Pittz, E. P., & Bello, J. (1971) *Arch. Biochem. Biophys. 146*, 513-524.
Pittz, E. P., & Timasheff, S. N. (1978) *Biochemistry 17*, 615-623.
Rogers, J. A., & Tam, T. (1977) *Can. J. Pharm. Sci. 12*, 65-70.
Scatchard, G. (1946) *J. Am. Chem. Soc. 68*, 2315-2319.
Schachman, H. K., & Lauffer, M. A. (1949) *J. Am. Chem. Soc. 71*, 536-541.
Stockmayer, W. H. (1950) *J. Chem. Phys. 18*, 58-61.
Tanford, C. (1961) *Physical Chemistry of Macromolecules*, Wiley, New York, NY.
Timasheff, S. N., & Townend, R. (1961) *J. Am. Chem. Soc. 83*, 470-473.
Timasheff, S. N., Mescanti, L., Basch, J. J., & Townend, R. (1966a) *J. Biol. Chem. 241*, 2496-2501.
Timasheff, S. N., Townend, R., & Mescanti, L. (1966b) *J. Biol. Chem. 241*, 1863-1870.
Timasheff, S. N., Lee, J. C., Pittz, E. P., & Tweedy, N. (1976) *J. Colloid Interface Sci. 55*, 658-663.
Townend, R., Winterbottom, R. J., & Timasheff, S. N. (1960) *J. Am. Chem. Soc. 82*, 3161-3168.
Witz, J., Timasheff, S. N., & Luzzati, V. (1964) *J. Am. Chem. Soc. 86*, 168-173.

Amino Acid Sequence of the Calcium-Dependent Photoprotein Aequorin†

Harry Charbonneau,*,‡ Kenneth A. Walsh,‡ Richard O. McCann,§ Franklyn G. Prendergast,∥ Milton J. Cormier,§ and Thomas C. Vanaman⊥,#

Department of Biochemistry, University of Washington, Seattle, Washington 98195, Department of Biochemistry, University of Georgia, Athens, Georgia 30602, Department of Pharmacology, Mayo Foundation, Rochester, Minnesota 55901, and Department of Microbiology and Immunology, Duke University, Durham, North Carolina 27710

Received April 23, 1985

ABSTRACT: The Ca(II)-dependent photoprotein aequorin produces the luminescence of the marine coelenterate *Aequorea victoria*. The complete amino acid sequence of aequorin has been determined. A complete set of nonoverlapping peptides was produced by cyanogen bromide cleavage. These peptides were aligned by using the amino-terminal sequence of the intact protein and the sequences of selected arginyl and lysyl cleavage products. Although the aequorin preparations employed in these studies were homogeneous by sodium dodecyl sulfate–polyacrylamide gel electrophoresis, the presence of a minimum of 3 isotypes was demonstrated by the location of 17 sites of sequence microheterogeneity. Two amino acid variants were observed at each of 16 positions while 1 position had 3 different replacements. The protein as isolated has 189 amino acids with an unblocked amino terminus. According to the sequence reported here, the molecular weight of the apoprotein is 21 459 while that of the holoprotein is 21 914. The molecule possesses three internally homologous domains which were judged to be EF-hand Ca(II) binding domains by several different criteria. Aequorin is homologous to troponin C and to calmodulin. These findings demonstrate that aequorin is a member of the Ca(II) binding protein superfamily.

Aequorin is the protein responsible for the bioluminescence of the marine coelenterate *Aequorea victoria* (Arai & Brinckman-Voss, 1980). Aequorin was first isolated from these jellyfish by Shimomura et al. (1962) and was shown to produce a blue luminescence in response to added Ca(II). The luminescent reaction of aequorin and other photoproteins is initiated solely by the addition of Ca(II) and requires neither additional molecular oxygen nor any other cofactor for production of the energy (about 60 kcal/mol) released as light and heat.

Aequorin is comprised of a single polypeptide chain and contains 1 mol of a tightly bound chromophore termed coelenterate-type luciferin [cf. reviews by Cormier (1978), Blinks et al. (1976), and Cormier et al. (1974)]. This chromophore is a substituted dihydropyrazinimidazolone ring system (M_r 423) which has been found to be a common substrate in the bioluminescent reaction of the photoproteins as well as a large number of diverse marine organisms (Shimomura & Johnson, 1978; Ward & Cormier, 1975; Hori et al., 1977). Upon the addition of Ca(II), the luciferin of aequorin is converted to oxyluciferin plus CO_2. Coelenterate-type oxyluciferin is a 2,3,6-substituted pyrazine compound which is generated by the oxidative decarboxylation of luciferin. The light emission (λ_{max} = 469 nm) occurs from the first singlet excited state of oxyluciferin (Shimomura & Johnson, 1973, 1975a; Hori et al., 1973). Studies of other related bioluminescent systems have clearly established the dependency of light emission on the concentration of molecular oxygen (Cormier et al., 1975). Consequently, the lack of an O_2 dependency for the aequorin reaction has led to the postulate that aequorin has some form of tightly bound oxygen at a site that is independent of the luciferin binding site (Cormier, 1978). In this model, the binding of Ca(II) to aequorin triggers luminescence by inducing a conformational change in the protein that allows the oxygen species to react with luciferin to produce luminescence. An alternative hypothesis, based largely on chromophore extraction experiments and chemical modification of luciferin derivatives, speculates that the oxygen occurs as a linear peroxide that covalently links the luciferin to an amino acid

†Supported by National Institutes of Health Grants to K.A.W. (GM15731), T.C.V. (NS10123), and F.G.P. (GM30178) and by the University of Georgia Research Foundation (to M.J.C.).
‡University of Washington.
§University of Georgia.
∥Mayo Foundation.
⊥Duke University.
#Present address: Department of Biochemistry, Albert B. Chandler Medical Center, University of Kentucky, Lexington, KY 40536-0084.

0006-2960/85/0424-6762$01.50/0

side chain of the protein (Shimomura & Johnson, 1978). In vitro, the photoprotein does not turn over since the reaction produces a stable complex of oxyluciferin and protein. However, upon removal of the oxyluciferin product, aequorin can be regenerated by the addition of synthetic luciferin in the presence of 2-mercaptoethanol and oxygen (Shimomura & Johnson, 1975b).

The luminescence of aequorin is initiated by micromolar Ca(II) concentrations. The Ca(II) dependency of the light emission is best described by a two-state model in which there are three equivalent Ca(II) sites with a dissociation constant of 0.1 μM (Allen et al., 1977). The finding that Mg(II) competitively inhibits the Ca(II) activation suggests that at least one of the Ca(II) sites also binds Mg(II) (Blinks et al., 1976). Since aequorin is an intracellular, high-affinity Ca(II) binding protein, it has been suggested as a possible member of the superfamily of "EF-hand" proteins that includes troponin C, calmodulin, and parvalbumins (Kretsinger, 1980).

Although the chemistry of the light reaction itself is well characterized, little is known about the luciferin binding site, the nature of the protein-bound oxygen species, or how the binding of Ca(II) initiates this reaction. In order to provide a structural basis for elucidating these mechanisms, we have determined the complete amino acid sequence of aequorin. Although the data fit a single structure of 189 residues, a surprisingly large number of sites (17) in the sequence has 2 (or more) alternate residues. This sequence microheterogeneity indicates the presence of at least three isotypes of aequorin. The results suggest that aequorin is a member of the EF-hand family of proteins with three internally homologous EF-hand Ca(II) binding domains.

Materials and Methods

N^{α}-(*p*-Tosyl)-L-phenylalanine chloromethyl ketone–trypsin (TPCK–trypsin),[1] *Staphylococcus aureus* V8 protease, soybean trypsin inhibitor, and endopeptidase Lys-C were purchased from Millipore, Miles Laboratories, Worthington, and Boehringer-Mannheim, respectively. Carboxypeptidase Y was a generous gift of Dr. M. Ottesen (Carlsberg Laboratories, Copenhagen), and lysylendopeptidase from *Achromobacter lyticus* was a gift from Wako Pure Chemicals, Japan. Trifluoroacetic acid, cyanogen bromide, and citraconic anhydride were purchased from Pierce Chemical Co. Iodoacetic acid (Sigma) was twice recrystallized from chloroform. Iodo[^{14}C]acetic acid was obtained from New England Nuclear, and acetonitrile was purchased from Burdick and Jackson.

