# Chemical structure of the hexapeptide chromophore of the Aequorea Green-fluorescent protein

**ARTICLE** *in* BIOCHEMISTRY · MARCH 1993

Impact Factor: 3.02 · DOI: 10.1021/bi00056a003 · Source: PubMed

# Chemical Structure of the Hexapeptide Chromophore of the *Aequorea* Green-Fluorescent Protein†

Chris W. Cody,‡,§ Douglas C. Prasher,‖,△ William M. Westler,⊥ Franklyn G. Prendergast,# and William W. Ward*,‡

*Department of Biochemistry & Microbiology, Rutgers University, Cook College, New Brunswick, New Jersey 08903, Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, Department of Biochemistry, University of Wisconsin—Madison, Madison, Wisconsin 53901, and Department of Biochemistry & Molecular Biology, Mayo Foundation, Rochester, Minnesota 55901*

*Received December 17, 1991; Revised Manuscript Received June 4, 1992*

ABSTRACT: The green-fluorescent proteins (GFP) are a unique class of proteins involved in bioluminescence of many cnidaria. The GFPs serve as energy-transfer acceptors, receiving energy from either a luciferase–oxyluciferin complex or a $Ca^{2+}$-activated photoprotein, depending on the organism. Upon mechanical stimulation of the organism, GFP emits green light spectrally identical to its fluorescence emission. These highly fluorescent proteins are unique due to the nature of the covalently attached chromophore, which is composed of modified amino acid residues within the polypeptide. This report describes the characterization of the *Aequorea victoria* GFP chromophore which is released as a hexapeptide upon digestion of the protein with papain. The chromophore is formed upon cyclization of the residues Ser-dehydroTyr-Gly within the polypeptide. The chromophore structure proposed here differs from that described by Shimomura [(1979) *FEBS Lett. 104*, 220] in a number of ways.

Green in vivo light emission in the jellyfish *Aequorea victoria* occurs from a noncatalytic green-fluorescent protein (GFP)[1] via energy transfer from calcium-triggered aequorin. Morise et al. (1974) first identified GFP as the energy acceptor and ultimate emitter in *Aequorea* bioluminescence. *Aequorea* GFP exists as a monomer having a $M_r$ of 27 000 (Shimomura, 1979) to 30 000 (Prendergast & Mann, 1978). GFP is extremely conformationally stable (Nageswara et al., 1980; Shimomura & Shimomura, 1981; Roth & Ward, 1983), with denaturation occurring only under very harsh conditions (Ward & Bokman, 1982; Bokman & Ward, 1981).

*Aequorea* GFP absorbs light maximally at 395 nm and exhibits a smaller absorbance peak at 470 nm (Morise et al., 1974; Ward et al., 1980). The fluorescence emission spectrum peaks at 509 nm with a shoulder at 540 nm (Morise et al., 1974). Upon denaturation, GFP loses its fluorescence and undergoes large absorption spectral changes (Bokman & Ward, 1981; Ward & Bokman, 1982). Spectrophotometric titration of denaturated GFP results in an absorption peak at 380 nm at acidic pH and one at 445 nm at basic pH. The pH-dependent transition shows an isobestic point at 405 nm with a p$K$ of 8.1 (Ward et al., 1980). Protease-derived chromopeptides from GFP have been isolated that display the same pH-dependent spectral characteristics exhibited by denatured GFP (Shimomura et al., 1979; Ward et al., 1980).

Spectral characteristics of GFP are due to a covalently bound chromophore. According to a model proposed by Ward et al. (1980), different absorption spectra of GFPs from various sources could be attributed to changes in noncovalent interactions between the chromophore and other parts of the protein. Precise folding of the protein around the chromophore would confer a unique set of spectral characteristics to native GFP including fluorescence.

Under some conditions, the loss of fluorescence that accompanies denaturation of GFP appears to be irreversible; however, renaturation of denatured *Aequorea* GFP (up to 90%) has been demonstrated (Bokman & Ward, 1981; Ward & Bokman, 1982). Surpin and Ward (1989) have recently demonstrated a thiol requirement for maximum renaturation. Renaturation data imply that the chromophore remains intact when GFP is denatured; upon proper protein refolding, the chromophore returns to a suitable environment and fluorescence is restored.

Shimomura (1979) proposed the structure of the GFP chromophore to be a substituted 4-(*p*-hydroxybenzylidene)-5-imidazolone moiety (structure I, Figure 1). We have re-examined the GFP chromophore for four reasons. (1) The structure identification proposed by Shimomura (1979) was based on analogy with a model compound, the structure of which was not supported with adequate physical evidence. (2) Amino acid analysis of the chromopeptide indicated the presence of glycine which was not incorporated into the proposed structure (Shimomura, 1979). (3) Almost no quantitative data to support the proposed structure were presented by Shimomura (1979) and the identification of the C-terminus of the peptide was left ambiguous. (4) McCapra et al. (1988) argues convincingly that Shimomura's proposed structure cannot totally explain the spectroscopic properties of the chromophore. We have elucidated the complete amino acid sequence and structure of the papain-derived chromophore-containing hexapeptide, which differs from that proposed by Shimomura (1979), and provide an explanation

† A paper of the journal series New Jersey Agricultural Experiment Station. This work was performed as part of NJAES Project No. 01102.

* To whom correspondence and reprint requests should be addressed.

‡ Rutgers University, Cook College.

§ Present address: Department of Biochemistry, John Hopkins School of Hygiene & Public Health, Baltimore, MD 21205.

‖ Woods Hole Oceanographic Institution.

△ Present address: USDA/APHIS, Building 1398, Otis ANGB, MA 02542.

⊥ University of Wisconsin—Madison.

