination of results of these experiments on the 332-nm-absorbing peptides

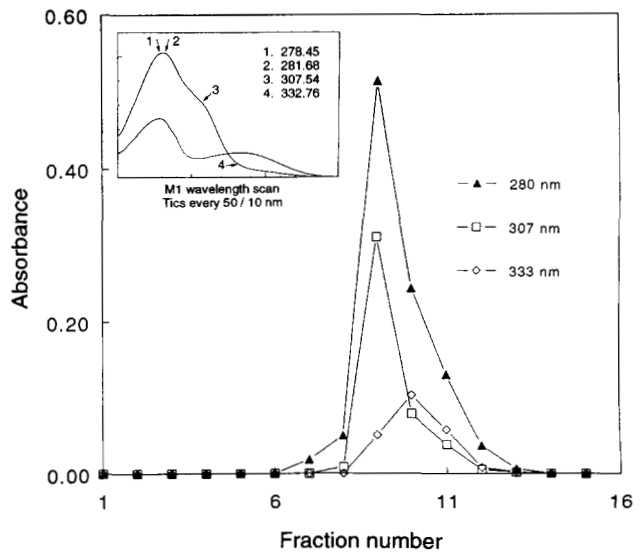

Fig. 9. Gel filtration of the proteinase K digest of the native yellow protein. The inset shows the absorbance spectra of fractions 9 and 10.

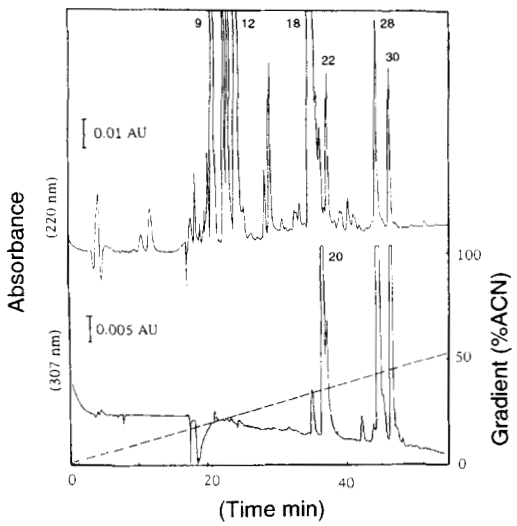

Fig. 10. HPLC separation of fraction 10 of the proteinase K digest of the native yellow protein. Only the fractions showing 307-nm absorbance were further investigated by mass and sequence analysis.

is the most direct proof that the chromophore of the yellow protein is covalently bound to Cys 69. The mass peak at 1,071.4 Da is smaller in intensity than the mass peak at 536.2 Da. Given the amino acid composition of the

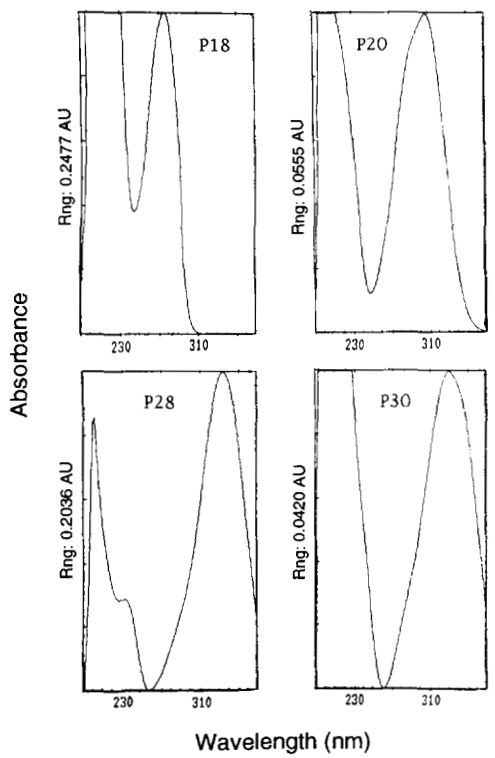

Fig. 11. Absorption spectra, registered on line by photodiode array detection, of fractions 18, 20, 28, and 30 obtained by proteinase K digestion of the yellow protein (Fig. 10).