Preparation of CM-aequorin. Aequorin was isolated from several hundred thousand individual jellyfish (*Aequorea victoria*) obtained in the late summer during collections conducted over a period of several years at the marine laboratories of the University of Washington, Friday Harbor, WA. Aequorin was purified by the procedures described by Blinks et al. (1978) except that a Sephadex G-75 (superfine) column was used in the second gel filtration step instead of the suggested Sephadex G-50 column.

S-Carboxymethylated and ^{14}C-labeled aequorin was prepared by incubating 75–420 nmol of protein in 0.55 mL of 6 M guanidine hydrochloride, 9.1 mM dithioerythritol, 0.2 M Tris, and 0.01 M EDTA, pH 8.2, under N_2 for 1–2 h at 37 °C. An aliquot from a freshly prepared solution of 116 mM iodo[^{14}C]acetic acid ($\sim$100 μCi) in 6 M guanidine hydrochloride, 0.2 M Tris, and 0.01 M EDTA, pH 8, was added to give a 2.9-fold molar ratio over total thiols in the solution. This mixture was incubated in the dark at room temperature for 30 min and quenched with 50 μL of 2-mercaptoethanol. In some cases, 15% SDS–PAGE (Laemmli, 1970) of [^{14}C]-CM-aequorin indicated low molecular weight contaminants, suggesting the possibility that limited proteolysis by contaminating proteases occurred during preparation. When such fragments were observed, they were removed by separating the alkylation reaction mixture on a Sephacryl S-200 column (1.6 × 90 cm) equilibrated in 6 M guanidine hydrochloride, 10 mM Tris, and 1 mM EDTA, pH 8, and eluted at a flow rate of 5.8 mL/h. The major protein fractions were pooled, dialyzed vs. 10 mM NH_4HCO_3, and lyophilized. When contaminating fragments were absent, the alkylation reaction mixture was dialyzed in the dark with 10 mM NH_4HCO_3 and lyophilized. The purity of preparations used for sequence analysis was greater than 95% as judged by SDS–PAGE.

To label all sulfhydryl groups which are not disulfide bonded, aequorin (60 nmol) was first alkylated with iodo[^{14}C]acetic acid (20 μCi), without reduction, in 1.4 mL of 6 M guanidine hydrochloride, 0.5 M Tris, and 0.01 M EDTA, pH 8.2, in the dark for 40 min at room temperature. For reduction, 10 mg of dithiothreitol was added, and the sample was incubated at room temperature for 1 h. Nonradioactive iodoacetic acid (27.5 mg in the 6 M guanidine hydrochloride buffer above) was added, and the sample was incubated at room temperature for 30 min in the dark. The reaction was quenched by the addition of 100 μL of 2-mercaptoethanol, dialyzed extensively against 10 mM NH_4HCO_3 in the dark, and lyophilized.

Specific Cleavage of [^{14}C]CM-aequorin. Cleavage at methionine residues (Gross, 1967) was performed by reacting the protein with cyanogen bromide (100-fold molar excess over methionine residues) in 70% (v/v) formic acid.

CM-aequorin was cleaved at lysines by incubating the protein with endopeptidase Lys-C in 50 mM NH_4HCO_3 and 4 M urea, pH 8, for 4–5 h at 37 °C using a protease:substrate weight ratio of 1:100. For the preparation of peptide K15, the protein was incubated with lysylendopeptidase from *Achromobacter lyticus* (1:100 protease:substrate) in 2 M urea and 50 mM Tris, pH 9, for 5 h at 37 °C.

For cleavage at arginine residues, CM-aequorin was N-citraconylated (Atassi & Habeeb, 1972) and digested with TPCK–trypsin. Citraconylation was performed over a period of 1.5 h at room temperature by adding 150 μL of citraconic anhydride in small aliquots to CM-aequorin (400 nmol) dissolved in 6 M guanidine hydrochloride. The pH of the reaction was maintained between 8.0 and 8.5 by the manual addition of small aliquots of 2 M NaOH. The citraconylated protein was separated on a Sephadex G-25 column in 100 mM NH_4HCO_3, pH 8.5, and the pooled product digested immediately with TPCK–trypsin. TPCK–trypsin was added in two aliquots over a 4-h period at 37 °C (final ratio 1:25 protease:substrate). Digestion was terminated by the addition of 1 μg of soybean trypsin inhibitor/μg of trypsin. Citraconyl groups were removed by incubating the digest in 10% (v/v) formic acid for 3.5 h at 37 °C. The digestion mixture was run on a Sephadex G-10 column in 10 mM NH_4HCO_3 and lyophilized.

Tryptic subdigestion of a large peptide, R7, employed the same trypsin digestion procedure described above. Peptide K15 was subdigested by cleavage at glutamyl bonds using

[1] Abbreviations: CM, carboxymethyl; HPLC, high-performance liquid chromatography; TPCK, N^{α}-(*p*-tosyl)-L-phenylalanine chloromethyl ketone; EDTA, ethylenediaminetetraacetic acid; Tris, tris(hydroxymethyl)aminomethane; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; TFA, trifluoroacetic acid; Pth, phenylthiohydantoin.

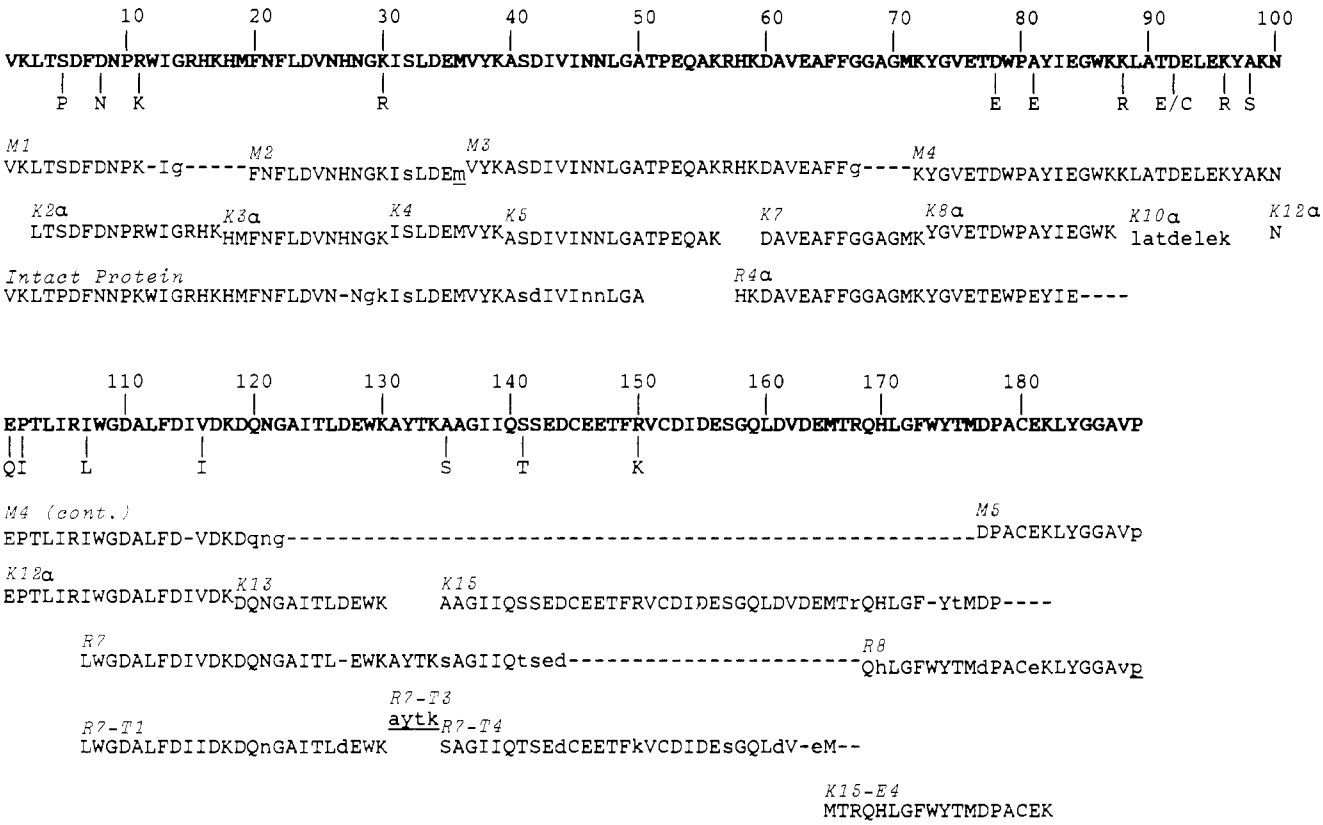

FIGURE 1: Summary of the proof of sequence of aequorin. The proven sequences of specific peptides (named in italics) are given in one-letter code below the summary sequence (bold type). The letters below the summary sequence give the minor variant residue(s) observed at each of 17 positions showing microheterogeneity. The prefixes M, K, and R denote peptides generated by cleavage of the CM protein at methionyl, lysyl, and arginyl bonds, respectively. Subpeptides are identified in hyphenated suffixes, with T and E designating enzymatic cleavage by trypsin or *S. aureus* protease, respectively. A Greek symbol following a peptide name indicates that it is one of a series of variant peptides arising from the microheterogeneity observed in aequorin (see Figure 2). Peptide sequences in capital letters were proven by automated Edman degradation while those given in lower case letters indicate that only a tentative identification could be made. A dash designates a residue within a peptide which could not be identified. Residues which are underlined were deduced only from compositional analysis.

staphylococcal V8 protease in 50 mM NH_4HCO_3 with a protease:substrate weight ratio of 1:40 at 37 °C for 8 h and then lyophilized.