# Mayo Foundation.

[1] Abbreviations: BTI, [bis(trifluoroacetoxy)iodo]benzene; DFP, diisopropyl fluorophosphate; EDTA, ethylenediaminetetraacetic acid; *gfp*, DNA encoding GFP; GFP, green-fluorescent protein; HOHAHA, homonuclear Hartmann–Hahn; HPLC, high-performance liquid chromatography; $M_r$, relative molecular mass; PITC, phenylisothiocyanate; PMSF, phenylmethanesulfonylfluoride; RT, retention time in seconds; TLC, thin-layer chromatography; TSP, (trimethylsilyl)propionic acid.

0006-2960/93/0432-1212$04.00/0 

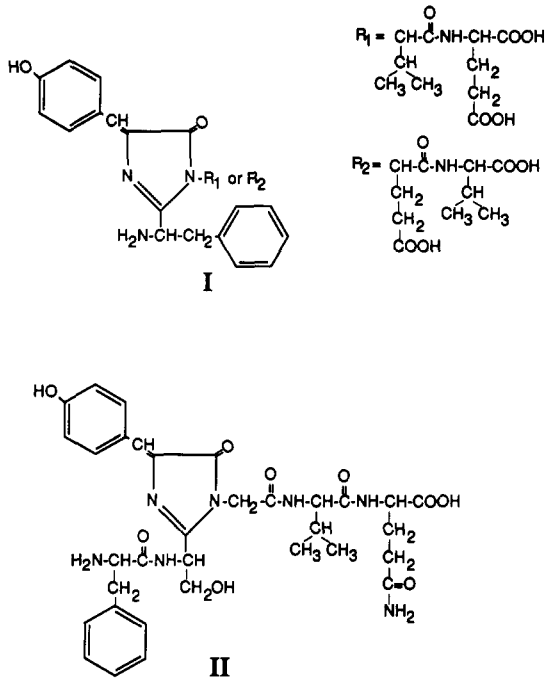

FIGURE 1: Proposed structures of *Aequorea* GFP hexapeptide chromophore: (I) from Shimomura (1979); (II) the chromophore structure supported here; (II) the chromophore structure supported here. The heterocyclic ring is composed of Ser-dehydroTyr-Gly.

for the generation of the chromophore from the primary sequence of apoGFP.

## EXPERIMENTAL PROCEDURES

*A. Aequorea GFP Purification.* Specimens of the hydromedusan jellyfish *A. victoria* were collected at the University of Washington's Friday Harbor Laboratories, Friday Harbor, Washington, during the late summer and early fall of several years. Initial dissection and biochemical extraction were performed according to Blinks et al. (1976) as modified by A. F. Roth and W. W. Ward (unpublished). *Aequorea* GFP was purified to electrophoretic homogeneity using a four-step chromatographic sequence employing Bio-Gel P-100, DEAE Bio-Gel, and Sephadex G-75 superfine (A. F. Roth and W. W. Ward, unpublished). Electrophoretically pure GFP exists as three major isoproteins each with an absorption ratio ($A_{395}/A_{280}$) between 1.10 and 1.25.

*B. Peptide Purification.* Green-fluorescent protein was denatured prior to protease digestion by heating the protein solution to 90 °C for 5 min. Buffer composition at this stage varied but was generally 1 mM Tris, 1 mM EDTA, pH 8.0. Following denaturation, the GFP solution was allowed to cool to room temperature before treatment with a protease. Alternatively, the native protein was denatured with dilute acetic acid (pH 3–4).

Papain and pronase were obtained from Sigma Chemical Co., St. Lous, MO. Carboxypeptidase (DFP- or PMSF-treated) was obtained from Worthington Biochemicals, Freehold, NJ.

Papain (Sigma Type IV) digestions were carried out in 1 mM sodium phosphate, 1 mM EDTA, 5 mM cysteine at pH 6.2 and 37 °C. Pronase (Sigma Type XIV) digestions were carried out in 10 mM sodium phosphate, 0.02% sodium azide, pH 7–7.5 at 37 °C. Carboxypeptidase digestions were carried out either in 25 mM Tris, 0.5 M sodium chloride, pH 7.5, or in 0.2 M *N*-ethylmorpholine, pH 7.5. Protease digestions were halted by acidifying the buffer to a pH between 1 and 3 with phosphoric or hydrochloric acid. The samples were stored at −20 °C until further use.

Carboxypeptidase activity was assayed using the method of Folk and Shirmer (1963) as revised by Decker (1977). Other proteases were assayed by using Azocoll (Calbiochem-Behring Corp.).

A Bio-Rad ODS-5S reverse-phase HPLC column (4 × 250 mm) was used except where noted. High-performance liquid chromatography was carried out on a Spectra Physics SP8000. Column eluent was monitored with two detectors in series. The optical unit built into the SP8000 was used with a 365-nm filter to monitor the chromophore; an LDC Spectromonitor II variable wavelength detector was used to monitor peptide absorbance at 210 nm. All HPLC runs were performed in acidic buffer solutions, so the GFP chromophore was always in the 380-nm-absorbing form (see Results).

Standard HPLC conditions included the following. The column was equilibrated with a starting buffer of 10 mM sodium phosphate at pH 4 at a flow rate of 1 mL/min. Peptide-containing samples were injected onto the column in starting buffer. Five minutes after sample injection, a linear acetonitrile gradient (+1% acetonitrile/min, flow rate 1 mL/min) was initiated until all components in the mixture were eluted. Following each HPLC run, the column was rinsed with 100% acetonitrile.

*C. Analytical Spectra.* A Cary 17-D recording spectrophotometer and a Bausch and Lomb Spectronic 2000 were used interchangeably for fixed-wavelength absorbance measurements or for spectral scans.

Plasma desorption mass spectra were obtained on a BIO-ION Nordic (Uppsala, Sweden), BIN-10K californium-252 plasma desorption time-of-flight mass spectrometer (Macfarlane & Torgerson, 1976; Sundqvist et al., 1984) using an accelerating voltage of −12 kV for positive ions and a flight tube length of 15 cm. Spectra were generally acquired for approximately 1 h. The recorded time-of-flight spectra were converted to mass spectra using the time centroids for $H^+$ and $Na^+$ as calibration peaks.