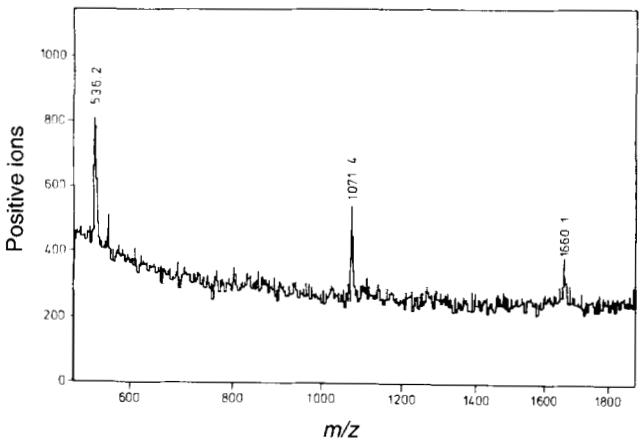

Fig. 12. Plasma desorption mass spectrum of peptide 28 from the peptide mixture "fraction 10" of the proteinase K digest.

peptide, we should conclude that this value may represent a dimeric form of the peptide Val-Ala-Pro-Cys (chromophore). This conclusion is also supported by the general observation for PDMS spectra of small peptides that the $M + 2H^+/2$ mass values are smaller than the MH^+ values (Cotter, 1988). We have no idea, however, how such a dimer would be formed.

Discussion

Sequence comparisons

There is no obvious sequence similarity between PYP and any other protein. However, the three-dimensional structure (Kinemage 1; McRee et al., 1989) of PYP suggests a structural relationship to proteins of the P2 superfamily of small lipid binding proteins such as the intestinal fatty acid binding protein (I-FABP) (Scapin et al., 1992) and P2 myelin protein (P2) (Jones et al., 1988). Alignment of PYP with rat and human I-FABP (Alpers et al., 1984; Sweetser et al., 1987) and cellular retinoic acid and retinol binding proteins (CRABP and CRBP) (Sundelin et al., 1985a,b), also members of this family, is shown in Figure 13. PYP is only 8–14% similar to the individual proteins in the P2 family, CRBP II, CRBP, and AP2 (Sundelin et al., 1985a; Hunt et al., 1986; Li et al., 1986), which is very close to what is expected for randomly scrambled sequences (Table 3). There is much greater similarity among the other proteins in this family, and there are residues that are conserved in all the proteins but PYP. An aromatic residue at position 8 (numbering as in Fig. 13) and the conserved residues Glu 15 and Asn 16 were the basis for the alignment chosen for the PYP. However, the other proteins have glycines at positions 46 and 67 that have been shown to be involved in hairpin turns in P2 and I-FABP. These glycines appear to be absent in PYP based on the alignment of Figure 13,

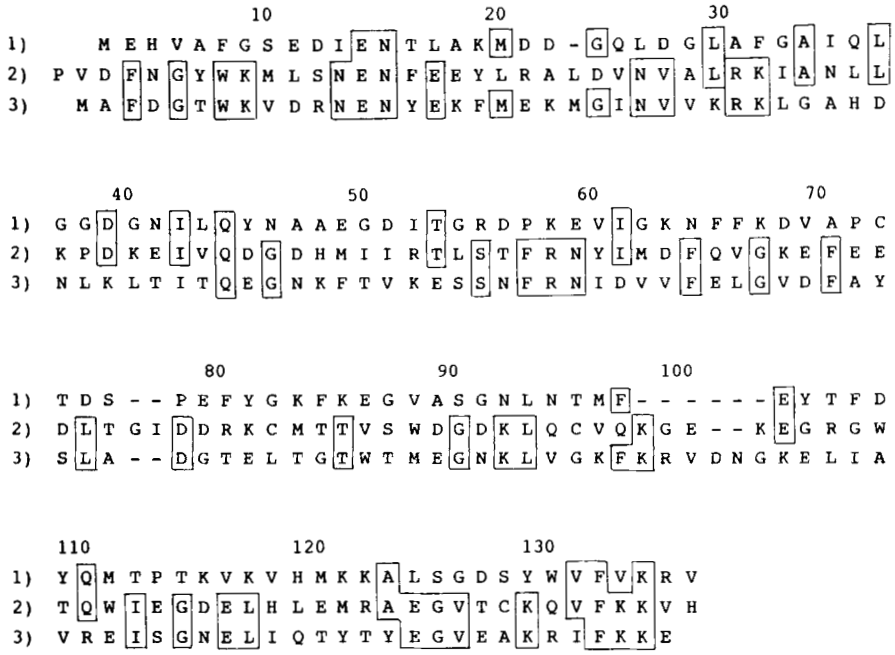

Fig. 13. Sequence alignment of *E. halophila* photoactive yellow protein (1) with rat cellular retinol binding protein (2), and rat intestinal fatty acid binding protein (Alpers et al., 1984) (3). Identical residues are boxed.