Peptides were purified by reverse-phase HPLC using a Varian 5000 liquid chromatograph or, in some cases, a Waters Associates system (Model 720 system controller, Model 440 absorbance detector, and Model 6000 pumps). Unless otherwise stated, a buffer system (Mahoney & Hermodson, 1980) of 0.1% trifluoroacetic acid with an acetonitrile gradient was used with an Ultrapore RPSC-C3 column (Altex). Gel permeation HPLC was performed with TSK G-2000SW and G-3000SW (LKB) columns with buffer containing 6 M guanidine hydrochloride and 10 mM sodium phosphate at pH 6.

Automated Sequence Analysis and Amino Acid Analysis. The sequence analysis of intact protein or relatively large amounts of peptide (>10 nmol) was performed with a Model 890B Beckman sequencer, updated as described by Bhown et al. (1980), using 0.55 M Quadrol buffer and polybrene (Tarr et al., 1978). Pth derivatives of amino acids were identified by using reverse-phase HPLC on a Dupont Zorbax ODS column as described by Hunkapiller & Hood (1978). Peptides available at the 2–10-nmol level were sequenced on a Model 890C Beckman sequencer employing a program (Brauer et al., 1975) using 0.16 M Quadrol buffer and polybrene. Pth-amino acids were identified by using the reverse-phase HPLC system described by Ericsson et al. (1977). An Applied Biosystems Model 470A gas phase sequencer (Hunkapiller et al., 1983) was used for sequence analysis when there was less than 1.5 nmol of peptide available. Pth-amino acids from the gas phase instrument were identified by using an IBM cyano column as described by Hunkapiller & Hood (1983). Two peptides, M3 and M5, which were small and appeared to wash out of the cup with Quadrol buffer programs, were sequenced on the Model 890B Beckman sequencer using a program adapted for use with dimethylallylamine buffer and polybrene as suggested by Klapper et al. (1978).

Amino acid analyses were performed on a Dionex Model D-500 or a Beckman Model 6300 amino acid analyzer.

Analysis of Sequence Homologies. A search of the National Biomedical Research Foundation data bank for homologous protein sequences and the optimal alignment of related sequences was done on a VAX/VMS computer using the ALIGN and SEARCH programs described by Dayhoff et al. (1983). The application of these programs in our laboratories for the analyses of related proteins has been described in more detail by Reimann et al. (1984).

RESULTS

General Strategy. Much of the sequence of aequorin (66%) was identified from the sequence analysis of the intact CM protein and of five nonoverlapping cyanogen bromide fragments. The overlaps and the remainder of the sequence were provided by analysis of selected peptides obtained by cleavage at either arginine or lysine. A detailed summary of the proof of sequence is given in Figure 1.

During the course of our sequence studies, it became clear that our preparations, although homogeneous by molecular weight analysis, were actually comprised of a minimum of three isotypes of aequorin. These isotypes were revealed when

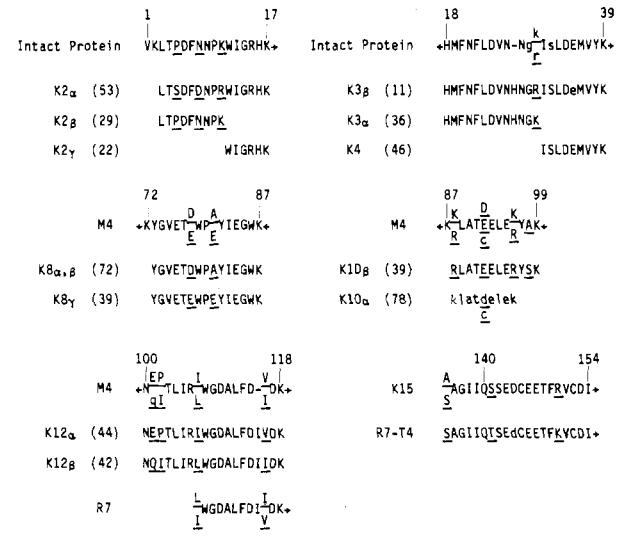

FIGURE 2: Supporting evidence for microheterogeneity in six regions of aequorin. For each region, the upper line denotes residues observed after direct Edman degradation of the protein or a large fragment. Below these are sequences of smaller variant peptides which provide support for the identification of sites of microheterogeneity. Variant residues are underlined. The yields of variant peptides are shown in parentheses after the peptide name. Peptides are named as described in the legend to Figure 1. The one-letter code denotes residues unequivocally identified by automated Edman degradation. When two residues were found in one degradation cycle, the upper one was judged to be the most abundant in that cycle. Arrows indicate that only a portion of the sequence is given and that the sequence continues in the direction of the arrow. Lower case letters indicate tentative identifications. Lys-87 (in italics) was observed as an additional residue at the amino terminus of a minor peptide (approximately 20%) contaminating K10α.

two or more different amino acids were unequivocally identified at the same position in the sequence. Seventeen variant loci are indicated in Figure 1 where the minor variant amino acid in each case is shown beneath the major (boldface) variant. Detailed data supporting the assignments at these heterogeneous positions are shown in Figure 2 and are discussed in greater detail below.

Cleavage at Methionyl Bonds. The products of cyanogen bromide treatment of CM-aequorin were dissolved in 6 M guanidine hydrochloride and separated directly by reverse-phase HPLC chromatography as shown in Figure 3. The amino acid compositions of these peptides and that of the intact protein (Table I) demonstrate that these five fragments constitute a complete set of unique nonoverlapping peptides. With the exception of the large fragment, M4, all peptides were obtained in yields ranging from 57% to 61% (Table I). The large size of M4 (M_r 12 130) may account for the lower yield (42%) and its elution at high acetonitrile concentration. As suggested by the presence of methionine in the amino acid composition, M4 contains an internal Met–Thr bond (residue 165) which was resistant to cleavage. The absence of significant quantities of a peptide representing residues 166–176 suggests that little or no Met–Thr cleavage occurred under our experimental conditions.

The M1 fraction (Figure 3) appeared to contain significant contaminants as shown by the amino acid composition (Table I) which gave high values for Gly and Ala. Peptide M1 required further purification by HPLC chromatography on an Ultrapore RPSC column using a buffer system of higher pH (solvent A, 10 mM potassium phosphate, pH 6; solvent B, 50% acetonitrile and 5 mM potassium phosphate, pH 6). Elution with a linear gradient of solvent B at a flow rate of 1.0 mL/min produced three peaks, two of which eluted at 36%

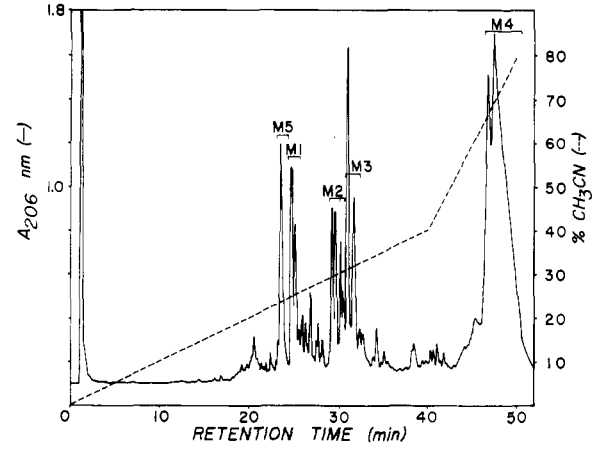

FIGURE 3: HPLC separation of a cyanogen bromide digest of [^{14}C]CM-aequorin (19.8 nmol). Separation was performed on an Altex Ultrapore RPSC-C3 column (4.6 × 75 mm) equilibrated in 0.1% TFA and eluted with a linear gradient of acetonitrile at 2.0 mL/min at room temperature. The lyophilized digest was dissolved in 6 M guanidine hydrochloride and injected directly into the column. The identity of purified peptides is shown by using the prefix M as described in Figure 1.

