Complete Chemical Structure of Photoactive Yellow Protein: Novel Thioester-Linked 4-Hydroxycinnamyl Chromophore and Photocycle Chemistry†,#

Manuel Baca, Gloria E. O. Borgstahl, Maurice Boissinot,‡ Patrick M. Burke,§ DeWight R. Williams,∥ Kelly A. Slater,∥ and Elizabeth D. Getzoff∥

Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

Received August 23, 1994; Revised Manuscript Received October 19, 1994®

ABSTRACT: The unique ability of photoactive proteins to capture and use energy from a photon of light depends on the chromophore, its linkage to the protein, and the surrounding protein environment. To understand the molecular mechanisms by which a chromophore and protein interact to undergo a light cycle, we are studying photoactive yellow protein (PYP), a 14-kDa water-soluble photoreceptor from Ectothiorhodospira halophila with a photocycle similar to that of sensory rhodopsin. Here, we report the cloning and sequencing of the pyp gene and the chemical identification of both the chromophore and its covalent linkage to the protein. Elemental composition data from high-resolution mass spectrometry of a proteolytically derived chromopeptide, pH titrations and UV-visible spectroscopy of the protein-bound and chemically released chromophore, and fragmentation mass spectrometry of the liberated chromophore amide were combined with results from the 1.4-Å-resolution protein crystal structure to identify the chromophore in PYP as a 4-hydroxycinnamyl group covalently bound to the sole cysteine residue via a thioester linkage. While 4-hydroxycinnamate is a metabolic product of the phenylpropanoid pathway and a key molecule in plant stress response, this is the first report of covalent modification of a protein by this group. In the dark (yellow) state of PYP, the protein stabilizes the chromophore as the deprotonated phenolate anion. By combining our biochemical characterization of the chromophore with other published observations, we propose a chemical basis for the photocycle: following the initial absorption of a photon, the photocycle of PYP involves protonation of the chromophore to a neutral phenol form corresponding to the observed photobleached intermediate.

Photoactive yellow protein (PYP)† isolated from Ectothiorhodospira halophila is a small soluble protein of 14 kDa with a distinct bright yellow color, resulting from an absorbance maximum at 446 nm (Meyer, 1985). The action spectrum of the repellent response to blue light in E. halophila is centered at 472 nm (Meyer, 1985).
halophila matches the absorption spectrum of PYP (Sprenger et al., 1993), implicating PYP as the photoreceptor for the negative phototactic response. PYP undergoes a photocycle (Miller et al., 1993; Meyer et al., 1991, 1989, 1987) similar to that of the membrane-bound proteins bacteriorhodopsin, halorhodopsin, and the two sensory rhodopsins from the unrelated halophilic archaeabacterium Halobacterium halobi-

um. In the photocycle of PYP, the protein is excited to an intermediate I, with a red-shifted absorbance (\( \lambda_{\text{max}} = 495 \) nm) in less than 10 ns. The second intermediate (I, \( k \sim 10^4 \text{ s}^{-1} \)) is bleached to 340 nm. The photocycle is completed by a slower return to the dark state (\( k \sim 2 \text{ s}^{-1} \)) (Meyer et al., 1987). The absorbance maximum of the PYP holoprotein at 446 nm is shifted to about 340 nm upon denaturation (Meyer et al., 1987) or at pH below 2.7 (Meyer, 1985). The PYP photocycle involves the net uptake of one proton during formation of the fully bleached intermediate, followed by an equivalent proton release upon return of PYP to the ground state (Meyer et al., 1993). The unknown chromophore is covalently linked to Cys69 and has a mass of 147 Da (Van Beeumen et al., 1993). The unbleached ground-state structure of PYP has been determined to 1.4-Å resolution (G. E. O. Borgstahl, D. R. Williams, and E. D. Getzoff, manuscript in preparation), revising the initial chain tracing (McRee et al., 1989). PYP has an \( \alpha/\beta \) fold consisting of an N-terminal 28-residue helical lariat structure and a central 45-residue, chromophore-bound helical loop section located on opposite sides of a six-stranded \( \beta \)-sheet (G. E. O. Borgstahl, D. R. Williams, and E. D. Getzoff, manuscript in preparation).

Here, we report the cloning and DNA sequencing of the complete pyp gene. Furthermore, through combined analysis of the properties of both the protein-bound and chemically released chromophore, the high-resolution mass determination of a proteolytically derived chromopeptide fragment, and the 1.4-Å resolution electron density image, we have identified the chromophore in PYP as a 4-hydroxycinnamyl group covalently bound to Cys69 via a thioester linkage.