Homonuclear Hartmann–Hahn spectra (Rance, 1987) were acquired on a Bruker AM-500 spectrometer at the National Magnetic Resonance Facility at Madison, WI. Proton chemical shifts are referenced to internal TSP. HOHAHA data were acquired with the proton decoupler channel for proton excitation. The decoupler channel and the receiver were referenced to a common frequency.

*D. Amino Acid Analysis.* Amino acid analysis was performed using the method from a Waters PICO Tag system following pre-column dervitization with PITC. Glutamine analysis was performed independently by the method of Soby and Johnson (1981) using the reagent [bis(trifluoroacetoxy)-iodo]benzene (BTI).

*E. p-Hydroxybenzaldehyde Analysis.* A 2-cm Dowex 50W-X3 cation-exchange column was equilibrated with 2 mM phosphoric acid (pH 2.8). Following application of 1 mL of acid hydrolysate of the hexapeptide, the column was rinsed with 1 mL of 2 mM phosphoric acid. The unretained column void (~2 mL) was collected and analyzed by HPLC with comparison to a *p*-hydroxybenzaldehyde standard. For HPLC analysis of *p*-hydroxybenzaldehyde, the Bio-Rad ODS-5S column was used with a mobile phase consisting of water and acetonitrile. Following equilibration of the column with water, the sample was injected in dilute aqueous phosphate buffer. Five minutes after injection of the sample, a linear gradient of acetonitrile (+2%/min, flow rate 1 mL/min) was initiated. The HPLC column eluent was monitored at 220 nm.

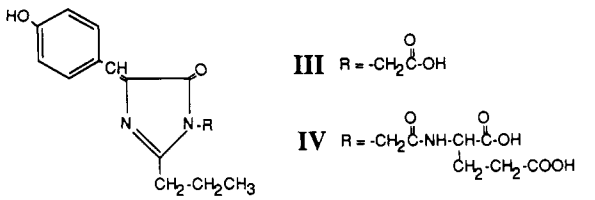


FIGURE 2: Proposed structures of synthetic model chromophores III and IV.

*F. Synthetic Model Chromophores.* A model compound (structure III, Figure 2) was prepared and initially extracted according to the method of Shimomura (1979). The crude mixture containing the desired compound was dissolved in ethyl acetate and extracted with 1 N hydrochloric acid. The pH of the aqueous solution was adjusted to 1 with sodium hydroxide, and the mixture was then extracted with ethyl acetate (Shimomura, 1979). Beyond this point, the procedure of Shimomura was no longer followed. The sample was evaporated to dryness and then dissolved in a mixture of 50% methanol and 50% 1 mM boric acid, pH 9.2. The product was purified by chromatography of DEAE Bio-Gel in a 1:3 mixture of methanol and 1 mM boric acid, pH 9.2, using an increasing gradient of NaCl in the same methanolic/aqueous buffer solution. Purification was completed by HPLC using standard conditions with the Bio-Rad ODS-5S column. In agreement with the data of Shimomura (1979), standard amino acid analysis of compound III (6 N HCl, 110 °C, 24 h) revealed 2 mol of glycine/mole assuming compound III has the same extinction coefficient at 370 nm as the limit digest chromopeptide has at 380 nm. The structure of compound III has been verified by mass spectroscopy and NMR (Ward et al., manuscript in preparation).

A second synthetic model chromophore (structure IV, Figure 2) was prepared according to the same procedure (Shimomura, 1979) except that glycylglutamate was used in place of glycylglycine.

Purification of compound IV followed the sequence used for compound III through the ethyl acetate extraction. Then, the dried product was dissolved in 1 M potassium phosphate buffer, pH 2, and applied to a Sephadex LH-20 column equilibrated in the same buffer. After passsage of several column volumes of buffer through the column, compound IV was eluted from the column with a mixture of 1:1 methanol and 1 mM boric acid, pH 9.2. Purification was completed by HPLC using the Bio-Rad ODS-5S column under standard conditions.

## RESULTS

*A. Papain Digestions of Aequorea GFP.* Following heat denaturation at 90 °C for 5 min, pure *Aequorea* GFP (1 mg/mL) was incubated with papain for a period of 12 h. Three doses of papain (each 15% by weight with respect to GFP) were added at time intervals of 0, 4, and 8 h. Aliquots (100 μL) were removed from the digestion mixture 15 min and 4 h after each addition of papain. HPLC analysis showed that one major chromopeptide, the papain limit digest peptide, is released by papain digestion of *Aequorea* GFP under the most extensive digestion conditions employed. It eluted as a sharp peak at retention time 2330 s. The limit digest chromopeptide is apparently readily released from *Aequorea* GFP by papain, as it was the main 365-nm-absorbing peak seen at any time sampled, including the shortest digestion time. For preparative work, a longer incubation time (24 h) with 15% papain was routinely employed, as the resulting peptide mixture proved easier to purify.

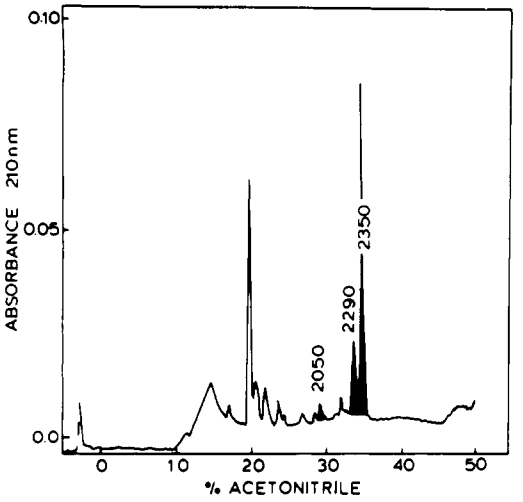


FIGURE 3: Preparative HPLC purification of papain-derived *Aequorea* GFP chromopeptide (step 6). The 210-nm-absorbing peaks are drawn to scale; the 365-nm-absorbing peaks are shaded (not to scale). Retention times are in seconds.