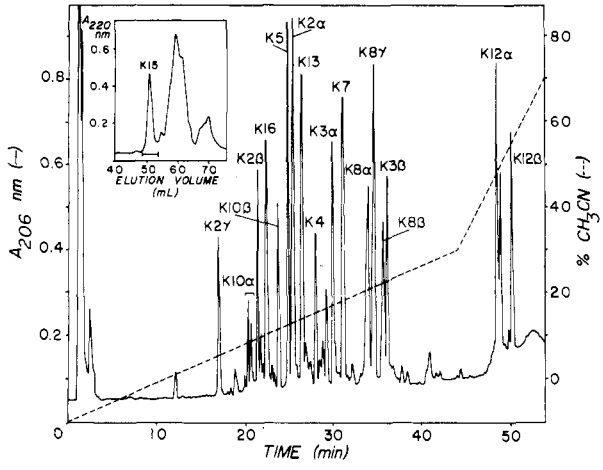

FIGURE 4: HPLC separation of peptides produced by the cleavage of [^{14}C]CM-aequorin (9.8 nmol) with endoproteinase Lys-C. Chromatography conditions were as in Figure 3 but with a flow rate of 1.0 mL/min. The digest in 4 M urea was injected directly into the column. The inset shows the HPLC separation of a large lysyl peptide, K15, after a separate digestion of CM-aequorin (12 nmol) with *Achromobacter* Lys-cleaving enzyme. K15 was isolated by using gel filtration on a series of three TSK columns (LKB) linked in the order TSK-G3000SW, G2000SW, G2000SW. The columns were eluted with 6 M guanidine hydrochloride containing 10 mM sodium phosphate buffer (pH 6.0) at a flow rate of 0.5 mL/min.

and 39% solvent B and contained M1 as shown by sequence and amino acid composition. The overall yield of M1 was less than 10%.

Cleavage at Lysyl Bonds. [^{14}C]CM-aequorin was cleaved specifically at lysyl bonds by using endopeptidase Lys-C. The resulting fragments were completely resolved and obtained in good yield by reverse-phase HPLC chromatography as shown in Figure 4. As Table I shows, the set of fragments purified by this procedure was nearly complete. The small peptides K1, K6, K9, K11, and K14 were not needed for sequence analysis and were purposely disregarded. Peptide K15 is large and acidic (pI = 3.8, M_r 5470) and was not recovered by this separation procedure due to its insolubility at the pH of the buffer system (pH 2–3). Peptide K15 was recovered with a yield of 60% from a digestion of [^{14}C]CM protein with the lysyl endopeptidase of *Achromobacter*. As the largest lysyl peptide, K15 was easily separated by using gel permeation

Table I. Amino Acid Composition[a] of Peptides Isolated After Cleavage with Endopeptidase LysC[b] and CNBr[b]

Peptide		K1	K2$_\alpha$	K3$_\alpha$	K4	K5	K6	K7	K8$_\alpha$	K9	K10$_\alpha$	K11	K12$_\alpha$	K13	K14	K15	K16	Whole Protein		M1	M2[c]	M3	M4[c]	M5
Residues		1-2	3-17	18-30	31-39	40-56	57-59	60-72	73-87	88	89-96	97-99	100-118	119-130	131-134	135-182	183-189	from Sequence	from Hydrolysis[d]	1-19	20-36	37-71	72-176	177-189
Asp/Asn	(D/N)		2.9(3)	3.8(4)	1.1(1)	2.6(3)		1.0(1)	1.0(1)		1.0(1)		4.0(4)	2.5(3)		6.6(6)		(27)	27.0	2.4(3)	5.0(5)	3.9(4)	12.9(14)	1.0(1)
Thr	(T)		0.9(1)			0.9(1)			0.7(1)		0.7(1)		1.4(1)	0.8(1)	(1)	2.8(3)		(10)	10.1	0.8(1)		1.0(1)	6.9(8)	
Ser	(S)		0.9(1)	0.7	0.8(1)	0.9(1)							0.7			2.5(3)		(6)	8.5	0.8(1)	1.0(1)	1.2(1)	3.6(3)	
Glu/Gln	(E/Q)		0.8	0.4	1.0(1)	2.5(2)		1.5(1)	2.1(2)		2.2(2)		3.4(1)	2.4(2)		8.6(9)	0.6	(20)	24.1	1.5	1.1(1)	3.2(3)	16.6(15)	1.0(1)
Pro	(P)		0.9(1)			1.5(1)			0.8(1)				1.2(1)			1.5(1)	1.2(1)	(6)	5.9	1.7(1)		1.5(1)	1.9(2)	1.9(2)
Gly	(G)		1.2(1)	1.5(1)		1.3(1)		3.1(3)	1.8(2)		0.5		2.0(1)	1.2(1)		3.6(3)	2.2(2)	(15)	15.9	1.7(1)	1.2(1)	4.2(4)	8.9(7)	2.1(2)
Ala	(A)		0.7	0.7		3.1(3)		3.3(3)	0.9(1)		0.9(1)	(1)	1.6(1)	1.4(1)	(1)	3.3(3)	1.2(1)	(16)	13.7	0.8		5.1(6)	7.2(8)	1.9(2)
CMCys	(C)															2.5(3)		(3)	2.9				1.8(2)	1.0(1)
Val	(V)	(1)	0.4	1.2(1)	1.0(1)	1.0(1)		1.1(1)	0.9(1)				1.1(1)			2.2(2)	1.1(1)	(10)	9.4	1.3(1)	1.1(1)	3.0(3)	4.8(4)	1.2(1)
Met	(M)			0.5(1)	0.7(1)			0.7(1)					0.6			1.5(2)		(5)	3.9				0.4(1)	
Ile	(I)		1.0(1)	0.4	0.9(1)	1.6(2)			0.7(1)				2.4(3)	0.9(1)		2.3(3)		(12)	11.6	0.8(1)	1.0	1.7(2)	7.6(8)	
Leu	(L)		0.9(1)	1.0(1)	0.9(1)	1.0(1)					1.2(2)		2.1(2)	1.0(1)		2.0(2)	1.0(1)	(12)	13.1	1.1(1)	2.1(2)	1.5(1)	7.4(7)	1.0(1)
Tyr	(Y)			0.5	0.7(1)				1.2(2)			(1)			(1)	0.7(1)	0.8(1)	(7)	6.4			0.6(1)	4.3(5)	1.0(1)
Phe	(F)		0.8(1)	1.5(2)				1.4(2)					0.9(1)			1.7(2)		(8)	7.9	0.6(1)	2.1(2)	1.9(2)	2.9(3)	
His	(H)		0.9(1)	1.7(2)			(1)									0.8(1)		(5)	4.8	1.3(2)	0.8(1)	1.2(1)	1.0(1)	
Lys	(K)	(1)	0.7(1)	0.9(1)	0.8(1)	1.1(1)	(1)	1.0(1)	1.3(1)	(1)	1.0(1)	(1)	1.1(1)	0.9(1)	(1)	1.3(1)		(15)	15.1	2.0(2)	0.7(1)	2.7(3)	7.8(8)	1.0(1)
Arg	(R)		1.7(2)	0.7			(1)						1.0(1)			2.0(2)		(6)	7.1	1.3(2)	0.6	1.2(1)	3.9(3)	
Trp[e]	(W)		(1)						(2)				(1)	(1)		(1)		(6)		(1)			(5)	
Hser[f]																				+ (1)	+ (1)	+ (1)	+ (1)	
Total Residues		2	15	13	9	17	3	13	42	1	8	3	11	12	4	48	7	(189)		19	17	35	105	13
Yield %		-	53	36	46	84	-	69	42	-	78	-	44	81	-	60	77			58	57	60	42	61

[a] Residues per peptide from amino acid analysis (6 N HCl, 110°C, 24 h). Values of 0.3 or less are not reported. Numbers in parentheses are derived from the sequence (Figure 1).

[b] The CNBr peptides (M1, 3, and 5) and LysC peptides (K2$_\alpha$-K16) were isolated by HPLC chromotography as shown in Figure 3 and Figure 4.

[c] The samples used for analysis of peptides M2 and M4 were isolated by G-50 gel filtration followed by HPLC chromatography, while the yields reported for these two peptides are those obtained using the methods shown in Figure 3.