METHODS

Protein Production and Purification. E. halophila strain BN9626, kindly provided by T. E. Meyer, was grown in 15-L carboys as described previously (Meyer, 1985) and yielded about 4–5 g of cells (wet weight)/L. In a Waring blender at low speed, 180–200 g of cell pellets was homogenized in Tris/EDTA buffer [10 mM Tris buffer, pH 7.5, 2 mM ethylenediaminetetraacetic acid (EDTA)] plus 1.0 mM phenylmethylsulfonyl fluoride. The homogenized cells were then passed twice through a French pressure cell at 18 000 psi. The cell debris was removed by centrifugation at 16 000g for 5–6 h and supernatant dialyzed into Tris/EDTA buffer. This step was repeated on the pellet after sonication in Tris/EDTA buffer to extract additional PYP from the cell debris. Following dialysis, the combined lysate was loaded onto a DE-52 anion-exchange column (Whatman) and washed with Tris/EDTA buffer until the absorbance of the eluant at 280 nm was negligible. Then, yellow fractions containing PYP were eluted with 80 mM NaCl and 50 mM Tris, pH 7.5, and further purified with a 75% ammonium sulfate precipitation. After dialysis of the supernatant into Tris/EDTA buffer, PYP was concentrated over a YM-10 Amicon membrane to ≤50 mL, loaded onto an HQ-poros anion-exchange column (Perseptive Biosystems), and eluted at 25–50 mM NaCl with a NaCl gradient in Tris/EDTA buffer. All visibly yellow fractions were then dialyzed into Bis-Tris buffer, pH 6.5, and further purified by chromatofocusing with Polybuffer 74 (diluted 1:10) at pH 3.8 using a mono P column (Pharmacia). The PYP was then run on a G-75 sizing column (Pharmacia) to remove the ampholytes, dialyzed into 20 mM sodium phosphate buffer or Hepes buffer, pH 7.0, and concentrated to ∼20 mg/mL.

Cloning the pyp Gene. Two degenerate oligonucleotides were designed to amplify the PYP coding region from E. halophila chromosomal DNA by the polymerase chain reaction (PCR) using Taq polymerase (Innis et al., 1990). The oligonucleotide sequences were chosen by back-translation from the PYP amino acid sequence (Van Beeumen et al., 1993) with the codon usage patterns for highly expressed genes in enteric bacteria. Two regions with minimal genetic code degeneracy were selected. The 16-fold degenerate primer 5'ATACCATGGARGAYATYGARAACAC3' (where R = purine and Y = pyrimidine) included a NcoI restriction site and codons corresponding to PYP amino acid residues 9–14. The 32-fold degenerate primer 5'ATGTGACTANGCYTTTCATRTG3' (where N = purine or pyrimidine) included a SalI site and a stop codon as well as the complement of the sequence coding for residues 108–112. Amplification was achieved by 30–35 cycles of PCR with denaturation at 94 °C for 1 min, annealing at 37 °C for 2 min, and extension at 72 °C for 3 min. The main PCR product corresponded to the expected size for the pyp gene between the codons for amino acids 9 and 112. The 315 bp between the NcoI and SalI sites of the PCR product were cloned between the corresponding sites of the pPHSD1Lac1 (Getzoff et al., 1992) for production of a probe to be used in identifying the pyp gene from a partial genomic library and between the same sites of pBluescriptII SK− (Stratagene) for sequencing. Additional PCR amplifications and recloning of the 315-bp fragment were done using the high-fidelity Vent polymerase (New England Biolabs).

E. halophila genomic DNA was prepared according to the miniprep method of Ausubel et al. (1992). Preliminary Southern blotting experiments performed on complete restriction enzyme digests of E. halophila chromosomal DNA revealed that the smallest piece of DNA hybridizing with the cloned PCR product was a 2.5-kb PstI fragment. Consequently, DNA fragments resulting from PstI digestion of the E. halophila genomic DNA were fractionated by size with a sucrose gradient, and those containing fragments of approximately 2.5 kb of DNA were purified and ligated into pBluescriptII KS− (Stratagene) for sequencing. Additional PCR amplifications and recloning of the 315-bp fragment were done using the cloned PCR product as probe.