Table I: Preparative Purification of Papain-Derived *Aequorea* GFP Chromopeptide

| purification step | $A_{395}$ units | $A_{380}$ units | recovery |
|---|---|---|---|
| 1. initial | 14.1 | | |
| 2. dialysis vs 1 mM EDTA, pH 6.2 | 13.5 | | 96% |
| 3. denaturation, 90 °C, 5 min | | | |
| 4. papain digest, 15% by wt, 24 h | | 11.0 | 76% |
| 5. concentration and Bio-Gel P-2 chromatography | | 7.4 | 67% |
| 6. concentration and HPLC | | 4.8 | 65% |
| cumulative | | 4.8 | 32% |

*B. Preparative Purification of Papain-Derived Aequorea GFP Chromopeptide.* Table I summarizes the purification scheme which includes two chromatography separations using Bio-Gel P-2 (step 5) and reverse-phase HPLC (step 6). On the basis of absorbance at 380 nm and the extinction coefficient for denaturated GFP (Ward, 1981), 76% of the chromophore was recovered following papain digestion. At pH 3, the chromopeptide is only weakly charged, interacting fairly strongly with the Bio-Gel P-2 column beads by hydrophobic attraction. Thus, it elutes at nearly twice the column volume, effecting a very efficient separation from other peptides. The P-2 pool (42–60 mL) contained 7.4 absorbance units at 380 nm (67% recovery).

Figure 3 shows the chromatogram of the HPLC separation (step 6) for the P-2 pool, monitored at both 210 nm and 365 nm. The limit digest chromopeptide eluted at retention time 2350 s and had higher absorbance at 210 and 365 nm than any other peptide. Two minor chromopeptides eluted at 2290 and 2050 s (Figure 3). Absorption spectra of all three chromopeptides, normalized to the 380-nm absorption peak, are shown in Figure 4. The limit digest chromopeptide (2350 s) exhibits the lowest relative absorbance at all wavelengths below 350 nm, evidence that this is the smallest peptide with an intact chromophore that is released by papain treatment. Total yield of chromophore-containing peptide, based on published extinction coefficients for *Aequorea* GFP (Ward, 1981), was 32%. Pure papain-derived chromophore-containing peptide is characterized by its low absolute absorbance in the UV from 230 nm to 300 nm.

*C. Papain Treatment of Aequorea GFP Isoforms. Aequorea* GFP is known to exist as several isoforms. Three

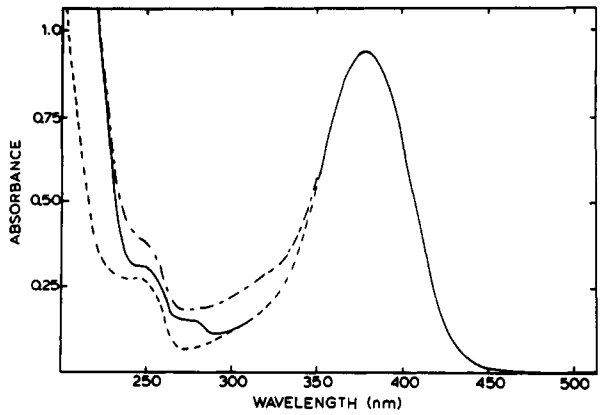


FIGURE 4: Absorption spectra of three different papain-derived GFP chromopeptides. The chromopeptides are the 365-nm-absorbing species indicated in Figure 3. The spectra were recorded at pH 4: RT 2350 (- - -); RT 2290 (—); RT 2050 (– – –)

major isoforms designated m, l, and c have been identified and highly purified (A. F. Roth and W. W. Ward, unpublished). The possibility that different GFP isoforms may give rise to different papapain-generated chromophore-containing peptides was investigated. Between 0.2 and 0.5 mg of each of the three GFP isoforms was digested separately with 5% (w/w) papain for a period of 3 h. Each digestion mixture was chromatographed on a 1 × 21 cm P-2 Bio-Gel column equilibrated with 0.01 M sodium phosphate, pH 3. A single 380-nm-absorbing peak eluted from the column in all three cases and was collected and concentrated for HPLC. HPLC analysis using an Excalibar ODS-5S column resulted in a single 365-nm-absorbing peak eluting at retention time 2440 s.[2] To verify that the chromopeptide was in fact identical from the three isoforms, equal amounts of each peptide were mixed together. The resultant pool was chromatographed on HPLC yielding a single 365-nm-absorbing peak eluting at 2440 s.

To be sure that all 380-nm-absorbing material had eluted from the Bio-Gel P-2 column, the HPLC analysis was repeated omitting the size-exclusion chromatographic step. In each case, a single 365-nm-absorbing peak eluting at 2440 s was observed. Papain must cleave all three GFP isoforms to the same limit chromopeptide. We reject, therefore, the hypothesis that multiple 365-nm-absorbing peaks, seen in some papain digestions, arise from different isoforms.

*D. Characteristics of Aequorea GFP Limit Digest Chromopeptide.* The absorption spectrum (Figure 5) of the papain limit digest chromopeptide is identical to that described by Shimomura (1979). At acidic pH, the peptide exhibits an absorption maximum at 380 nm which shifts to 445 nm as the pH is raised (Figure 5). This pH-dependent absorbance shift has an isosbestic point at 405 nm. These spectral characteristics are identical to those of denaturated *Aequorea* and *Renilla* GFP in 6 M GuHCl (Ward et al., 1980), indicating that proteolytic treatment has not altered the conjugation system of the "denatured chromophore".

Plasma desorption mass spectroscopy of the limit digest chromopeptide yielded the molecular ion ($MH^+ = 680.9$ where 680.7 was expected). No fragmentation pattern was observed.