[d] From the hydrolysis of S-carboxymethylated aequorin for 24, 48, and 72 hr. Thr and Ser values extrapolated to t = 0. Val and Ile values were from the 72 hr analysis.

[e] Tryptophan was not determined.

[f] Homoserine and homoserine lactone values for CNBr peptides were not quantitated; however, a (+) symbol is used to indicate that these residues were observed in good yield.

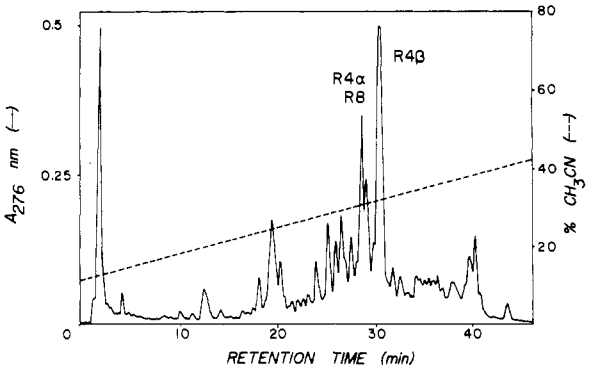

FIGURE 5: Separation of selected peptides after cleavage at arginyl bonds. Following the tryptic digestion of citraconylated aequorin (323 nmol), the citraconyl groups were removed, and the fraction soluble in 0.1% TFA was separated on an Ultrapore RPSC-C3 column (4.6 × 75 mm) at a flow rate of 1.0 mL/min using 0.1% TFA–acetonitrile buffers as described in Figure 3.

HPLC in 6 M guanidine hydrochloride (inset, Figure 4).

In addition to those lysyl peptides used in the proof of structure (Figure 1), several additional distinct peptides, arising from sequence microheterogeneity, were isolated by HPLC (Figure 4) and are denoted as a structurally related set by the use of suffixed Greek symbols (e.g., K8α, K8β, etc.).

Cleavage at Arginyl Bonds. N-Citraconylated [^{14}C]CM-aequorin was specifically cleaved at arginine residues with trypsin. When the decitraconylated trypsin digest was suspended in 0.1% TFA, centrifugation produced a pellet which was comprised primarily of one major peptide with a molecular weight of 8500 as shown by SDS–PAGE. This peptide, R7, was further purified by gel filtration in 10 mM NH_4HCO_3 on a Sephadex G-50 (superfine) column (1.0 × 118 cm) at a flow rate of 8.0 mL/h. Peptide R7 was located in the major peak (A_{230nm}) which eluted between 43 and 47 mL.

The peptides that were soluble in 0.1% TFA were separated by reverse-phase HPLC as shown in Figure 5. Although several other peptides representing different regions of the sequence were isolated, Figure 5 identifies only those peptides used in establishing the sequence (Figure 1). Peptides R4α and R8 coeluted and were separated on a μBondapak phenyl column (Waters) with a TFA–acetonitrile gradient (data not shown). Each eluted in two peaks that showed no significant differences in their amino acid compositions; when combined, they showed no heterogeneity upon sequence analysis. Since R4α and R8 both contain tryptophan, it is likely that partial oxidation of tryptophan yielded two forms of each peptide differing in chromatographic properties. The low yields (5%) for both R4α and R8 can be attributed to the use of two HPLC separations for purification.

Subdigestion of R7 and K15. Peptide R7 (residues 107–167) was subdigested with trypsin to identify its C-terminal residues. The tryptic subdigest was separated into soluble and insoluble fractions in 0.1% TFA. The acid-insoluble fraction was dissolved in phosphate buffer, pH 6.8, and separated by HPLC on an Ultrapore RPSC-C3 column using a phosphate buffer system (solvent A, 10 mM potassium phosphate, pH 6.8; solvent B, 50% acetonitrile and 5 mM potassium phosphate, pH 6.8) at a flow rate of 1.0 mL/min. The acid-soluble fraction was separated by using the TFA–acetonitrile buffer system on the same column. The tetrapeptide R7-T3 was acid soluble and eluted just after the column void volume. R7-T1 (residues 107–130) was recovered in both the soluble and insoluble fractions and eluted as two major peaks in the TFA–acetonitrile system (data not shown). Amino acid composition data suggested that these two peptides were nearly identical except for differences resulting from microheterogeneity at positions 107 and 116 (see Figures 1 and 2). Peptide R7-T4 in the acid-insoluble fraction also eluted as several peaks which showed no difference in amino acid composition and no heterogeneity by sequence analysis of the combined fractions.

The staphylococcal V8 protease digest of K15 was resolved into four major peaks on an RPSC-C3 column using a 0.1% TFA–acetonitrile gradient (1.3% acetonitrile/min) at a flow rate of 2.0 mL/min. K15-E4, used for sequence determination, eluted last at approximately 25% acetonitrile.

Sequence Analysis. Edman degradation of the intact CM protein proceeded for 50 residues, but 7 identifications were tentative, and 1 residue was not identified at position 27 (Figure 1). This amino-terminal sequence agrees completely with the 15-residue sequence reported earlier by Prendergast & Mann (1978). This sequence included three methionines which aligned peptides M1, M2, and M3. The sequences of M2, M3, and K4 completed this region. The remaining C-terminal residues of M3 were provided by the overlapping Lys-C peptide K7 (residues 60–72). The sequence of K7 ended with Met-Lys, indicating that M4 was the contiguous fragment. Peptide R4α provided the confirming overlap, and the M4 sequence extended the structure to Asp-114.

Sequence analysis of the large arginyl peptide R7 overlapped M4 by 16 residues, identifying position 115 as Ile and extending the sequence to Leu-126. With K13 and the remaining sequence of R7, the sequence extended to Gln-140.

The sequence of the Lys-C peptide K15 completed the unidentified sequence of M4, overlapped R7, and extended the sequence to Phe-172 except for a weak identification of Arg-167. R8 provided an overlap of K15 to M5, the only product of CNBr cleavage that did not contain homoserine and thus the putative C-terminal fragment. The combination of K15, R8, and M5 provided a complete protein sequence with the exception of weak identifications of Arg-167 and of Pro at the C-terminus.

Arg-167 was confirmed by the sequence of the staphylococcal protease V8 subdigestion product of K15, K15-E4. The occurrence of proline at the C-terminus is consistent with the composition of M5 and its proven sequence. Furthermore, the C-terminus of R8 was also tentatively identified as Ala-Val-Pro on the basis of sequence data and amino acid composition. Confirmation of the C-terminal Gly-Gly-Ala-Val-Pro sequence was obtained by digesting 4.0 nmol of the intact CM protein with a 1:100 molar ratio of carboxypeptidase Y to substrate in 0.1% SDS and 20 mM sodium acetate, pH 4.9, for 1 h at 25 °C. This released the following equivalents of amino acids: Pro, 1.3; Val, 1.0; Ala, 1.1; Gly, 1.8; Tyr, 1.2. Similar results were obtained when the C-terminal peptide M5 and R8 were digested under similar conditions but in the absence of SDS.

The sequence data described above provide an unambiguous alignment of the various fragments within the 189-residue chain of aequorin. However, more than 1 residue was seen at 17 different loci in this proof of sequence, suggesting an unusual level of sequence microheterogeneity.

Proof of Microheterogeneity. Figure 2 details the sequence data which provide the evidence for sequence microheterogeneity and give a measure of the relative yields of variant peptides. Evidence for microheterogeneity was first noted during the sequence analysis of the whole protein and its large fragments (e.g., M4, R-7), when in an otherwise homogeneous sample a single cycle of the Edman degradation produced substantial quantities of two different Pth-amino acids. Further evidence for this microheterogeneity was obtained

during the analysis of variant fragments isolated after cleavage at lysine. As illustrated in Figure 2, pairs of peptides were isolated in good yield which were identical at all positions except those displaying sequence microheterogeneity. In some cases, where a lysine residue was in the variant, an additional peptide (e.g., K2α or K4) was obtained in good yield. Both variant amino acids were usually observed in a single Edman degradation, but in some cases (e.g., residues 5, 8, 11, 98, and 141), only one was observed. In the latter cases, the dual assignments are supported by the observation of each of the variants in separate peptides (Figure 2). Due to their characteristically low yield, it is likely that Pth-Ser or Thr could be missed if it were the minor of two residues present in a single cycle.