although the three-dimensional structure of PYP shows that there are turns at about the same places as in P2 and I-FABP (McRee et al., 1989). Until a higher-resolution refined structure for PYP becomes available for comparison with I-FABP, no precise alignments or additional conclusions about sequence homology can be drawn.

The photocycle of PYP is remarkably similar to those of the sensory rhodopsins and to a lesser degree to bacteriorhodopsin from *Halobacterium* (Meyer et al., 1987, 1989; Hoff et al., 1992; Miller et al., 1992). We therefore also considered the possibility that PYP might have some similarity to the structure of those proteins. However, the proteins from *Halobacterium* have molecular weights about 23,000–26,000, they are completely membrane bound, and they are primarily helical. In these properties, they are more like the visual pigments of animals, which have molecular weights about 40,000–45,000 and which are also membrane bound and primarily helical. The retinal is covalently bound to a lysine near the C-terminus in both bacteriorhodopsins and the visual pigments, and the seven membrane-spanning helices are distributed over the length of the sequences (Henderson et al., 1990). The secondary structure of PYP is all β with no helix (McRee et al., 1989), and there is no sequential concentration of hydrophobic residues as observed in the membrane-spanning segments of the rhodopsins. We conclude therefore that there is no structural relationship between PYP and the rhodopsins.

Table 3. *Matrix of amino acid sequence identities for pairwise comparison of the P2 superfamily of hydrophobic molecule binding proteins*[a]

	CRBP II	CRBP	CRABP	PYP	L-FABP	I-FABP	H-FABP	AP2
CRBP	56							
CRABP	37	42						
PYP	13	14	9					
L-FABP	27	23	27	10				
I-FABP	31	31	29	8	29			
H-FABP	40	33	43	11	25	31		
AP2	37	35	38	11	22	28	63	
P2	35	37	43	14	21	28	56	66

[a] Proteins are those mentioned in Figure 13 (CRBP and I-FABP) as well as rat cellular retinol binding protein II (CRBP II), bovine cellular retinoic acid binding protein (CRABP), human intestinal fatty acid binding protein (H-FABP), rat liver fatty acid binding protein (L-FABP), mouse murine adipocyte protein (AP2), and bovine P2 protein from peripheral nervous system myelin (P2).

Some features of the tertiary structure

The three-dimensional structure of PYP (McRee et al., 1989) shows that eight aromatic residues (Phe 75, Tyr 76, Phe 79, Phe 92, Tyr 94, Phe 96, Tyr 118, and Trp 119) are spatially oriented to form a hydrophobic patch on the surface, which is surrounded by both acidic and basic residues. Based on kinetics measurements (Meyer et al., 1990), it appears that the protein undergoes a conformational change upon absorption of light and that a hydrophobic site is exposed to solvent. Perhaps it is the above site that becomes even more exposed in the bleached protein than in the ground-state protein. This will have to await higher-resolution three-dimensional analysis of ground-state and bleached protein for clarification.

The charged residues in PYP are not uniformly distributed over the surface, but acidic and basic residues are localized on opposite sides of the protein (Fig. 14; Kinemage 2). Seven basic residues in the C-terminal quarter of the sequence (Lys 104, Lys 106, His 108, Lys 110, Lys 111, Lys 123, and Arg 124) define the weakly basic side of the protein, whereas the 20 acidic residues and the remaining 8 basic residues are on the opposite highly acidic face. This is very unusual for a soluble protein and is not found in fatty acid binding proteins. Instead, it is reminiscent of mitochondrial cytochrome *c* and of bacterial cytochromes c_2, which have the functionally important basic residues surrounding the exposed heme edge—where electron transfer takes place—and all of the acidic residues located on the back side (Meyer & Kamen, 1982). By analogy, the asymmetric distribution of surface charge suggests an important functional role for the clustered charges in PYP.