One mole of acid-hydrolyzed limit digest chromopeptide (RT 2350, Figure 3) was found to contain 1 mol each of Glx, Val, and Phe and 2 mol of Gly (Table II, part A). Under standard hydrolysis conditions, 0.19 mol of Ser was detected;

[2] Note, the Excalibar column produced slightly longer retention times than the Bio-Rad column.

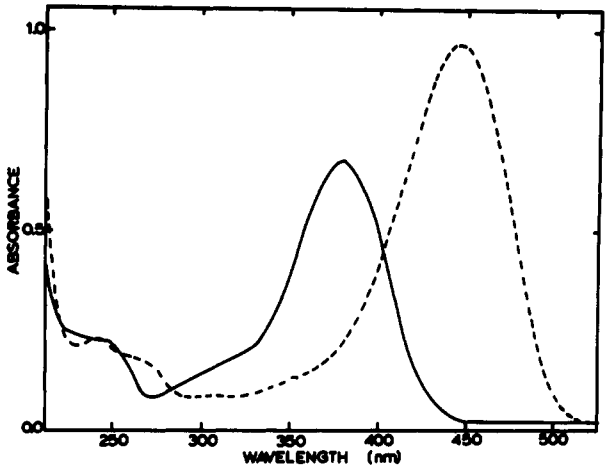


FIGURE 5: Absorption spectra of purified *Aequorea* GFP papain-derived chromopeptide. The chromopeptide is that which had a retention time of 2350 s in Figure 3. Spectra were recorded at pH 4 (—) and pH 11 (- - -).

Table II: Amino Acid Analysis of Chromopeptides[a] from *Aequorea* GFP

| amino acid | A. papain-derived peptides RT 2350 | RT 2290 | RT 2050 | B. peptides following secondary proteolytic treatment[b] RT 1900 (carboxypeptidase-treated) | RT 1920 (pronase-treated) |
|---|---|---|---|---|---|
| Asp | 0.15 (0)[c] | 0.17 (0) | 0.50 (0–1) | | |
| Thr | 0.09 (0) | 1.6 (2) | 0.54 (0–1) | | |
| Ser | 0.19 (0)[d] | 0.0 (0) | 0.65 (0–1) | | |
| Glx | 1.0 (1)[e] | 1.0 (1) | 1.0 (1) | (0) | (0) |
| Gly | 1.9 (2) | 1.9 (2) | 2.3 (2) | (2) | (2) |
| Ala | 0.16 (0) | 0.06 (0) | 0.35 (0) | | |
| Val | 0.94 (1) | 1.0 (1) | 0.99 (1) | (0) | (1) |
| Ile | 0.12 (0) | 1.9 (2) | 0.17 (0) | | |
| Leu | 0.05 (0) | 0.0 (0) | 0.42 (0) | | |
| Tyr | 0.05 (0) | 1.4 (1) | 0.40 (0) | | |
| Phe | 1.2 (1) | 3.1 (3) | 3.2 (3) | (1) | (0) |

[a] Purified through step 6 and identified according to their retention times. [b] The papain limit digest chromopeptide (RT 2350) was treated further. Serine recovery was negligible under the harsh hydrolysis conditions (6 N HCl, 110 °C, 24 h). [c] Amino acids were calculated as moles per mole of chromopeptide. The first number in each column is the integrated value; the number in parentheses is the nearest integer. [d] Shorter hydrolysis times yielded successively greater yields of serine. At $3^1/_2$ h hydrolysis time, the yield was 0.49. [e] Independent analysis for glutamine by the method of Soby and Johnson (1981) showed that Glx is all in the form of glutamine in the intact chromopeptide.

however, as hydrolysis time was shortened to 3.5 h, the yield increased to 0.49. Analysis for glutamine indicated very clearly that the amino acid reported as Glx in Table II entirely derived from glutamine. Separate analysis for *p*-hydroxybenzaldehyde in the hydrolysate yielded 0.88 mol of *p*-hydroxybenzaldehyde/mol of chromopeptide.

Amino acid analysis of the two minor chromopeptides (RT 2290 and 2050, Figure 3) revealed the same five amino acids, but there was an enrichment in phenylalanine and several other amino acids (Table II). The extra phenylalanine(s) could account for the relatively high absorbances seen below 300 nm but cannot account for the high absorbance between 300 nm and 340 nm especially characteristic of the peptide that elutes at 2050 s (Figure 4). Thus, indications are that the two additional 365-nm-absorbing peptides are larger than the main papain-derived chromopeptide and also that the peptide eluting at retention time 2050 s may have undergone some chromophore degradation.

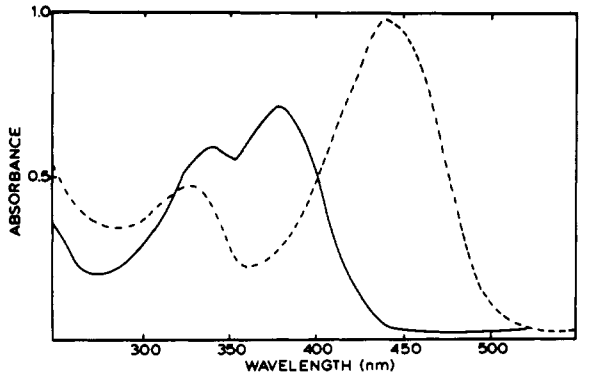


FIGURE 6: Absorption spectra of carboxypeptidase-digested papain limit digest GFP chromopeptide, HPLC purified (Bio-Rad ODS-5S column) using standard conditions. Spectra were recorded at pH 4 (—) and pH 11 (- - -).

To obtain amino acid sequence information, purified papain-derived limit digest chromopeptide was further digested with two other proteases. Pronase cleaved the chromopeptide to a new 365-nm-absorbing product that eluted at 1920 s under standard HPLC conditions and exhibited an unaltered absorption spectrum in the near UV. Amino acid analysis of the purified pronase digest chromopeptide yielded 1 mol of Val and 2 mol of Gly/mol of chromopeptide (Table II, part B). Glutamic acid and Phe were not detected, and Ser was present as a trace.