DNA Sequencing. Both the 315-bp PCR clones (resulting from amplifications with Taq and Vent polymerase) and a 1540-bp segment of the 2.5-kb genomic clone containing the pyp gene were sequenced on an Applied Biosystems automatic sequencer. Due to the GC-rich nature of the sample, the cycle sequencing reaction protocol recommended by the manufacturer (Applied Biosystems) was modified by adding 2.5% Triton X-100 "hydrogenated" (Calbiochem) and
4 additional units of AmpliTaq (Perkin-Elmer) and by raising the denaturation temperature to 98 °C. Additional internal sequencing primers were synthesized to allow complete sequencing of both DNA strands up to nucleotide 1094. Beyond this point, sequence data were acquired by sequencing only the coding strand with a minimum of two independent sequencing reactions. The sequence data were analyzed using the GCG (Genetics Computer Group) and Intelligenetics Suite computer software packages.

Mass Spectrometry. All mass spectrometry (MS) experiments, with the exception of high-resolution spectra, were performed on a Sciex API-III triple quadrupole ion-spray instrument. Sample introduction was usually by direct infusion into the spectrometer using a syringe pump (Harvard Applications) operating at 5 µL/min. For liquid chromatography/mass spectrometry (LC/MS) experiments, an Applied Biosystems 140B dual syringe pump HPLC was used in conjunction with an Applied Biosystems aquapore C8 reverse-phase column (2.1 × 100 mm). Linear acetonitrile gradients in 0.1% aqueous TFA were used at a flow rate of 150 µL/min, and approximately one-third the column eluant was directed into the mass spectrometer. For tandem mass spectrometry (MS/MS) experiments, collision-induced dissociation (CID) of the parent precursor ion was effected by collision with argon gas. The collision gas thickness was ~8 × 10^{14} atoms/cm², with the collision energy (quadrupole-2 rod offset voltage) set at ~50 V. High-resolution mass spectra were obtained on a VG ZAB-2VE mass spectrometer using a fast atom bombardment (FAB) ionization source and 3-nitrobenzyl alcohol as the sample matrix. Bradykinin was used as an internal mass calibrant (monoisotopic mass: 1059.5614) in determining the high-resolution mass of the chromopeptide fragment.

Proteolytic Digestion and Peptide Sequencing. PYP (130 µM solution in 10 mM ammonium bicarbonate, pH 7) was digested with 6 µM trypsin (Sigma) for 7 h at 37 °C. The digestion mixture was analyzed by mass spectrometry using both direct infusion and LC/MS. PYP was also digested with Staphylococcus aureus V8 protease (Boehringer Mannheim) using the same conditions as for trypsin, except in 100 mM sodium phosphate, pH 7.8. The V8 protease digestion mixture was analyzed by LC/MS.

PYP (5.44 µg in 4 µL of H₂O) was digested with the arginine-specific protease clostripain (Sigma, 2.5 µg or 0.25 unit in 16 µL of 10 mM Tris, pH 7.5, 2 mM DTT, and 1 mM CaCl₂, activated for 2.5 h in this buffer) for 2 h at 25 °C. The proteolytic fragment containing residues 53–124 was then isolated from the digestion mixture by PAGE on a 20% acrylamide gel blotted onto a ProBlot membrane in 10 mM CAPS, pH 10, 10% methanol, 5% SDS, and 5 mM DTT and N-terminally sequenced with an Applied Biosystems peptide sequencer.

Cleavage and Purification of Chromophore. Removal of the chromophore was done by two methods. In the first method, a 14 µM solution of PYP was successively dialyzed overnight at 4 °C in the presence of 1, 14, and 140 mM hydroxyamine adjusted to pH 7.0 with NaOH. In the second approach, the chromophore was released at room temperature by dilution of PYP into 4.5 M GuHCl and 0.5 M ammonium chloride, pH 10. Progress of the cleavage reaction at room temperature was monitored by observing the change in chromophore absorbance from 398 to 348 nm with a Hewlett-Packard 8542A diode array spectrophotometer. Following cleavage, the reaction mixture was acidified with an equal volume of glacial acetic acid, and the liberated chromophore was purified by reverse-phase HPLC on a C4 column (Vydac, 1.0 × 25 cm) using linear gradients of acetonitrile in 0.1% aqueous TFA. UV monitoring at 214 and 310 nm was used to distinguish between peaks containing the liberated chromophore and the apoprotein.

Titration Experiments. Spectrophotometric titrations were performed on protein-bound and chemically liberated chromophore and on 4-hydroxycinnamic acid. Unfolded holoprotein was titrated by diluting solutions of native protein 1:100 into 4 M GuHCl solutions buffered between pH 3 and pH 10.5 at 0.5 pH unit intervals. Titrations of chemically liberated chromophore and 4-hydroxycinnamic acid (Aldrich) were performed by diluting an aqueous solution of chromophore 1:100 into 33 mM solutions of Tris/Bis/Tris/CAPS (1:1:1) buffered between pH 8.25 and pH 9.75 at 0.25 pH unit intervals.