In summary, the similarity of the PYP sequence with the amino acid sequences of both the P2 and particularly the lipocalin superfamily of the lipid binding proteins is marginal or even nonexistent. This does not exclude the possibility, however, that the yellow protein may have evolved by divergent evolution (Stavenga et al., 1991). It certainly suggests that nature has evolved a common structural solution to the problem of forming an enclosed binding site for a small organic molecule. Secondly, although there appears to be similarity in the photoactivity of the rhodopsins and the yellow protein, there is no structural similarity either at the primary or at the tertiary structure level.

The mass of the chromophore

In the paper on the three-dimensional structure of PYP, it was suggested that Lys 111 may be involved in binding the chromophore responsible for the photoactivity (McRee et al., 1989). The fact that the molecular mass of the native protein is 146.9 Da higher than the one calculated from the determined sequence and that, on the other hand, one particular peptide from the Lys-C digest of the native protein (Kc2) likewise had a molecular weight dif-

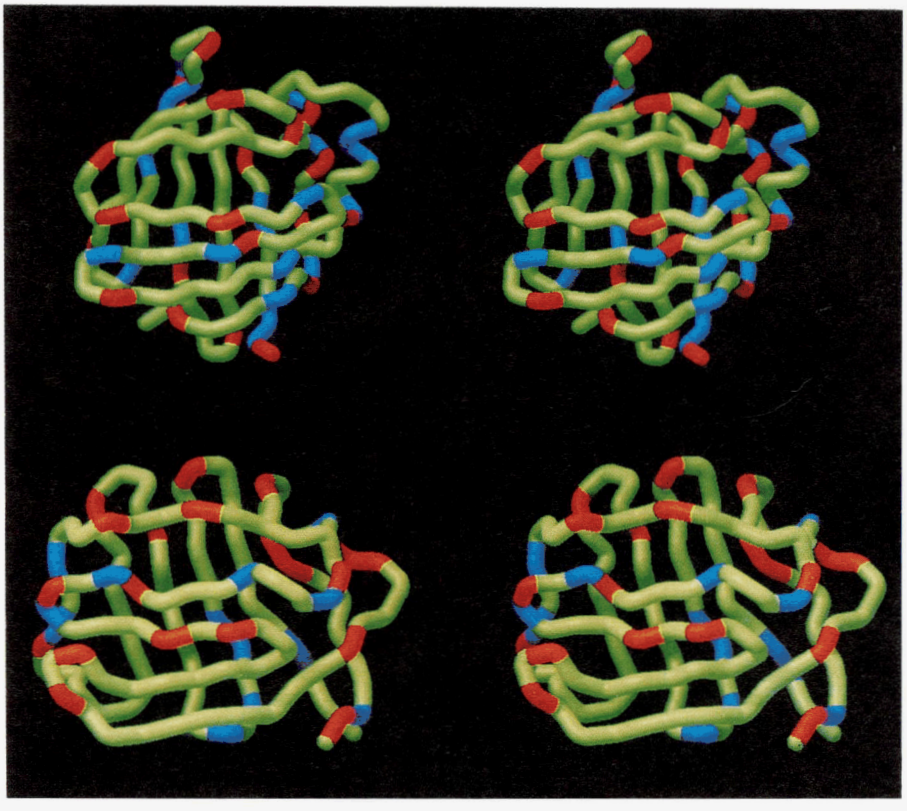

Fig. 14. Stereo views of fatty acid binding protein (**top**) and photoactive yellow protein (**bottom**). The paths of the main chains are shown as cylinders colored by residue charge: red, negatively charged residues; blue, positively charged residues; green, noncharged residues. Both proteins form a clamlike shell of perpendicular β-strands. A large cluster of negatively charged residues is along the top of the molecule in this view, and a cluster of positive residues is found in the lower back portion of the molecule.

ference with the calculated value of 147.0 Da indicated that the chromophoric group is covalently bound to this peptide. From the results of sequence, amino acid, and mass analyses of the small 332-nm-absorbing peptides P28 and P30 that were obtained from the proteinase K digest of the native yellow protein, we can unambiguously conclude that the chromophore is covalently bound to Cys 69 and that its molecular mass is 146.9 Da. This conclusion is further supported by the molecular mass of peptide Kc5 (12,146.9 Da), which originated from a single cleavage of the protein chain at Lys 17, and of the whole protein, which give an average mass value of the chromophore of 146.0 ± 0.5 Da.