Carboxypeptidase treatment of the limit digest chromopeptide resulted in the generation of a smaller chromopeptide that eluted at 1900 s under standard HPLC conditions. The HPLC-purified chromopeptide showed an altered absorption spectrum (Figure 6). A newly generated peak absorbing near 340 nm was seen under both basic and acidic conditions. Amino acid analysis of this chromopeptide after acid hydrolysis revealed 1 mol of Phe and 2 mol of Gly/mol of chromopeptide, indicating that carboxypeptidase removed Glx and Val from the carboxyl-terminal end of the peptide (Table II, part B).

*E. 2D NMR.* Results of two-dimensional NMR are consistent with structure II (Figure 1). Homonuclear Hartmann–Hahn spectra of the aromatic and aliphatic regions are presented in Figure 7A,B, respectively. The cross-peaks in the spectra appear at the intersection of the chemical shifts of protons that are scalar coupled. The cross peak at (8.0, 6.93 ppm) in Figure 7A is due to the coupling of the $\delta$ and $\epsilon$ protons of the tyrosine ring. The peak at (8.0, 8.0 ppm) is shifted far downfield from the normal position in tyrosine (6.8 and 7.1 ppm). This may be indicative of the extra conjugation supplied by the vinyl group and the imidazolone ring. The single peak on the diagonal at 7.29 ppm is probably the single $\beta$ proton of the tyrosine vinyl group. This chemical shift is also consistent with a conjugated vinyl group. The peaks near 6.9–7.1 are due to the phenylalanine aromatic ring protons. When the region is expanded, the peak at 7.1 is a doublet and, therefore, is probably the $\delta$ protons of the ring. They are coupled to the $\epsilon$ protons which are located at 7.4 ppm. The $\zeta$ proton overlaps with the highfield peak of the tyrosine ring protons at 6.93 ppm.

The quality of the aliphatic region of the two-dimensional HOHAHA spectrum (Figure 7B) is not as good as the aromatic region due to limited amounts of material. Therefore, the following assignments described should be regarded as tentative. The peak at 4.85 ppm (labeled Ser) has a cross-peak with one at 3.8 ppm. This could be due to the coupling between the $\alpha$ and $\beta$ protons of an intact serine residue. No vinyl protons are observed other than the vinyl proton observed

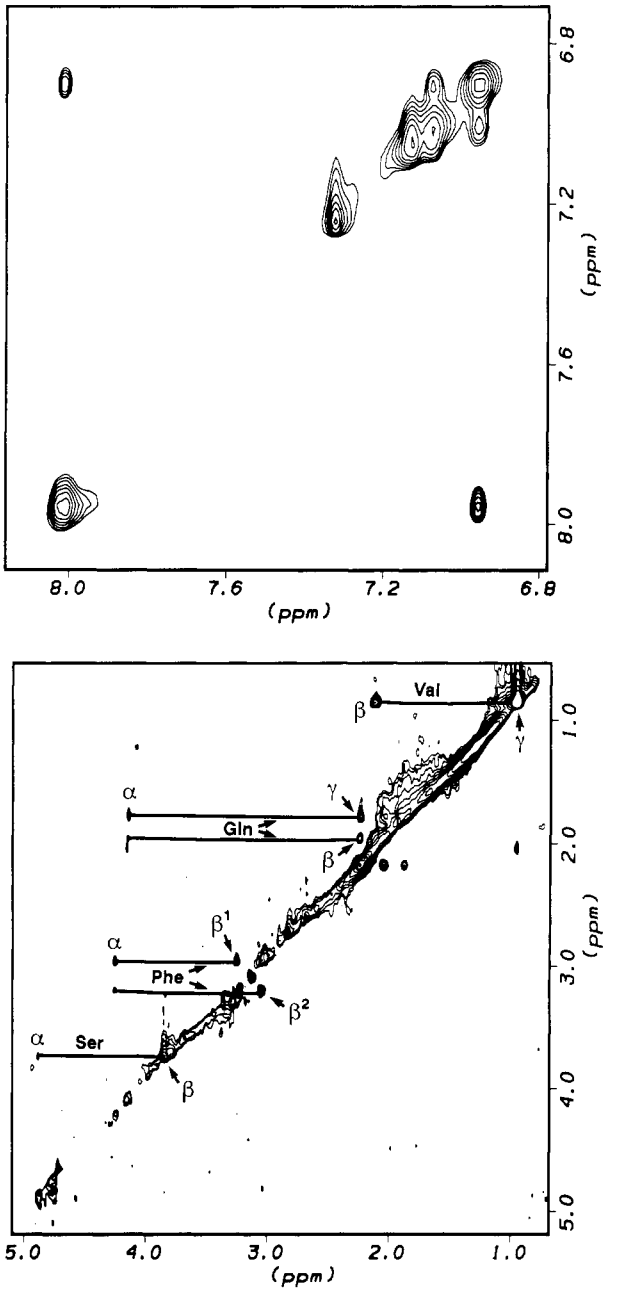


FIGURE 7: HOHAHA maps of the papain-derived GFP chromopeptide. (A, Top) Aromatic region. (B, Bottom) Aliphatic region. The Greek letters identify the atoms involved in the correlation. The spectrometer frequency was 500 MHz, and the spectral width (both dimensions) was 6024 Hz. The total experiment consisted of 512 blocks (40 scans/block), each of 4096 words. The mixing time was 28 ms with a total recycle time of 0.138 seconds.

with the tyrosine (Figure 7A). The cross-peak (3.0, 3.3 ppm, labeled Phe in Figure 7B) is probably due to the $\beta$ protons of the phenylalanine side chain. The peaks near the diagonal in the region around 2.0–2.3 ppm are consistent with the protons of the glutamine side chain. Last, the peak at (0.9, 2.1 ppm) is probably due to the coupling between one or both valine methyl protons and the $\beta$ methine proton. Cross-peaks are also observed in the vertical columns near 4–4.5 ppm. These peaks may be used for the assignment of $\alpha$ protons. Some of the peaks may be "relay" peaks between $\gamma$ and $\alpha$ protons. Relay peaks arise between two protons that are not directly coupled but have a common coupling partner. This effect is a result of the HOHAHA pulse sequence. The glycine protons do not give rise to any visible peaks, possibly because they may overlap with the serine $\beta$ protons.

