*Protein Science* (2000), *9*:2085–2093. Cambridge University Press. Printed in the USA.


# Structure of the $Ca^{2+}$-regulated photoprotein obelin at 1.7 Å resolution determined directly from its sulfur substructure

ZHI-JIE LIU,[1,3] EUGENE S. VYSOTSKI,[1,2] CHUN-JUNG CHEN,[1,3] JOHN P. ROSE,[1] JOHN LEE,[1] AND BI-CHENG WANG[1]

[1]Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602
[2]Institute of Biophysics, RAS (SB), Krasnoyarsk 660036, Russia

(RECEIVED May 9, 2000; FINAL REVISION August 3, 2000; ACCEPTED August 18, 2000)

**Abstract**

The crystal structure of the photoprotein obelin (22.2 kDa) from *Obelia longissima* has been determined and refined to 1.7 Å resolution. Contrary to the prediction of a peroxide, the noncovalently bound substrate, coelenterazine, has only a single oxygen atom bound at the C2-position. The protein-coelenterazine 2-oxy complex observed in the crystals is photo-active because, in the presence of calcium ion, bioluminescence emission within the crystal is observed. This structure represents only the second de novo protein structure determined using the anomalous scattering signal of the sulfur substructure in the crystal. The method used here is theoretically different from that used for crambin in 1981 (4.72 kDa) and represents a significant advancement in protein crystal structure determination.

**Keywords:** bioluminescence; crystallography; obelin; photoprotein; single wavelength anomalous scattering; solvent flattening; sulfur phasing

The calcium-regulated photoprotein obelin, a 22.2 kDa single subunit protein, consists of 195 amino acid residues (Illarionov et al., 1995). Obelin together with aequorin and other photoproteins that originate mainly from marine bioluminescent coelenterates (Morin, 1974) are a subfamily of calcium-regulated photoproteins that is part of the larger calcium-binding protein family (Moncrief et al., 1990). The defining property of these proteins is that the calcium ion binding triggers the bioluminescence response through the oxidative decarboxylation of a tightly bound coelenterazine-oxygen substrate (Shimomura & Johnson, 1978). All photoproteins show high sequence homology and contain three "EF-hand" calcium-binding sites (Charbonneau et al., 1985; Illarionov et al., 1995; Tsuji et al., 1995).

Some calcium-binding proteins, for example sarcoplasmic calcium-binding protein, also have three EF-hand sites (Hermann & Cox, 1995). However, the majority of the calcium-binding proteins contain an even number, 2 or 4, EF-hand sites (calmodulin, troponin C, etc.). Comparison of the amino acid sequences of the photoproteins aequorin, clytin, mitrocomin, and obelin with those for the luciferin-binding protein of *Renilla* and human calmodulin, shows that the second calcium-binding site of calmodulin is missing in the corresponding sequence position in the photoproteins. Also, the sequence distance between the second and third calcium-binding sites in the photoproteins is found to be identical to the distance between the third and fourth calcium-binding sites in calmodulin, suggesting that these proteins have a common evolutionary origin (Tsuji et al., 1995).

The three-dimensional (3D) crystal structures of two bioluminescent proteins, bacterial luciferase (Fisher et al., 1995) and firefly luciferase (Conti et al., 1996), have been published. However, these structures are without the substrates that are required for the bioluminescence reaction. Consequently, not much insight into the nature of the catalytic center or the mechanism of the bioluminescence process has been obtained. The structural information of a calcium-regulated photoprotein with the coelenterazine-oxygen molecule bound is therefore very desirable, because it represents a natural and stable reaction intermediate, which is a snapshot of the bioluminescent pathway.

This paper describes the crystal structure of the bioluminescent photoprotein obelin from *Obelia longissima* containing a coelenterazine-oxygen molecule at 1.73 Å. A new approach was use for the structure solution where the sulfur atom substructure naturally present in the protein is used as a phasing probe in the structure determination. This technique can be viewed as a significant advancement in methods of protein structure determina-

Reprint requests to: B.C. Wang, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602; e-mail: wang@BCL1.bmb.uga.edu.
[3]These authors contributed equally to this work.

tion. Additionally, a brief comparison with a 2.3 Å structure of the related photoprotein aequorin that has just appeared (Head et al., 2000) is also made.