How the chromophore is bound to the apoprotein

Addressing the question in which way the chromophore is linked to Cys 69, it should be remembered that Cys 69 has been identified in the appropriate peptides of the different digests as the carboxymethyl derivative after prior reduction and carboxymethylation of the native protein. This implies that the chromophore is linked to the cysteine by a chemical bond that is cleaved by dithiothreitol. In order to confirm the latter conclusion, we also determined the mass of the yellow protein by ESMS after reduction with this thiol reagent. Figure 15 shows that the mass of the "reduced" protein is indeed 147.2 Da lower than that of the native form. It should be mentioned that thioesters as well as disulfides should be cleaved by thiols, although the former should also be labile to acid hydrolysis. A thioether would not be cleaved by either treatment. The chromophore was not released from the protein nor from peptides P28 or P30 by treatment with 25% trifluoroacetic acid for 1 h at 25 °C, which suggests that it is bound via a disulfide rather than a thioester.

We point out that, as far as we are aware, there exists no other example of a cofactor being linked to its apoprotein via a disulfide bridge.

Structure of the chromophore

Although we have not been able to determine the exact structure of the chromophore, it is possible, on the basis of the above-mentioned data, to speculate on this structure in a fairly precise way. Because the chromophore can be cleaved by mild reduction in the same way as the reduction of a disulfide bridge, it is evident that one atom of the chromophore should be a sulfur atom. This leaves us with a mass of 115 Da for the remainder of the chromophore. The absorbance maximum at 332 nm of the peptides P28 and P30 (Fig. 11) is an indication that the chromophore very likely contains a conjugated system of at least four double bonds. Because we obtained no evidence from either the amino acid composition or the amino acid sequence analysis that fraction P20 contains peptide material, we believe that the absorbance of this fraction at 307 nm is due to the chromophoric molecule alone or, at the most, of the chromophore bound to the cysteine.

Recalling that the three-dimensional structure is related to those of the fatty acid binding and the retinol binding proteins, one may be inclined to think that the chromophore is related to the ligands bound by these proteins, i.e., being either a straight-chain unsaturated hydrocarbon or an isoprene derivative. The nearest isoprene derivative would be geranyl mercaptan ($C_{10}H_{15}S$-, 167.2

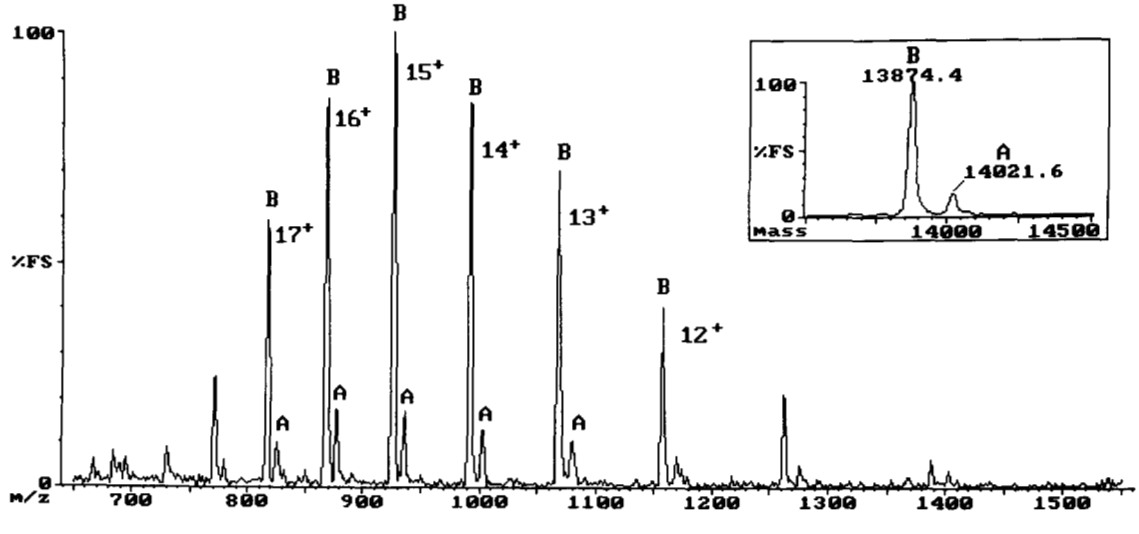

Fig. 15. Electrospray ionization mass spectrum of the native yellow protein after reduction with dithiothreitol (B). The *m/z* values for compound A refer to the mass of the native protein without reduction (see Fig. 6).