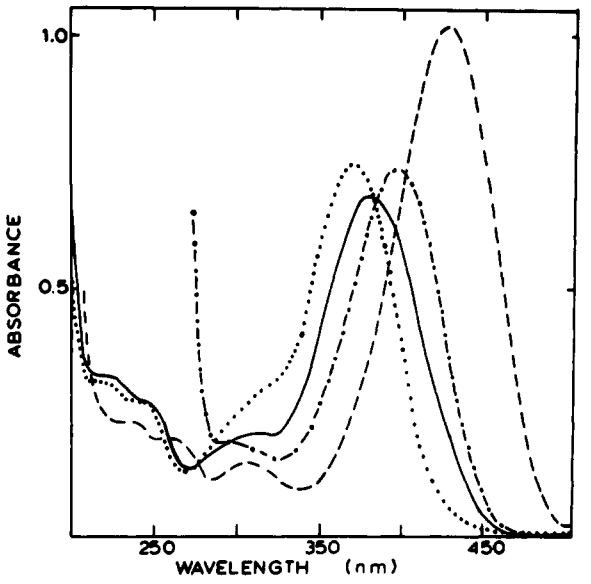


FIGURE 8: Absorption spectra of purified synthetic model compound III. Spectra were recorded at pH 1 (– · –), pH 2 (—), pH 3 (· · ·), or pH 11 (- - -).

*F. Characteristics of Synthetic Model Chromophores.* The absorption spectra of compounds III and IV (Figure 2) are similar, but not identical, to the absorption spectrum of intact *Aequorea* GFP papain-derived chromopeptide (Figure 8). At pH 12 compounds III and IV have absorption maxima at 430 nm, and at acidic pH (3–7.5) their maxima are at 370 nm. These absorption maxima differ significantly from the wavelength maxima of the limit digest chromopeptide (445 nm and 380 nm at basic and acidic pH, respectively). Further spectral differences between the native GFP chromophore and compound III were seen below pH 3. From pH 1 to 3, both native and synthetic chromophores have absorption maxima in the vicinity of 380 nm. The absorption spectrum of compound III, however, shows a peak whose exact position is pH dependent. As the pH is lowered below 3, the absorption maximum of compound III continuously red shifts. At pH 2, in agreement with Shimomura (1979), the absorption maximum of compound III is 380 nm. However, only near pH 2 do the absorption peaks of the limit digest chromopeptide and compound III match reasonably well. Above pH 2, the compound III absorption maximum is blue-shifted. At pH 1, the wavelength maximum for compound III is red-shifted to 395 nm. Thus, it is difficult to draw conclusions about the structure of the natural chromophore largely on the basis of spectral comparisons with synthetic model compounds as done by Shimomura (1979).

## DISCUSSION

While Shimomura (1979) digested large quantities of GFP (100 mg) with papain, we used less than 1 mg (33 nmol) of GFP in many experiments by developing a more efficient purification scheme. Shimomura's (1979) methodology consisted of butanol extraction from acidified water (pH 1) followed by TLC (three steps). We were unable to repeat these procedures satisfactorily; butanol extraction generated new spectrally-altered products and TLC resulted in extremely poor recovery of chromopeptide (data not shown). Chromatography at pH 3 of the papain digest on a Bio-Gel P-2 column followed by reverse-phase HPLC proved satisfactory for purification of the limit digest *Aequorea* GFP chromopeptide. Shimomura (1979) reported that a single product absorbing above 300 nm is generated when *Aequorea* GFP is digested with papain. We found that under similar digestion conditions, three chromophore-containing peptides could be detected by HPLC (Figure 3). One of these peptides (the limit digest chromopeptide) generally contained greater than 90% of the total 380 nm absorbance and it has an absorption spectrum very similar to that described by Shimomura (1979). Presumably, our limit digest chromopeptide is identical to the chromopeptide isolated by Shimomura.

The results of secondary proteolytic treatment of the papain limit digest chromopeptide with pronase or carboxypeptidase A are summarized in Table II, part B. Both carboxypeptidase and pronase generated smaller chromophore-containing peptides from the papain-derived limit-digest chromopeptide. These smaller peptides eluted at earlier retention times than the limit digest chromopeptide (Table II, part B); amino acid analysis showed that secondary protease treatment of the limit digest chromopeptide resulted in the cleavage of two amino acids from the peptide. Since carboxypeptidase A cleaved both valine and glutamate/glutamine from the peptide, one of them must be the carboxyl terminus of the peptide. Pronase cleavage of glutamate/glutamine and phenylalanine, but not valine, from the chromopeptide indicates that the amino terminus is phenylalanine and the carboxyl terminus is glutamate or glutamine.

The above results indicate that the carboxyl terminus of the peptide (Val-Glx) should be the correct identity of the R group in the structure of the chromopeptide proposed by Shimomura (1979) (structure I, Figure 1). The successful synthesis of glycylglycine and glycylglutamate model chromophores (Figure 2) supports the assignment of Shimomura (1979) on the functional portion of the chromophore as being 4-(*p*-hydroxybenzylidene)-5-imidazolone. However, several lines of evidence reported here suggest that structure I is not totally correct.