Sequence analysis of K10α revealed that this peptide fraction comprised a mixture of two species. The minor species (about 20% of the total) was one residue longer than the major species and had a Lys residue at the amino terminus. This suggests that cleavage after the second Lys of the Lys–Lys bond at positions 87–88 may not have been complete. The identification of Pth-Asp in cycle 4 from the Edman degradation of peptide K10α was supported by the yield of 1.0 equiv of Asp from amino acid hydrolysis.

The presence of Cys at position 92 is supported only by the detection of ^{14}C radioactivity from [^{14}C]Pth-CM-Cys; the Pth derivative itself was not directly identified at position 92. Sequence analysis of M4 and R4β revealed the presence of significant ^{14}C radioactivity in the cycle corresponding to position 92. R4β was radioactive, and amino acid analysis revealed the presence of 0.7 equiv of CM-Cys. The amino acid composition of R4β suggests that this peptide is derived from residues 58–96. Further evidence for this Cys comes from the sequence of the K10α mixture. Radioactivity from [^{14}C]-CM-Cys was detected in cycle 5 along with a strong yield of Pth-Glu, suggesting that the minor peptide starting at Lys had Cys at position 92, while the Glu was assigned as position 93 from the shorter, major species. The amount of radioactivity was consistent with the Cys being derived from the minor peptide as opposed to the major.

Sequence analysis of R7-T4 showed only a weak identification of Lys at position 150. This identification was considered tentative since the yield of Pth-Lys at that cycle was about 50% of that expected. The amino acid composition of R7-T4 suggests that this position should be Arg since 1.8 and 0.3 equiv of Arg and Lys, respectively, were observed. Thus, it appears that an arginine at position 150 in R7-T4 may have been missed. The sequence of K15 clearly showed Arg at position 150. The Arg/Lys replacements at position 150 must be considered a relatively weak identification of microheterogeneity since it rests mainly on the observation of a weak signal for Lys obtained with only one peptide, R7-T4.

The assignment of the major and minor amino acid replacements at positions displaying microheterogeneity was difficult. The major amino acid at these sites appeared to depend on the preparation of aequorin used, suggesting that the ratios of isotypes varied from one preparation to another. Since the lysyl peptides and M4 were isolated from the same preparation of aequorin, the major variants observed from the sequence of M4 were generally consistent with those expected from the yields of the lysyl peptides (Figure 2). Thus, the variant identified as the major sequence in Figure 1 is somewhat arbitrarily chosen and may not represent the actual sequence of any aequorin isotype.

Discussion

As isolated, apoaequorin has a free amino terminus and is comprised of 189 residues. The molecular weight for the apoprotein sequence (Figure 1) is 21 459 while that for the holoenzyme with chromophore plus oxygen would be 21 914. These values are in reasonable agreement with the molecular weight of 19 500 determined by Prendergast & Mann (1978).

The structure of aequorin was deduced from a complete set of cyanogen bromide fragments which were overlapped by the amino-terminal sequence of the intact protein, selected arginyl peptides (R4α, R7, and R8), and the lysyl peptide K15. The nearly complete set of peptides derived from cleavage at lysine was used to confirm tentative identifications and to provide evidence for sequence microheterogeneity. Ninety-three percent of the residues were identified on two or more peptides, and all overlaps essential to the proof included 10 or more residues. Three of the weakest assignments in the sequence are Arg-57, Arg-150, and Asp-163 which were observed in only one peptide; however, the assignments of Arg-57 and Arg-150 are supported by amino acid composition data and specificity of enzymatic cleavage.

Several peptide bonds in aequorin were found to be resistant to chemical and enzymatic cleavage. The Met–Thr bond between residues 165 and 166 was not cleaved by cyanogen bromide in 24 h. The Lys–Asp bond of the Asp-Lys-Asp sequence (residues 117–119) and the Arg/Lys–Val bond (150–151) were not cleaved by trypsin in the subdigestion of R7. In each case, neighboring acidic residues probably account for the resistance of the bonds to trypsin cleavage.

The major isotypes of aequorin have three cysteine residues, but a minor isotype containing a fourth cysteine at position 92 was detected. Carboxymethylation studies without prior reduction showed that all four Cys residues could be alkylated in the absence of reducing agent and that there are no disulfide bonds in aequorin. Kemple et al. (1984) and Shimomura & Johnson (1978) have reported the presence of one reactive sulfhydryl which is essential for bioluminescence, but its location is not known. In this regard, it is interesting to note that Cys-145 has a negatively charged microenvironment that could give its sulfhydryl group an unusual pK_a and selective reactivity toward charged species.

Sequence Microheterogeneity. Although the aequorin preparations used in our studies were homogeneous with respect to molecular weight (as shown by SDS–PAGE), sequence analysis revealed the presence of 17 sites of microheterogeneity (Figures 1 and 2). One minor isotype was indicated by the trace of cysteine at residue 92, but the ratios at other replacement sites were as high as 1:1 (Figure 2). Obviously, with 17 sites of microheterogeneity, the potential number of isotypes is enormous. Interestingly, many of the amino acid replacements appear to be linked with one another (Figure 2). For example, if Asp is the variant at position 78, then Ala is the variant at position 81 (see peptides K8α and K8β). This linkage of variants reduces the number of isotypes which must be present to account for this microheterogeneity. Nonetheless, this protein represents an unusually complex mixture of closely related isotypes.

The weakest evidence involves the Arg/Lys replacement at position 150, but the remaining assignments are made with confidence. Of course, the inherent limitations of our sequencing procedures and Pth-amino acid detection methods allow for additional undetected replacements at some of the 17 sites or elsewhere. It should be noted that our studies produced no evidence for isotypes with insertions or gaps in the molecule. The carboxypeptidase Y digest of CM protein gave no evidence of heterogeneity at the C-terminus. However, amino-terminal analysis of the intact protein indicated the

presence of some sequences (5–20%) beginning with leucine at position 3. It is not clear whether these shorter molecules represent real isotypes or simply artifacts of proteolytic nicking during purification.

Ten of the amino acid replacements are of a conservative nature. The most radical changes are the Ser/Pro and Pro/Ile changes at positions 5 and 102, respectively. Lys/Arg replacements were seen at five sites; Asp/Glu and Ala/Ser replacements occurred at two positions each. Four replacements involve changes in charge; these include the amide changes at positions 8 and 101, the Ala/Glu exchange at position 81, and the Cys replacement at position 92. These replacements could account for up to five electrophoretically distinct species. However, it should be noted that the vast majority of the potential isotypes resulting from the 17 replacements would be electrophoretically neutral. All of the replacements except Pro/Ile and Asp or Glu/Cys can be accounted for by a single base change. It is interesting to note from Figure 1 that some of the sites of microheterogeneity are clustered; e.g., there are nine sites within the region encompassing residues 78–110.

The presence of isotypes with a large number of sites of variance is relatively unusual and might be explained in several ways: (1) the presence of multiple genes coding for aequorin, (2) the presence in the population of multiple alleles at one locus, (3) both (1) and (2), or (4) the presence of closely related photoproteins in our preparations due to the collection of related species with similar morphology. We cannot eliminate the latter possibility even though efforts were made to collect only *Aequorea victoria*. The sequences of peptides K12α and K12β (Figure 2) prove that one pair of isotypes possesses a minimum of four replacements in one chain. Generally, alleles of a single locus do not differ at more than two sites (Doolittle, 1979; Gutman et al., 1975); hence, multiple alleles are unlikely. Thus, we assume that the most likely explanation involves multiple genes.

Multiple isotypes have been identified in purified samples of aequorin (Blinks et al., 1976; Prendergast & Mann, 1978) and another photoprotein, mnemiopsin (Ward & Seliger, 1974). Twelve luminescent bands with p*I*'s ranging between 4.2 and 4.9 have been observed in crude extracts of *Aequorea* whereas extracts prepared from a single organism contain at least eight electrophoretically distinct isotypes (Blinks et al., 1976). The presence of eight active forms of aequorin in a single individual suggests the presence of multiple aequorin genes. However, even though the eight species of aequorin were all active, there is a possibility that some of these forms could have been produced artifactually by proteolysis during extraction.

It seems unlikely that the amino acid replacements identified in our aequorin preparations produce inactive protein species, since up to 12 luminescent isotypes have been observed (Blinks et al., 1976). However, the multiple forms identified at the sequence level may suggest the presence of aequorin isotypes showing different quantum yields of luminescence or distinctive rates of decay. Ward & Seliger (1974) have described variants of the photoprotein mnemiopsin which display distinct decay kinetics. In fact, it may be an adaptive advantage for the organism to possess multiple genes coding for aequorin molecules with different kinetic properties which could be differentially expressed during development or in response to environmental changes.