Da; structure I in Fig. 16). Although this structure would be the logical choice, being related to an intermediate in the biosynthesis of retinol, the mass is far too large to be considered. Isoprenylation of an internal cysteine residue is well documented in the literature, e.g., in the case of the P-21 ras-protein and ras-related proteins (James & Olson, 1990), although the isoprenyl group is bound to the cysteine via an ether bond (and not via a disulfide bridge) prior to cleavage of the last three residues of the CAAX-box by a carboxypeptidase. Protein prenylation is discussed in a recent "perspective" article (Marshall, 1993). Also documented is the case of bacterial lipoproteins in which the side chain of the N-terminal cysteine residue is modified by diacylglycerol prior to cleavage of the signal peptide (Nielsen & Lampen, 1982; Gotschlich & Seiff, 1987). Also in this case the linkage is of the thioether type.

The straight-chain hydrocarbon that most closely matches the mass of the chromophore is the nine-carbon mercaptan containing four conjugated double bonds ($C_9H_{11}S$-; structure II in Fig. 16). The total mass is 151 Da, a value 4 Da larger than that measured for the chromophore, which is also beyond the limits of accuracy of the electrospray ionization method used to determine the mass of the protein (0.03% instead of 0.01%). This structure is also less likely from a biosynthetic standpoint than is an isoprene, because odd carbon and highly unsaturated fatty acids are not common.

A structure that does fit the measured mass of the protein within the expected accuracy, however, is a mercaptoindole (C_8H_6NS-, 148.1 Da; structure III in Fig. 16). Difficulties with this compound are that there is no known biosynthetic route to a mercaptoindole and that an aromatic ring would not exhibit photoactivity such as one would obtain through isomerization about a double bond in retinol and in the unsaturated hydrocarbons.

Although thioesters appear to be ruled out by failure to release the chromophore through mild acid hydrolysis, they more closely match the other properties of the chromophore. For example, geranoic and 10-carbon unsaturated fatty acids (structures IV and V in Fig. 16) have masses of 149 and 147 Da, which is closer to the value of the chromophore (147 Da). Both could exhibit photoactivity, and biosynthesis would be simpler. Until PYP is cloned and overproduced in an appropriate expression system, insufficient protein is available to complete the identification of the chromophore through standard chemical analyses, such as NMR, IR, and fragmentation mass spectroscopy. In regard to the overproduction of PYP, it should be noted that specific enzymes are probably required for biosynthesis of the chromophore and to attach it to the apoprotein. Therefore, it may be necessary to clone it into an organism that is capable of producing PYP, which at present is limited to only three species, *E. halophila*, *R. salexigens*, and *C. salexigens*, none of which has been characterized genetically.

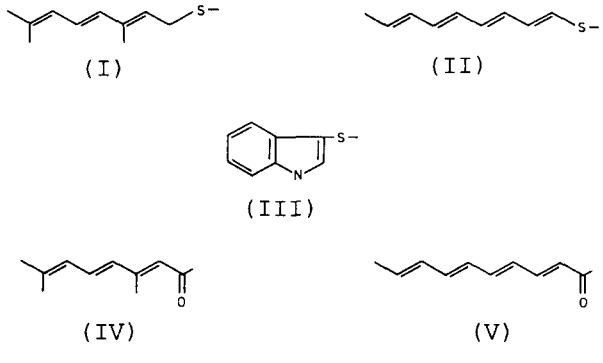

Fig. 16. Residue structures discussed as possible candidates for the structure of the chromophore of the yellow protein. The native chromophore very likely contains an SH group.