Crystallographic Refinement. The atomic structure of PYP, including all 125 amino acid residues, the 4-hydroxycinnamyl chromophore in thioester linkage with Cys 69, and 94 well-defined water molecules, has been determined and refined to 1.4-Å resolution (G. E. O. Borgstahl, D. R. Williams, and E. D. Getzoff, manuscript in preparation). The diffraction data with a signal-to-noise ratio (S/N) greater than 3 are 90% complete from 20- to 1.5-Å resolution, and more than 50% complete for the highest (1.4–1.5-Å) resolution data. The chromophore model was built into a 1.4-Å resolution F₂ - Fᵢ electron density map with phases calculated from the protein atoms only. Following crystallographic refinement with X-PLOR (Brünger et al., 1987), the crystallographic residual error (R-factor) between these data from 20- to 1.4-Å resolution and the complete model is 18.6%, and the root-mean-square deviations from ideality of bond lengths and angles are 0.012 Å and 1.57°, respectively, and there are no deviant main-chain torsion angles on the Ramachandran plot.

RESULTS

Cloning and Sequencing. The 315-bp PCR product including amino acids 9–112 of PYP was amplified from E. halophila chromosomal DNA and used to identify a 2.5-kb PstI clone containing the pyp gene from a partial genomic library. The pyp probe hybridized to only one DNA band in Southern blots of genomic DNA digested with a variety of restriction enzymes, suggesting that pyp occurs as a single-copy gene in the E. halophila chromosome.

A 1540-bp segment of the 2.5-kb PstI genomic clone was sequenced and analyzed for all possible open reading frames (ORF) on each strand (Figure 1). The G+C composition of this sequence is 67%, which is in the range expected for the genus (Stacey et al., 1989), and the codon usage shows a strong preference for codons ending with G (33%) or C (61%). The ORF encoding PYP was found between two other ORF’s encoding putative unknown proteins (Figure 1). The pyp ORF (bases 769–1143) is preceded by a potential ribosome binding site very similar to the canonical Shine-Dalgarno sequence, and it is followed by a pair of nearly perfect inverted repeats (bases 1157–1170 and 1174–1187) that could have a regulatory role at the transcriptional level. The partial translation product of the first ORF (bases 1–660) showed significant homology with E. coli D-amino acid...
dehydrogenase (a 47-kDa flavoprotein) with 37% identity and 61% similarity for a 218 amino acid overlap. Interestingly, a 47-kDa flavoprotein has been isolated from *Halobacterium halophilum* (Meyer, 1985). A third partial ORF (bases 1205-1540), located downstream from the first ORF, may also encode a protein since it is preceded by a potential ribosome binding site, but the sequence is not significantly homologous to any genes or proteins in GenBank 83.0 or EMBL 38.0 databases.

Translation of the nucleotide sequence of the *pyp* gene and flanking sequences. Numbering corresponds to nucleotides. For three open reading frames (ORF), the nucleotide sequence is translated into one-letter amino acid code below the first nucleotide of each codon.

![Sequence Alignment](image)

**Improved PYP Purification.** Modifications of the PYP purification protocol improved both protein yield and purity. To increase the efficiency of cell lysis, the resuspended cell pellets were homogenized in a Waring blender and then put through the French press twice. To extract all PYP from the cells, centrifugation and dialysis steps were repeated on the sonicated cell debris pellet. The 75% ammonium sulfate precipitation, additional anion-exchange chromatography, and chromatofocusing steps were added to improve purity. The yield was 66 μg of PYP/g of cells. The ratio of the 280:446 nm absorbances averaged 0.5, equivalent to the purity obtained in the best fractions of the original purification procedure, and has been as low as 0.4. By ion-spray MS, the measured mass of purified PYP is 14020 Da, consistent with its previously measured PI of 4.3 (McRee et al., 1986). PYP purified by using the chromatofocusing protocol improved both protein yield and purity. From acidic to slightly alkaline pH the chromophore absorption is maximally absorbed at 340 nm, but at more alkaline pH, crystals that do not twin.

**Chromophore Chemistry.** We measured the UV-visible absorption spectra of denatured PYP as a function of pH. From acidic to slightly alkaline pH the chromophore maximally absorbs at 340 nm, but at more alkaline pH, the absorption shifted to 398 nm (Table 1, lines 3 and 4). The pKₐ of this transition is 9.0 ± 0.5.