## Results

### *Overall structure*

The structure of the obelin molecule, as shown in Figure 1A, is highly compact and globular with a radius of ~25 Å. The molecular structure is formed by two sets of four helices designated A (16–29), B (39–54), C (58–74), and D (85–105) in the N-terminal domain, and E (110–122), F (132–142), G (148–157), and H (168–180) in the C-terminal domain (Fig. 1A,B). Both the N- and C-terminal domains can be thought of as a cup whose insides are lined with hydrophobic residues. The overall structure of obelin can then be considered as two "cups" joined at their rims (Fig. 1A). The coelenterazine-oxygen substrate resides in an internal cavity, which is surrounded by hydrophobic residues from the eight helices (Fig. 1A).

Analysis of the obelin structure reveals that the obelin molecule consists of four sets of helix-turn-helix (HTH) structural motifs, which are characteristic of EF-hand calcium-binding domains. The A- and B-helices are joined by a 9-residue loop, which together with three residues of the B-helix forms the first expected calcium-binding site. The amino acid composition of this site is similar to the calcium-binding sites of other calcium-binding proteins (Kretsinger & Nakayama, 1993). The C- and D-helices are linked by a 10-residue loop (75CGMEYGKEIA84). The amino acid residues forming this loop are known to be unable to coordinate calcium (Kretsinger & Nakayama, 1993) and, consequently, this HTH motif II is not functional for calcium binding. This was also observed for cardiac troponin C (Van Eerd & Takahshi, 1976) and troponin C from crayfish (Wnuk et al., 1984). The E- and F-helices are connected through a 9-residue loop, which together with three residues of the F-helix forms HTH motif III (the second expected calcium-binding site). The G- and H-helices are linked via a 10-residue loop, nine of which together with three residues of the H helix forms HTH motif IV (the third calcium-binding site). The amino acid composition and structural organization of calcium-binding sites III and IV are typical for canonic calcium-binding sites (Kretsinger & Nakayama, 1993) and consequently, both HTH motifs of the C-terminal can be expected to bind calcium ions. All loops of the HTH motifs are exposed to the solvent.

The electrostatic potentials (Nicholls et al., 1991) of the surfaces of the three potential calcium-binding sites (HTH motifs I, III, and IV) are negative (Fig. 2A). These provide likely regions for binding of calcium ions. On the other hand, the electrostatic potential at the corresponding position of the loop of HTH motif II is neutral. This explains that while I, III, and IV HTH motifs can bind calcium ions, HTH motif II cannot. In addition, the surface electrostatic potential of the loops at the HTH motif III and IV is more negative than that of HTH motif I. This indicates that the calcium-binding constants for III and IV calcium-binding sites are probably higher than for site I.

As described previously, the N- and C-terminal domains form cup-like structures. The N-terminal cup is formed by the interactions between HTH motifs I and II while the C-terminal cup is formed by the interactions between HTH motifs III and IV. Intermotif interactions occur via the loops and the helices. Two loops of the adjacent HTH motifs interact via hydrogen bonding in the form of antiparallel $\beta$-sheets. The loops of motifs I and II interact through hydrogen bonds between Ile37(O) and Ile83(N) (2.81 Å), and Ile37(N) and Ile83(O) (3.04 Å). The loops of motifs III and IV form hydrogen bonds between Ile130(N) and Leu176(O) (2.88 Å), Ile130(O) and Leu166(N) (2.99 Å). Similar intermotif interactions have also been observed in other calcium-binding proteins (Strynadka & James, 1989).

Although no hydrogen bonding was identified between the eight helices in the obelin structure, several close interhelical contacts (<3 Å) were observed including a salt bridge between Arg112(NH1) of helix E and Asp169(OD1) of helix H (2.97 Å).