Finally, we stress the key role that the precise mass measurement (by ESMS) of the chromophore-containing peptide and of the native protein has played in the elucidation of the residue to which the chromophore is bound (cysteine) as well as of the nature of the binding (most likely a disulfide bridge). Because the protein was reduced prior to derivatization with iodoacetic acid, as is routinely done at the beginning of a sequencing project, this reduction resulted in the unexpected removal of the chromophore. Therefore, the chromophore binding residue was not detected because it eluted as a carboxymethyl derivative upon sequence analysis by Edman degradation. This is the opposite of what usually happens with covalent posttranslational modifications, namely that they are detected during Edman degradation by an unusual retention time of the residue to which it is bound or by the absence of any PTH derivative at all. It is the high-precision mass determination by ESMS of the native protein, after the unambiguous identification of each amino acid in the sequence, that has finally led us to the discovery of Cys 69 as the chromophore binding residue.

Supplementary material

On the Diskette Appendix, we supply the quantitative data for the PTH amino acid analyses of the more important peptides and of peptides that are not mentioned in Figure 1 but that all corroborate the sequence proposal. Together with these tables, we describe in more detail than in the main text the way the total sequence was obtained.

Acknowledgments

This work was supported by grant 2.0018.91 from the Belgian Fund for Joint Basic Research (to J.V.B.) and by grant DMB 8718678 from the National Science Foundation (to T.E.M.). W.D.H. is supported by the Dutch Organisation for Pure Research (NWO) via the Netherlands Foundation of Biological Research (BION).

References

Åkerström, B. & Lögdberg, L. (1990). An intriguing member of the lipocalin protein family: α-Microglobulin. *Trends Biochem. Sci. 15*, 240–243.

Alpers, D.H., Strauss, A.W., Ockner, R.K., Bass, N.M., & Gordon, J.I. (1984). Cloning of a cDNA encoding rat intestinal fatty acid-binding protein. *Proc. Natl. Acad. Sci. USA 81*, 313–317.

Cotter, R.J. (1988). Plasma desorption mass spectrometry: Coming of age. *Anal. Chem. 60*, 781A.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., & Whitehouse, C.M. (1990). Electrospray ionization – Principles and practice. *Mass Spectrom. Rev. 9*, 37–70.

Gotschlich, E.C. & Seiff, M.E. (1987). Identification and gene structure of an azurin-like protein with a lipoprotein signal peptide in *Neisseria gonorrhoeae*. *FEMS Lett. 43*, 253–255.

Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., & Downing, K.H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol. 213*, 899–929.

Hoff, W.D., Kwa, S.L.S., van Grondelle, R., & Hellingwerf, K.J. (1992). Low temperature spectroscopy of the photoactive yellow protein from *Ectothiorhodospira halophila*. *Photochem. Photobiol. 56*, 529–539.

Holden, H.M., Rypniewski, W.R., Law, J.H., & Rayment, I. (1987). The molecular structure of insecticyanin from the tobacco hornworm *Manduca sexta* L. at 2.6 Å resolution. *EMBO J. 6*, 1565–1570.

Hunt, C.R., Ro, J.H.-S., Dobson, D.E., Min, H.Y., & Spiegelman, B.C. (1986). Adipocyte P2 gene: Developmental expression and homology of 5′-flanking sequences among fat cell-specific genes. *Proc. Natl. Acad. Sci. USA 83*, 3786–3790.

James, G. & Olson, E.N. (1990). Fatty acylated proteins as components of intracellular signaling pathways. *Biochemistry 29*, 2623–2634.

Jones, T.A., Bergfors, T., Sedzik, J., & Unge, T. (1988). The three-dimensional structure of P2 myelin protein. *EMBO J. 7*, 1597–1604.

Li, E., Demmer, L.A., Sweetser, D.A., Ong, D.E., & Gordon, J.I. (1986). Rat cellular retinol-binding protein II: Use of a cloned cDNA to define its primary structure, tissue-specific expression, and developmental regulation. *Proc. Natl. Acad. Sci. USA 83*, 5779–5783.

Loo, J.A., Edmonds, C.G., Smith, R.D., Lacey, M.P., & Keough, T. (1990). Comparison of electrospray ionization and plasma desorption mass spectrometry of peptides and proteins. *Biomed. Environ. Mass Spectrom. 19*, 286–294.

Marshall, C.J. (1993). Protein prenylation: A mediator of protein–protein interactions. *Science 259*, 1865–1866.

McRee, D.E., Tainer, J.A., Meyer, T.E., Van Beeumen, J., Cusanovich, M.A., & Getzoff, E.D. (1989). Crystallographic structure of a photoreceptor protein at 2.4 Å resolution. *Proc. Natl. Acad. Sci. USA 86*, 6533–6537.