**Figure 1:** Nucleotide sequence of the *pyp* gene and flanking sequences.
judged from the change in absorption maxima. Analysis of broad protein peaks. The early-eluting peak could not be predicted for aminolysis of a chromophore-protein thioester bond by ammonia (Figure 2C). There was insufficient peaks absorbed at 214 nm but not at 350 nm, indicating that bance at 350 nm which presumably contained the liberated buffer in GuHCl resulted in a spectrophotometrically ob-

confirm the presence of the liberated chromophore, the early-eluting peak was characterized by LCMS and W spectros-

Apoprotein was, however, identified among the later eluting the chromophore was detached from the protein (Figure 2A). mophore as determined by LC/MS (which allowed the chromophore as an amide derivative. The late-eluting protein and 350 nm showed a small early-eluting peak with absor-

The reaction was deemed complete in approximately 4 h, as interpreted as chemical modification of the chromophore. The reaction mixture by HPLC at dual wavelengths of 214 and 350-nm monitoring clearly distinguishes the chromophore-bearing protein from other minor contaminants. Absorbances are not to the same scale, and each has been adjusted to adequately show the peaks of interest. Left pair of chromatograms show elution of native PYP. Right pair of chromatograms correspond to separation of PYP following aminolytic cleavage of chromophore by aqueous ammonium chloride, pH 10. Elution of the cleaved chromophore fragment (*) could be selectively identified by monitoring the column eluent at 310 nm, and this peak was collected for further study. Apoprotein (+) elutes within a broad peak and is detectable at 214 nm. (B) UV absorption spectra of cleaved chromophore, as purified by HPLC. Sample solvent was 25% acetonitrile in 0.1% aqueous TFA (pH 2). Absorption maxima occur at 224 nm and a broad peak centered at 300 nm. The ratio of 224:300-nm absorbance was 0.7. (C) MS analysis of liberated chromophore. HPLC-purified material was analyzed by LC/MS, and the mass spectrum shown corresponds to elution of the 310-nm absorbing species. The spectrum has been corrected for background ions by subtraction of the average background spectrum. As predicted for cleavage of a thioester protein—chromophore linkage by ammonia, the predominant m/z 164 peak corresponds with the protonated chromophore cleavage product ([M + H]+). The ion at m/z 147 is presumably a fragment ion resulting from loss of ammonia (NH3) from the protonated parent ion at 164 (also see Figure 3).

that the elution time of the mass 163 species matched that of the peak absorbing at 310 nm. The absorption spectrum of free chromophore in 25% acetonitrile and 0.1% TFA (pH 2) displayed maxima at 224 nm and at a broad peak centered at 300 nm (Figure 2B). While the extinction coefficients could not be determined due to unknown concentration, the ratio of the 224:300 nm absorbances was approximately 0.7: 1. The UV absorbance peak of the liberated chromophore shifted from 300 to ≈340 nm under alkaline conditions.

The mass of the chromophore was determined by ion-spray MS for both the whole protein and V8 protease-cleaved chromopeptide 66-VAPCTDSPE-74 (data not shown). The chromophore mass was derived from the difference (146 Da) between the experimentally measured mass and the mass calculated from the protein sequence alone. Taking into account that the calculated masses of apoprotein and peptide included the sulfhydryl proton of free cysteine and that this proton must be absent in the chromophore-bound polypeptides, the mass of the chromophore is 147 Da.

**Chemical Cleavage of the Chromophore from the Protein.** Hydroxylamine treatment of PYP cleaved the chromophore linkage to produce the apoprotein, as detected by MS analysis, but also caused other side reactions. Cleavage by hydroxylamine suggested a thioester linkage (Weimbs & Stoffel, 1992), rather than a disulfide or thioether linkage, to Cys69. To test this, PYP was treated with ammonia under alkaline conditions that were expected to result in thioester, but not disulfide, cleavage to the free thiol and the amide derivative of the chromophore (Connors & Bender, 1961; Jencks, 1969). If, in fact, the chromophore were linked via a thioester, then the predicted mass of the liberated chromophore amide would be 163, as calculated on the basis of the masses of acyl chromophore as attached to PYP (147 Da) and a terminal NH2 group (16 Da).