A cDNA coding for aequorin has been isolated by Prasher et al. (1985) using a synthetic oligonucleotide probe based on our sequence data. The preliminary nucleotide sequence of this cDNA clone (M. J. Cormier and D. Prasher, unpublished results) provides the protein sequence of a single isotype of apoaequorin. The sequence obtained from the cDNA shows generally good agreement with our protein sequence and is fully consistent with the sequence microheterogeneity reported here. The sequence obtained from the cDNA most closely resembles the minor sequence shown in Figure 1 since the minor variant was observed at 14 of the 17 sites of microheterogeneity. At two positions, 91 and 125, the cDNA sequence predicts Ser whereas Thr was observed in the protein sequence. These discrepancies can be explained by assuming sequence microheterogeneity which was not detected due to inherent difficulties in detecting minor amounts of Ser. At position 95, Lys was obtained from the cDNA data, and we observed Glu; this discrepancy may be due to further sequence microheterogeneity or to errors in the cDNA sequence.

The coding sequence from the cDNA clone has seven additional residues at the amino terminus (MTGEQYS) which were not observed in our protein sequence. It is not known whether this sequence is removed during cellular processing of the aequorin molecule or by endogenous protease action during purification. The observation of amino-terminal heterogeneity in our samples favors the latter possibility.

While this paper was being reviewed, the cloning and cDNA sequence analysis for aequorin were reported by Inouye et al. (1985). The amino acid sequence predicted from the DNA sequence is in complete agreement with the major sequence shown in Figure 1.

EF-Hand Domains. Intracellular high-affinity Ca(II) binding proteins such as calmodulin, troponin C, parvalbumin, and myosin light chains belong to a superfamily of proteins known as EF-hand proteins (Tufty & Kretsinger, 1975; Barker et al., 1978; Kretsinger, 1980). These proteins have either two or four Ca(II) binding regions which show sequence homology with the well-defined helix–loop–helix structures denoted as EF hands in the X-ray structure of carp parvalbumin (Kretsinger & Nockolds, 1973). The Ca(II) dependency of aequorin luminescence indicates the presence of three high-affinity Ca(II) sites, making aequorin a candidate for membership in the EF-hand superfamily.

When evaluated with the ALIGN program (Dayhoff et al., 1983), aequorin shows significant homology with all available sequences from the four-domain EF-hand proteins troponin C and calmodulin. Although the level of sequence identity was about the same throughout (23–26%), the alignment scores were higher with the calmodulins (6.1–7.0) than with the troponin C's (3.3–4.6). Aequorin was not homologous with myosin light chains except possibly with the G1 regulatory light chain of chicken smooth muscle (15% identity, alignment score of 3.3). Figure 6 illustrates the alignment of aequorin with mammalian calmodulin.

Inspection of the alignment of aequorin with the homologous Ca(II) binding proteins calmodulin and troponin C revealed that three segments of aequorin, A (residues 13–43), C (residues 106–136), and D (residues 142–172), have properties expected for EF-hand domains. Figure 7 illustrates the relative position of these segments in the aequorin molecule whereas Figure 6 demonstrates their alignment with the Ca(II) binding loops from domains I, III, and IV of calmodulin. These segments in aequorin satisfy most of the criteria established for EF-hand domains by Tufty & Kretsinger (1975). For example, all 3 segments have the invariant Gly and Glu residues in the 6th and 12th positions of the Ca(II) binding loop as well as the hydrophobic residue in the 8th position. As Figure 7 shows, the side chains of the residues in the positions

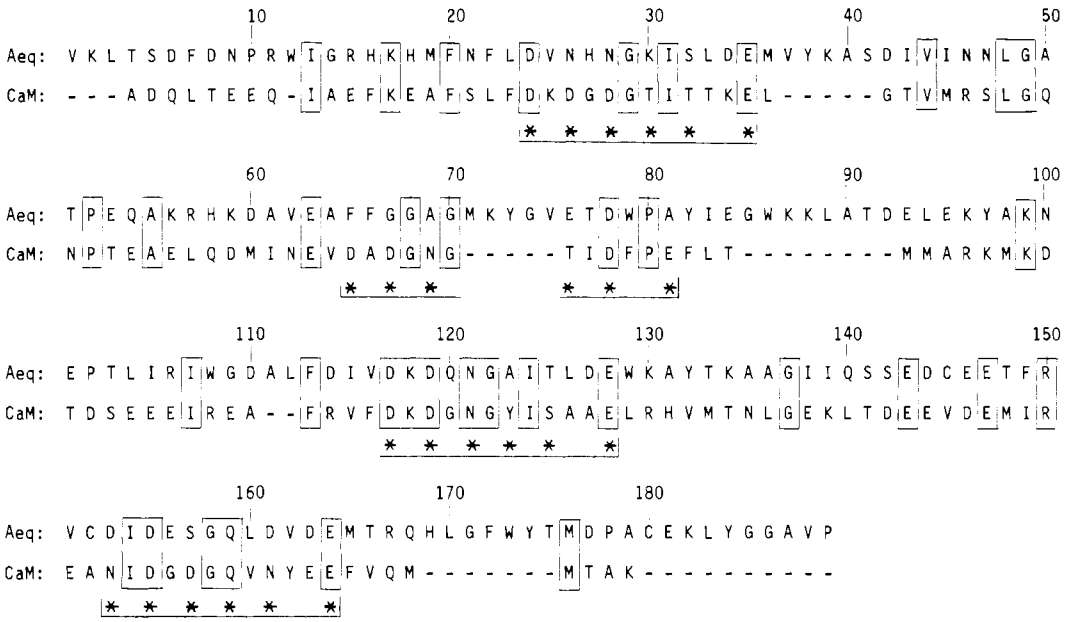

FIGURE 6: Sequence homology between aequorin and mammalian calmodulin. The sequence alignment shown was produced by the ALIGN program of Dayhoff et al. (1983) using a gap penalty of 10. Identical residues are enclosed in boxes. Asterisks indicate the positions of the six putative Ca(II) binding residues in the EF-hand domains of calmodulin.

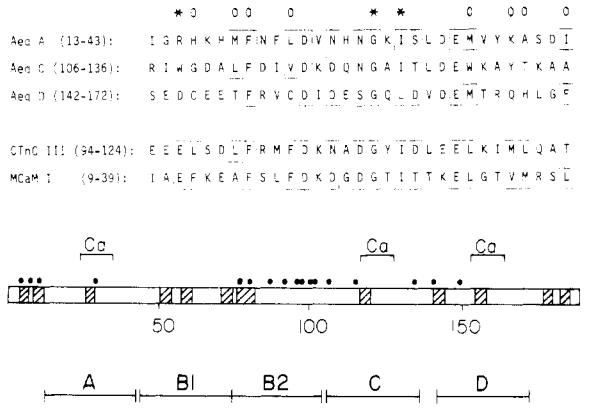

FIGURE 7: Proposed EF-hand domains of aequorin. (Upper panel) The sequences of three segments of aequorin (A, C, and D) are compared with representative EF-hand domains of bovine cardiac troponin C domain III (CTnC III) and mammalian calmodulin domain I (MCaM I). The vertical lines indicate the positions of the six residues providing the oxygen ligands for the octahedral coordination of calcium as observed in parvalbumin. The bulky hydrophobic residues in the flanking helical segments are designated by the symbol ϕ. Asterisks indicate the positions where residues are highly conserved, namely, glutamate (lower left) in a helical region, an invariant glycine in the loop, and a hydrophobic residue preceding the fifth ligand position. Residues enclosed in boxes fulfill 1 of the 16 criteria of an EF-hand domain set forth by Tufty & Kretsinger (1975). (Lower panel) Diagram of the aequorin sequence showing the relative position and location of the 31-residue, putative Ca(II) binding EF-hand domains, segments A, C, and D. Domains B1 and B2 each contain two 31-residue segments of rather different sequence. The hatched boxes show β-turn structures which were predicted by the Chou and Fasman procedure (Chou & Fasman, 1978; Argos et al., 1978). The solid dots above the bar indicate sites of microheterogeneity.

of the Ca(II) ligands all contain oxygen. However, the bulky hydrophobic residues usually observed in the helices on either side of the Ca(II) binding loops are missing in some positions, particularly in segment C (Figure 7). It is noteworthy that, with the exception of Lys/Arg at position 30, no sequence microheterogeneity was observed within the predicted Ca(II) binding loops. Generally, secondary structure predictions show a high probability of a β-turn segment in the first residue of the Ca(II) binding loops of EF-hand domains (Argos, 1977),

Table II: Internal Homology[a] of Aequorin

domains compared[b]	aequorin A (13–43)	aequorin C (106–136)	aequorin D (142–172)	aequorin B1 (44–74)	aequorin B2 (75–105)
aequorin A		5.9 (8)[c]	4.4 (6)	−1.0 (1)	0.0 (1)
aequorin C			3.5 (6)	−0.9 (5)	−0.6 (3)
aequorin D				−1.9 (2)	0.7 (6)
aequorin B1					0.6 (4)

[a] Homology is represented by alignment scores obtained from the ALIGN program of Dayhoff et al. (1983). Scores ≥3.0 are underlined; the probability of these scores occurring by chance is ≤0.0014. [b] Aequorin domains are defined diagrammatically in Figure 7. [c] Values in parentheses are the number of sequence identities observed.

and this appears to be the case in all three aequorin segments (see Figure 7).