Meyer, T.E. (1985). Isolation and characterization of soluble cytochromes, ferredoxins and other chromophoric proteins from the halophilic phototrophic bacterium, *Ectothiorhodospira halophila*. *Biochim. Biophys. Acta 806*, 175–183.

Meyer, T.E., Fitch, J.C., Bartsch, R.G., Tollin, G., & Cusanovich, M.A. (1990). Soluble cytochromes and a photoactive yellow protein isolated from the moderately halophilic purple phototrophic bacterium, *Rhodospirillum salexigens*. *Biochim. Biophys. Acta 1016*, 364–370.

Meyer, T.E. & Kamen, M.D. (1982). New perspectives on *C*-type cytochromes. *Adv. Protein Chem. 35*, 105–212.

Meyer, T.E., Tollin, G., Causgrove, T.P., Cheng, P., & Blankenship, R.E. (1991). Picosecond decay kinetics and quantum yield of fluorescence of the photoactive yellow protein from the halophilic purple phototrophic bacterium, *Ectothiorhodospira halophila*. *Biophys. J. 59*, 988–991.

Meyer, T.E., Tollin, G., Hazzard, J.H., & Cusanovich, M.A. (1989). Photoactive yellow protein from the purple phototrophic bacterium, *Ectothiorhodospira halophila*. *Biophys. J. 56*, 559–564.

Meyer, T.E., Yakali, E., Cusanovich, M.A., & Tollin, G. (1987). Properties of a water-soluble, yellow protein isolated from a halophilic phototrophic bacterium that has photochemical activity analogous to sensory rhodopsin. *Biochemistry 26*, 418–423.

Miller, A., Leigeber, H., Hoff, W.D., & Hellingwerf, K.J. (1993). A light dependent branching-reaction in the photocycle of the yellow protein from *Ectothiorhodospira halophila*. *Biochim. Biophys. Acta 1141*, 190–196.

Monaco, H.L., Zanotti, G., Spadon, P., Bolognesi, M., Sawyer, L., & Eliopoulos, E.E. (1987). Crystal structure of the trigonal form of bovine beta-lactoglobulin and of its complex with retinol at 2.5 Å resolution. *J. Mol. Biol. 197*, 695–706.

Newcomer, M.E., Jones, T.A., Åqvist, J., Sundelin, J., Eriksson, U., Rask, L., & Peterson, P.A. (1984). The three-dimensional structure of retinol-binding protein. *EMBO J. 3*, 1451–1454.

Nielsen, J.B.K. & Lampen, J.O. (1982). Membrane-bound penicillinases in gram-positive bacteria. *J. Biol. Chem. 257*, 4490–4495.

Pervaiz, S. & Brew, K. (1987). Homology and structure function correlations between α_1-acid glycoprotein and serum retinol-binding protein and its relatives. *FASEB J. 1*, 209–214.

Scapin, G., Gordon, J.I., & Sacchettini, J.C. (1992). Refinement of the structure of recombinant rat intestinal fatty acid-binding apoprotein at 1.2 Å resolution. *J. Biol. Chem. 267*, 4253–4269.

Stavenga, D.G., Schwemer, J., & Hellingwerf, K.J. (1991). Visual pigments, bacterial rhodopsins, and related retinoïd-binding proteins. In *Photoreceptor Evolution and Function* (Holmes, M.G., Ed.), pp. 261–349. Academic Press, London.

Sundelin, J., Anundi, H., Trägardh, L., Eriksson, U., Lind, P., Ronne, H., Peterson, P.A., & Rask, L. (1985a). The primary structure of rat liver cellular retinol-binding protein. *J. Biol. Chem. 260*, 6488–6493.

Sundelin, J., Das, S.R., Eriksson, U., Rask, L., & Peterson, P.A. (1985b). The primary structure of bovine cellular retinoic acid-binding protein. *J. Biol. Chem. 260*, 6494–6499.

Sweetser, D.A., Birkenmeier, E.H., Klisak, I.J., Zollman, S., Sparkes, R.S., Mohandas, T., Lusis, A.J., & Gordon, J.I. (1987). The human and rodent intestinal fatty acid binding protein genes. *J. Biol. Chem. 262*, 16060–16071.