Treatment of PYP holoprotein with pH 10 ammonium buffer in GuHCl resulted in a spectrophotometrically observed shift in chromophore absorption maxima from 398 to 348 nm, displaying an isosbestic point at 362 nm. Given that the pH remained constant at 10, this change in \( \lambda_{\text{max}} \) was interpreted as chemical modification of the chromophore. The reaction was deemed complete in approximately 4 h, as judged from the change in absorption maxima. Analysis of the reaction mixture by HPLC at dual wavelengths of 214 and 350 nm showed a small early-eluting peak with absorbance at 350 nm which presumably contained the liberated chromophore as an amide derivative. The late-eluting protein peaks absorbed at 214 nm but not at 350 nm, indicating that the chromophore was detached from the protein (Figure 2A). Apoprotein was, however, identified among the later eluting broad protein peaks. The early-eluting peak could not be analyzed by ion-spray MS due to intense low-mass noise caused by salt contamination.

**Characterization of the Liberated Chromophore.** To confirm the presence of the liberated chromophore, the early-eluting peak was characterized by LC/MS and UV spectros-

The table below shows the acid-base absorption maxima of PYP chromophore:

<table>
<thead>
<tr>
<th>Protein-Bound or Cleaved Chromophore</th>
<th>pH</th>
<th>( \lambda_{\text{max}} )</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native protein</td>
<td>7</td>
<td>446</td>
<td>Meyer, 1985</td>
</tr>
<tr>
<td>Native protein</td>
<td>2</td>
<td>345</td>
<td>Meyer, 1985</td>
</tr>
<tr>
<td>Denatured protein</td>
<td>10</td>
<td>398</td>
<td>this work</td>
</tr>
<tr>
<td>Denatured protein</td>
<td>7</td>
<td>( \approx 340 )</td>
<td>Meyer, 1987</td>
</tr>
<tr>
<td>NH3 cleaved</td>
<td>13</td>
<td>( \approx 340 )</td>
<td>this work</td>
</tr>
<tr>
<td>NH2 cleaved</td>
<td>2</td>
<td>300</td>
<td>this work</td>
</tr>
</tbody>
</table>

**Table 1: Acid–Base Absorption Maxima of PYP Chromophore**

The UV absorbance peak of the liberated chromophore shifted from 300 to ≈340 nm under alkaline conditions.

**FIGURE 2:** Chemical cleavage and characterization of chromophore. (A) HPLC profiles of PYP before and after chemical release of chromophore. Peaks were eluted from a C4 reverse-phase column using a 1%/min acetonitrile gradient in 0.1% aqueous TFA. Run times were 60 min. Simultaneous 214- and 350-nm monitoring clearly distinguishes the chromophore-bearing protein from other minor contaminants. Absorbances are not to the same scale, and each has been adjusted to adequately show the peaks of interest. Left pair of chromatograms show elution of native PYP. Right pair of chromatograms correspond to separation of PYP following aminolytic cleavage of chromophore by aqueous ammonium chloride, pH 10. Elution of the cleaved chromophore fragment (*) could be selectively identified by monitoring the column eluent at 310 nm, and this peak was collected for further study. Apoprotein (+) elutes within a broad peak and is detectable at 214 nm. (B) UV absorption spectra of cleaved chromophore, as purified by HPLC. Sample solvent was 25% acetonitrile in 0.1% aqueous TFA (pH 2). Absorption maxima occur at 224 nm and a broad peak centered at 300 nm. The ratio of 224:300-nm absorbance was 0.7. (C) MS analysis of liberated chromophore. HPLC-purified material was analyzed by LC/MS, and the mass spectrum shown corresponds to elution of the 310-nm absorbing species. The spectrum has been corrected for background ions by subtraction of the average background spectrum. As predicted for cleavage of a thioester protein—chromophore linkage by ammonia, the predominant m/z 164 peak corresponds with the protonated chromophore cleavage product ([M + H]+). The ion at m/z 147 is presumably a fragment ion resulting from loss of ammonia (NH3) from the protonated parent ion at 164 (also see Figure 3).
FAB-MS. Given the chromophore-protein linkage to be a fragment ions at and aromatic nitrogen heterocycles is typically much lower which would be characteristic of a phenol (Davis et al., 1981). The absor-

High-Resolution Mass Spectrometry. In order to determine the elemental composition of the chromophore, we measured the high-resolution mass (1063.427 ± 0.015 Da) of the V8 protease derived chromopeptide, 66-VAPCTDSPE-74, by FAB-MS. Given the chromophore—protein linkage to be a thioester, the mass of the liberated chromophore amide was therefore derived as 163.074 ± 0.015 Da. Limiting the elements to C, H, N, O, and S, consideration of all potential elemental compositions of the form C,H,N,O,S, of mass 163.074 ± 0.015 Da generates over 30 possibilities. However, all but two of these candidates can be eliminated by imposing three constraints. First, the cleaved chromophore should contain at least one oxygenc atom (in the primary amide bond). Second, the UV absorption properties of cleaved chromophore indicate aromatic or conjugated double bonds; accordingly, the minimum index of hydrogen def-