### *Structure of the coelenterazine-oxygen binding pocket*

The coelenterazine-oxygen substrate and surrounding residues (within 4 Å) are shown in Figure 2B. The binding pocket is highly hydrophobic and is formed by residues originating from helix A (Met25 and Leu29), helix B (Ile42, Ala46, and Ile50), helix C (Phe72), helix D (Phe88 and Trp92), helix E (Ile111, Trp114, Val118, and Phe119), helix F (Trp135), Ile144 from the loop linking helices F and G, and helix H (Met171 and Trp179). In addition, several hydrophilic side chains are directed into the pocket. These are His22, Tyr138, and His175 localized in helices A, F, and H, respectively, and Tyr190 located near the C-terminus of the protein. The most protein-coelenterazine contacts are observed between residues of helix E in HTH motif III including a short (3.45 Å) contact to the main chain of Gly115. Almost all the residues forming the coelenterazine binding pocket are conserved among all calcium-regulated photoproteins (Fig. 1B) and consequently, the structure of the coelenterazine-oxygen binding sites should be conserved for all calcium-regulated photoproteins.

Site-directed mutagenesis and direct chemical modifications of photoproteins have shown certain amino acid residues to be important for photoprotein bioluminescence. These are Trp, His, Cys, and the C-terminal Pro. Mutants of aequorin with replacement of Trp residues for Phe exhibited various luminescent activity and spectra including a W86F mutant that gave a bimodal emission spectrum with maxima at 455 and 400 nm (Ohmiya et al., 1992). It was suggested that Trp86 in aequorin may be involved in the generation of the product excited state during photoprotein luminescence (Ohmiya et al., 1992). The obelin structure supports this suggestion since Trp92, Trp114, Trp135, and Trp179 (Fig. 2B) are among the residues that make close contact with the coelenterazine-oxygen substrate. For example, the side chains of Trp92 and Trp179 "sandwich" the p-hydroxyphenyl ring of coelenterazine. The planes of the phenyl ring and the Trp92 indole are almost parallel. The side chains of Trp114 and Trp135 are localized near the hydroxybenzyl group of coelenterazine.

Calcium-regulated photoproteins contain from 3 to 5 Cys residues (Fig. 1B) and it was shown that one of them is important for the bioluminescence function of photoproteins (Kurose et al., 1989; Bondar et al., 1999). However, in the current structure no Cys residues are observed near the coelenterazine-oxygen molecule. Also there are no disulfide bonds. Probably the Cys residues play some role in apoprotein "charging" with coelenterazine or can stabilize an active conformation of the photoprotein molecule since chemical modification of cysteine initiates a slow bioluminescence. A low rate of bioluminescence could be a consequence of destabilizing the photoprotein conformation (Bondar et al., 1999).