As expected, the sequence of aequorin shows internal homology at the three EF-hand segments (Table II) with the greatest similarity between segments A and C. The remaining segments, B1 and B2, show little sequence identity to A, C, or D or to each other (Table II). On the basis of the data outlined above, we conclude that aequorin is a member of the EF-hand superfamily of proteins and that it has three, internally homologous, 31-residue segments (A, C, and D) which form EF hands.

Since segments A, C, and D appear to function as Ca(II) binding sites, it seems likely that segments B1 and B2 (residues 44–105) could provide the binding sites for luciferin and oxygen and could contribute at least some of the side chains required for catalysis of luminescence. Segment B of aequorin may be derived from the second of the usual four Ca(II) binding domains by evolution of a luciferin and/or oxygen binding domain and the insertion of approximately 35 residues. This hypothesis is supported by the finding of homology between aequorin and the four-domain proteins calmodulin and troponin C. Of the four Ca(II) binding sites in calmodulin and troponin C, the first and third domains are most closely related, as are the second and fourth domains. If aequorin evolved from such a precursor by divergence of the second domain, then segments A and C would be expected to be more closely related to each other than to segment D, and this is indeed the case (Table II). Moreover, the spacing between

segments C and D of aequorin is identical with that between domains III and IV of calmodulin. However, the lack of homology of B1/B2 to the other segments suggests an alternative mechanism for the evolution of aequorin involving the splicing of DNA coding for the B1/B2 segment into the site previously occupied by the second domain of a four-domain precursor gene.

One other precedent exists for an enzyme with both EF-hand domains and an identified enzymatic activity within the same polypeptide chain. This is the Ca(II)-dependent protease calpain (Ohno et al., 1984), and it is interesting to speculate on the evolution of aequorin and calpain. The 80K subunit of calpain probably arose by the fusion of a thiolprotease domain (homologous with papain) and a Ca(II) binding segment with four EF-hand domains (Ohno et al., 1984). In the case of aequorin, a similar splicing event could have formed the catalytic site, or the active site could have diverged from one EF hand within a four-domain precursor.

ACKNOWLEDGMENTS

We are grateful to David H. Ransom and Roger Wade for their technical assistance with various aspects of this work. We are particularly indebted to Dr. Koji Takio and Dr. Koiti Titani for many valuable discussions while completing this project. We thank Dr. Dennis Willows and the staff of the University of Washington Marine Laboratories, Friday Harbor, WA, for providing the facilities for the collection of *Aequorea*. We are indebted to Barbara Munro and Mary Woods for their assistance in preparing and typing the manuscript.

Registry No. Ca, 7440-70-2.

REFERENCES

Allen, D. G., Blinks, J. R., & Prendergast, F. G. (1977) *Science (Washington, D.C.) 195*, 996.

Arai, M. N., & Brinckmann-Voss, A. (1980) *Can. Bull. Fish. Aquatic Sci. 204*, 1.

Argos, P. (1977) *Biochemistry 16*, 665.

Argos, P., Hanei, M., & Garavito, R. M. (1978) *FEBS Lett. 93*, 19.

Atassi, M. Z., & Habeeb, A. F. S. A. (1972) *Methods Enzymol. 25*, 546.

Barker, W. C., Ketcham, L. K., & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, pp 273–283, National Biomedical Foundation, Washington, DC.

Bhown, A. G., Cornelius, T. U., Mole, J. E., Lynn, J. D., Tidwell, W. A., & Bennett, J. C. (1980) *Anal. Biochem. 102*, 35.

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

Blinks, J. R., Mattingly, P. H., Jewell, B. R., van Leeuwen, M., Harrer, G. C., & Allen, D. G. (1978) *Methods Enzymol. 57*, 292.

Brauer, A. W., Margolies, M. N., & Haber, E. (1975) *Biochemistry 14*, 3029.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45.

Cormier, M. J. (1978) in *Bioluminescence in Action* (Herring, P. J., Ed.) pp 75–108, Academic Press, New York.

Cormier, M. J., Hori, K., & Anderson, J. M. (1974) *Biochim. Biophys. Acta 346*, 137.

Cormier, M. J., Lee, J., & Wampler, J. E. (1975) *Annu. Rev. Biochem. 44*, 255.

Dayhoff, M. O., Barker, W. C., & Hunt, L. T. (1983) *Methods Enzymol. 91*, 524.

Doolittle, R. F. (1979) *Proteins (3rd. Ed.) 4*, 1.

Ericsson, L. H., Wade, R. D., Gagnon, J., McDonald, R. M., Granberg, R. R., & Walsh, K. A. (1977) in *Solid Phase Methods in Protein Sequence Analysis* (Previero, A., & Coletti-Previero, M. A., Eds.) pp 137–142, Elsevier/North-Holland, Amsterdam.

Gross, E. (1967) *Methods Enzymol. 11*, 238.

Gutman, G. A., Loh, E., & Hood, L. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 5046.

Hori, K., Wampler, J. E., Matthews, J. C., & Cormier, M. J. (1973) *Biochemistry 12*, 4463.

Hori, K., Charbonneau, H., Hart, R. C., & Cormier, M. J. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4285.

Hunkapiller, M. W., & Hood, L. E. (1978) *Biochemistry 17*, 2124.

Hunkapiller, M. W., & Hood, L. E. (1983) *Methods Enzymol. 91*, 486.

Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J., & Hood, L. E. (1983) *Methods Enzymol. 91*, 399.

Inouye, S., Noguchi, M., Sakaki, Y., Takagi, Y., Miyata, T., Iwanaga, S., Miyata, T., & Tsuji, F. I. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3154.

Kemple, M. D., Ray, B. D., Jarovi, G. K., Rao, B. D. N., & Prendergast, F. G. (1984) *Biochemistry 23*, 4383.

Klapper, D. G., Wilde, C. E., & Capra, J. D. (1978) *Anal. Biochem. 85*, 126.

Kretsinger, R. H. (1980) *CRC Crit. Rev. Biochem. 8*, 119.

Kretsinger, R. H., & Nockolds, C. E. (1973) *J. Biol. Chem. 248*, 3313.

Laemmli, U. K. (1970) *Nature (London) 227*, 680.

Mahoney, W. C., & Hermodson, M. A. (1980) *J. Biol. Chem. 255*, 11199.

Ohno, S., Emori, Y., Imajoh, S., Kawasaki, H., Kisaragi, M., & Suzuki, K. (1984) *Nature (London) 312*, 566.

Prasher, D., McCann, R. O., & Cormier, M. J. (1985) *Biochem. Biophys. Res. Commun. 126*, 1259.

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

Reimann, E. M., Titani, K., Ericsson, L. H., Wade, R. D., Fischer, E. H., & Walsh, K. A. (1984) *Biochemistry 23*, 4185.

Shimomura, O., & Johnson, F. H. (1973) *Tetrahedron Lett. 31*, 2963.

Shimomura, O., & Johnson, F. H. (1975a) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1546.

Shimomura, O., & Johnson, F. H. (1975b) *Nature (London) 256*, 236.

Shimomura, O., & Johnson, F. H. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2611.

Shimomura, O., Johnson, F. H., & Saiga, Y. (1962) *J. Cell. Comp. Physiol. 59*, 223.

Tarr, G. E., Beecher, J. F., Bell, M., & McKean, D. J. (1978) *Anal. Biochem. 84*, 622.

Tufty, R. M., & Kretsinger, R. H. (1975) *Science (Washington, D.C.) 187*, 167.

Ward, W. W., & Seliger, H. H. (1974) *Biochemistry 13*, 1491.

Ward, W. W., & Cormier, M. J. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2530.