Mass Spectrometric Fragmentation. Fragmentation patterns for the amide derivative of the free chromophore, as liberated by alkaline ammonia, were analyzed in a MS/MS experiment. The molecular ion at m/z 164 ([M + H]+) was fragmented by collision with argon gas to generate a well-defined set of daughter ions at m/z 147, 119, 91, and 65, along with minor ions at 103, 77, and 44 (Figure 3). Assuming a sequential loss of fragments to generate the observed ions, this CID spectrum indicates consecutive losses from the parent ion of 17 (NH3) followed by 28 (C=O) or, alternatively, loss of the charged [O=C=NH2]⁺ fragment (m/z 44). Consecutive ions at 91, 77, and 65 suggest the presence of a phenyl ring, while the loss of a second mass 28 fragment is consistent with a phenol substructure.

Furthermore, the fragment ions at m/z 91 and 77 do not support the presence of either annular nitrogen (aromatic heterocycle) or nitrogen external to an aromatic ring (aromatic amine). Loss of the mass 26 fragment (91 → 65) is undoubtedly due to ejection of the neutral HCN=CH, a feature commonly observed in the fragmentation of aromatic hydrocarbons.

MS fragmentation therefore provided additional proof that the chemically cleaved chromophore contained a primary amide, as predicted from aminolysis of a protein chromophore thioester bond by ammonia. In conjunction with the titration data, it provided strong evidence that the chromophore was phenolic, in which case the molecular formula of chemically cleaved chromophore amide can only be C8H7O2.

Crystallographic Chromophore Structure. The 1.4-Å-resolution "omit" electron density map (Figure 4) made with phases calculated from the PYP protein atomic model without the chromophore (G. E. O. Borgstahl, D. R. Williams, and E. D. Getzoff, manuscript in preparation) clearly defined the positions and bonding of the heavy (non-hydrogen) atoms in the chromophore. This map revealed the chromophore to be composed of a six-membered ring, with a single heavy atom substituent at position 1 of the ring and four heavy atoms in a connecting linker between position 4 of the ring and the sulfur of Cys69. All heavy atoms of the chro-

The PYP protein atomic model including the 4-hydroxy-

cinnamyl chromophore was refined against 1.4-Å diffraction data. The resulting stereochemistry (bond lengths are
indicated in Figure 4) confirm that O2 is a carbonyl oxygen and that there is a double bond between C7 and C8, thus extending conjugation from the aromatic ring through to the thioester bond. The refinement of the C1-O1 bond length was given special consideration. When the structure was refined with a 1.376Å restraint [typically used for Tyr hydroxyl (Brünger et al., 1987)] the refined bond length converged to 1.345Å. Setting the restraint to 1.310Å [value from potassium 4-(4-nitrophenyl)phenolate small molecule structure (Haase et al., 1991)] resulted in a value of 1.292Å. Finally, a restraint of 1.249Å [value typically used for Asp or Glu carboxylate (Brünger et al., 1987)] was used and resulted in a value of 1.240Å. This short C1-O1 bond length indicates a double bond and thereby a deprotonated state for O1 in the context of the folded protein.

Comparison of the Cleaved Chromophore to 4-Hydroxycinnamic Acid. Proof of the proposed chromophore structure was provided by comparing the properties of the chemically cleaved chromophore with commercially available 4-hydroxycinnamic acid. This compound differs from the liberated chromophore in that it is the free acid, while the latter is a primary amide (Figure 5A). The UV absorption spectrum and MS/MS fragmentation pattern of 4-hydroxycinnamic acid support the proposed chemical structure of the cleaved chromophore as 4-hydroxycinnamide (Figure 5B,C). Both compounds exhibit absorption maxima at about 225 and 300 nm with a similar 225:300-nm absorbance ratio. Fragmentation of the molecular ion of 4-hydroxycinnamic acid (Figure 5C) generated the same set of major daughter ions as observed for the cleaved chromophore (Figure 3). This indicates these two compounds are composed of the same structural elements. The pK of the phenolic hydroxyl in 4-hydroxycinnamic acid is 9.5 ± 0.3 (for transition between 310 and 340 nm), a value similar to that obtained for the cleaved chromophore (9.0 ± 0.3). The 0.5 increase in pK for 4-hydroxycinnamic acid relative to the 4-hydroxycinnamide chromophore is not surprising, given that deprotonation of the phenolic hydroxyl is somewhat disfavored by the negatively charged carboxylate group.