The data presented here from the protease digestions and mass spectral analysis, in addition to the deduced translation of a *gfp* cDNA (see below), support structure II for the chromopeptide (Figure 1). Amino acid analysis and enzymatic cleavage of the peptide demonstrated that a phenylalanine is present at the amino terminus and that the carboxyl-terminal sequence is Val-Glx. This amino acid sequence is consistent with the information derived from the *gfp1* cDNA which contains the nucleotide sequence encoding the peptide Phe-Ser-Tyr-Gly-Val-Gln (Prasher et al., 1992). The composition of the cDNA translation is in agreement with amino acid analysis of the chromopeptide which showed that the peptide also contains a glycine, a dehydrotyrosine, and possibly a serine residue (Table II). Mass spectral analysis of the chromopeptide indicates a parent ion of 680.9 mass units, significantly greater than the molecular weight (522) of the structure proposed by Shimomura (1979). These results led us to propose a new structure for the chromophore-containing peptide (structure II, Figure 1). The structure differs from that described by Shimomura (structure I, Figure 1) by containing 6 amino acids as opposed to 4, and by including a serine moiety instead of phenylalanine in the 2-position of the imidazolone ring and a glycine residue instead of glutamate or valine in the 1-position of the imidazolone ring. Furthermore, the carboxyl terminus is now unambiguously determined to be glutamine.

The HOHAHA results are consistent with the structure proposed here (Figure 1). The tyrosine side chain is modified to a vinyl group and the serine side chain is intact. The heterocyclic ring might arise from a cyclization reaction of the hexapeptide Phe-Ser-dehydroTyr-Gly-Val-Gln which

could be formed by a Schiff's base linkage of the serine carboxyl carbon and the glycyl amino group.

The posttranslational events required for chromophore formation are not yet understood. It is very unlikely that the chromophore forms spontaneously, but its formation probably requires some enzymatic machinery. It is not known whether the chromophore forms in concert with protein synthesis or whether it forms posttranslationally. Physical studies suggest that the chromopeptide in the native protein is buried within the polypeptide. Ward et al. (1980) have hypothesized that the chromophores of the *Aequorea* and the *Renilla* GFPs are protected from solvent by the polypeptide. Fluorescence polarization and oxygen quenching measurements also suggest that the chromophore is held rigidly within the polypeptide (Nageswara Rao et al., 1980). Thus, if the apoGFP polypeptide maintains a folded conformation like that of the fluorescent protein and the apochromopeptide is buried, its modification must likely occur prior to protein folding. Thus, the chromophore would form in concert with protein synthesis. If, on the other hand, the apoGFP has a very different conformation from that of the native protein, the apochromopeptide may be accessible to soluble modification factors. In this case, conversion of the apochromopeptide may be an intermediate step in the folding pathway and thus the chromophore would form posttranslationally. The biochemical requirements for chromophore formation can be addressed when recombinant apoGFP is available.

## ACKNOWLEDGMENT

We thank Dr. A. O. D. Willows, Director, Friday Harbor Laboratories, for the use of laboratory facilities. We acknowledge technical assistance from Peter Callahan, Willem Westerhuis, Ben Madden, and Tony Tsarboboulos.

## REFERENCES

Blinks, J. R., Prendergast, F. G., & Allen, D. G. (1976) *Pharmacol. Rev. 28*, 1–93.

Bokman, S. H., & Ward, W. W. (1981) *Biochem. Biophys. Res. Commun. 101*, 1372–1380.

Decker, L. A., Ed. (1977) *Worthington Enzyme Manual*, Worthington Biochemical Corp., Freehold, NJ.

Folk, J. E., & Schirmer E. W. (1963) *J. Bio. Chem. 238*, 3884–3894.

Macfarlane, R. D., & Torgerson, D. F. (1976) *Science 191*, 1084–1087.

McCapra, F., Razavi, Z., & Neary, A. P. (1988) *J. Chem. Soc., Chem. Commun.*, 790–791.

Morise, H., Shimomura, O., Johnson, F. H., & Winant, J. (1974) *Biochemistry 13*, 2656–2662.

Nageswara Rao, B. D., Kemple, M. D., & Prendergast, F. G. (1980) *Biophys. J. 32*, 630–632.

Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G., & Cormier, M. J. (1992) *Gene 111*, 229–233.

Prendergast, F. G., & Mann, K. G. (1978) *Biochemistry 17*, 3448–3453.

Rance, M. (1987) *J. Magn. Reson. 74*, 557–564.

Roth, A. F., & Ward, W. W. (1983) *Photochem. Photobiol. 37*, S71.

Shimomura, O. (1979) *FEBS Lett. 104*, 220–222.

Shimomura, O., & Shimomura, A. (1981) *Biochem. J. 199*, 825–828.

Soby, L. M., & Johnson, P. (1981) *Anal. Biochem. 113*, 149–153.

Sundqvist, B., Kamensky, I., Häkansson, J., Kjellberg, J., Salehpour, M., Widdiyasekera, S., Fohlman, J., Peterson, P. A., & Roepstorff, P. (1984) *Anal. Chem. 11*, 242–257.

Surpin, M. A., & Ward, W. W. (1989) *Photochem. Photobiol. 49*, 62S.

Ward, W. W. (1981) in *Bioluminescence and Chemiluminescence* (DeLuca, M. A., & McElroy, W. D., Eds.) pp 235–242, Academic Press, New York.

Ward, W. W., & Bokman, S. H. (1982) *Biochemistry 21*, 4535–4540.

Ward, W. W., Cody, C. W., Hart, R. C., & Cormier, M. J. (1980) *Photochem. Photobiol. 31*, 611–615.

Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W., & Reeves, S. C. (1982) *Photochem. Photobiol. 35*, 803–808.

Ward, W. W., Cody, C. W., Prasher, D. C., & Prendergast, F. G. (1989) *Photochem. Photobiol. 49*, 62S.

**Registry No. II**, 142895-83-8; **III**, 71642-46-1; **IV**, 142868-84-6.