DISCUSSION

The first step in photosensing is the capture of energy from a photon. Establishing the chemical nature of a chromophore, its covalent linkage to the protein, and the protein’s exact primary structure is of fundamental importance in defining the mechanism of protein photosensing. Here we have cloned and sequenced the DNA encoding PYP (Figure 1). The pyp gene is the first protein-coding gene ever sequenced from a member of the Ectothiorhodospira genus, the single genus of the Ectothiorhodospiraceae family of purple sulfur bacteria (Stacey et al., 1989). The Ectothiorhodospiraceae family is smaller and less well characterized than the other purple sulfur bacterial family, Chromatiaceae. Genetic studies have been recently undertaken for some species of the Chromatiaceae, especially for Chromatium vinosum; whereas in the Ectothiorhodospiraceae, only ribosomal RNA genes have been sequenced so far. The translated amino acid sequence from the pyp gene matches the previously published protein sequence (Van Beeumen et al., 1993) with one exception: Gln56 replaces Glu56. The DNA and revised protein sequences for PYP (Figure 1) not only establish the protein’s primary structure but also open the door to future mutagenesis studies.

In PYP, the covalent linkage of the chromophore to Cys69 is cleaved by dithiothreitol, suggesting a disulfide or thioester but not a thioether bond (Van Beeumen et al., 1993). On the basis of the inability of 25% aqueous TFA to cleave chromophore from PYP or its chromopeptides, these researchers further proposed that the chromophore is attached through a disulfide rather than a thioester bond and noted
Accelerated Publications

Figure 5: Chemical structure and spectroscopic properties of 4-hydroxycinnamic acid. (A) Chemical structures of 4-hydroxycinnamic acid (left) and the chemically cleaved chromophore 4-hydroxycinnamamide (right). (B) UV absorption spectra of 4-hydroxycinnamic acid. Sample solvent was 25% acetonitrile in 0.1% aqueous TFA (pH 2). Absorption maxima occur at 228 and 310 nm and the 228:310 nm absorbance ratio is 0.5. The absorption properties of 4-hydroxycinnamic acid are thus very similar to those of the cleaved chromophore (cf. Figure 2). (C) MS/MS fragmentation of 4-hydroxycinnamic acid. The protonated molecular ion at m/z 165 was fragmented by collision with neutral argon gas. Mass and identity of sequential fragment losses from the parent ion were 18 (H$_2$O), 28 (C=O), 28 (C=O), and 14 (CH$_2$) or 26 (HC=CH). These losses resulted in a set of daughter ions (147, 119, 91, 77, and 65) essentially identical to that obtained for the cleaved chromophore, both in mass and relative intensity (cf. Figure 3).

During the photocycle, PYP takes up a single proton upon bleaching and subsequently releases it upon return to the dark state (Meyer et al., 1993). When the protein is denatured at neutral pH, the covalently bound chromophore is protonated, as shown by pH titrations (Table 1, lines 3 and 4). The 340-nm absorption maximum characteristic of the protonated chromophore in the denatured protein has a striking resemblance to that of the bleached intermediate (I$_2$) of the PYP.
photocycle (Meyer et al., 1987). Likewise, very acidic pH titration of the native protein bleaches the absorption maximum of the chromophore (Table 1, line 2), indicating that the dark state chromophore is not only deprotonated but also buried and that the protonation at pH 2.7 is probably correlated to solvent exposure of the chromophore. The anionic form of the chromophore in the dark state and the 340-nm absorbance common to both the bleached PYP intermediate and the protonated protein-bound chromophore suggest that the PYP light cycle occurs through the protonation and deprotonation of the chromophore.

ACKNOWLEDGMENT

We thank Darren J. Murtari, Che-Fu Kuo, Jessica Bessler, and Nikhil Munshi for their contributions to the preparation and purification of PYP; Terry E. Meyer, John A. Tainer, and Duncan E. McRee for help in starting the PYP project at TSRI; Lisa Bibbs, Gary Siuzdak, and Stephen Kent for use of facilities and equipment; Mark Erlander for sequencing tips; and Victoria A. Roberts and Ulrich Genick for stimulating and thought-provoking discussions.

REFERENCES