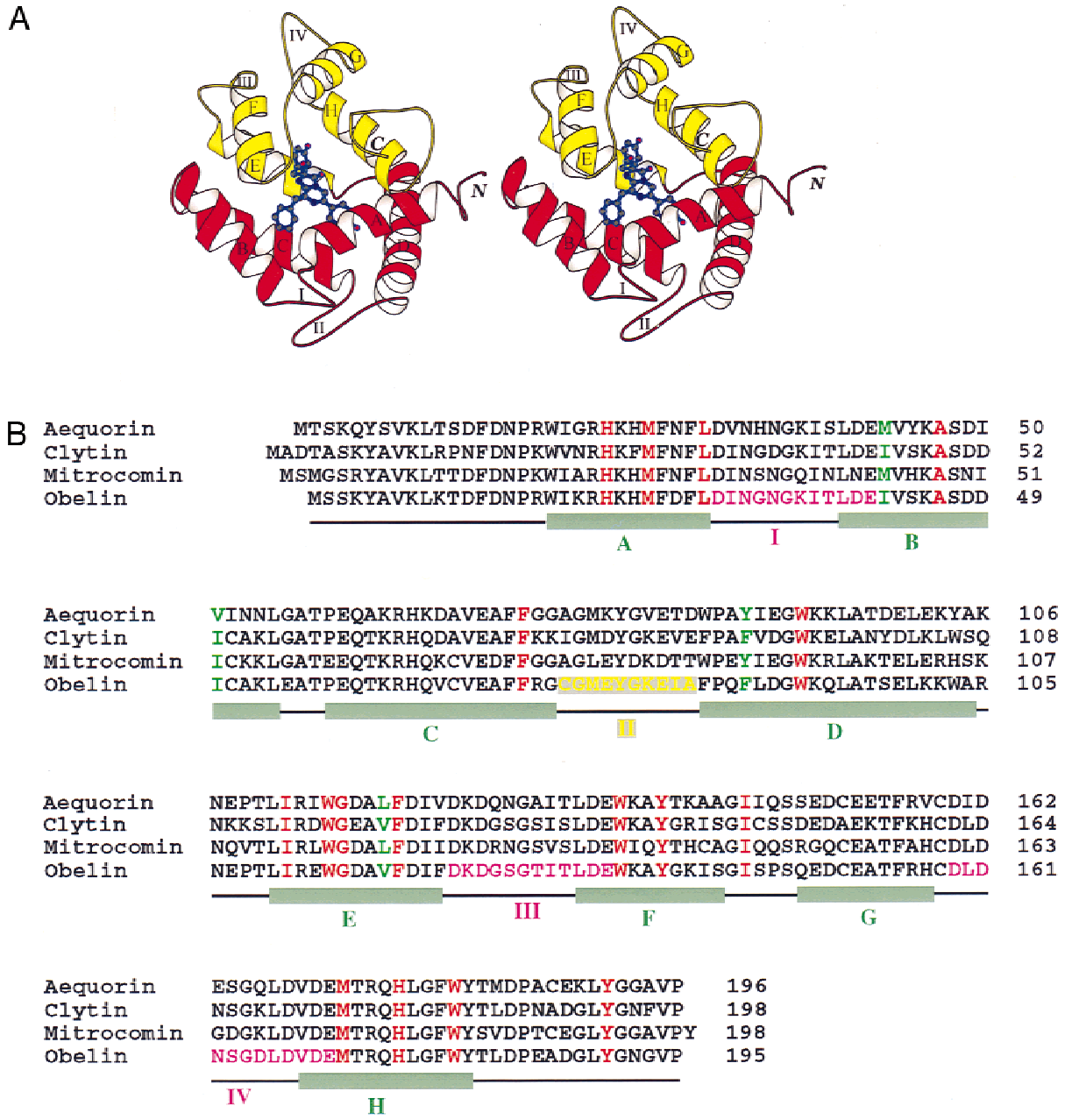


**Fig. 1. A:** A stereoview (Esnouf, 1997) of the overall crystal structure of obelin. The N- and C-terminal calcium-binding domains of the molecule are colored red and yellow, respectively. The coelenterazine-oxygen molecule is colored blue. The helices are marked by capital letters A through H. Roman numbers I through IV designate the loops of HTH motifs. The overall structure can be thought of as two "cups" joined at their rims. The top "cup" is formed by the N-terminal domain, which consists of helices A, B, C, and D, while the bottom "cup" is formed by the C-terminal domain, which consists of helices E, F, G, and H. **B:** A sequence alignment among the calcium-regulated photoproteins, aequorin (Inouye et al., 1985), clytin (Inouye & Tsuji, 1993), mitrocomin (Fagan et al., 1993), and obelin (Illarionov et al., 1995) showing the relative locations of residues forming the coelenterazine-oxygen binding pocket in obelin with respect to the secondary structure. Strictly conserved residues among these sequences, which are found in the substrate binding pocket, are shown in red. Other residues found in the substrate binding pocket are shown in green. The secondary structural features of the obelin sequence are shown below the sequence alignment. Helices A–H are shown green. The three predicted obelin calcium-binding loops (I, III, and IV) are shown in pink. Residues corresponding to the loop of the inactive HTH motif II are shown in dark yellow.

It has also been shown that one of His residues is important for photoprotein bioluminescence (Ohmiya & Tsuji, 1993; Bondar et al., 1999). It was suggested that a side chain of histidine could participate in molecular oxygen binding (Ohmiya & Tsuji, 1993). Figure 2B shows that His22 and His175 are in the vicinity of the coelenterazine-oxygen molecule. The N$\epsilon$ atom of His22 is very close (2.87 Å) to the oxygen atom of the coelenterazine phenyl ring implying a strong hydrogen bond or even complete proton

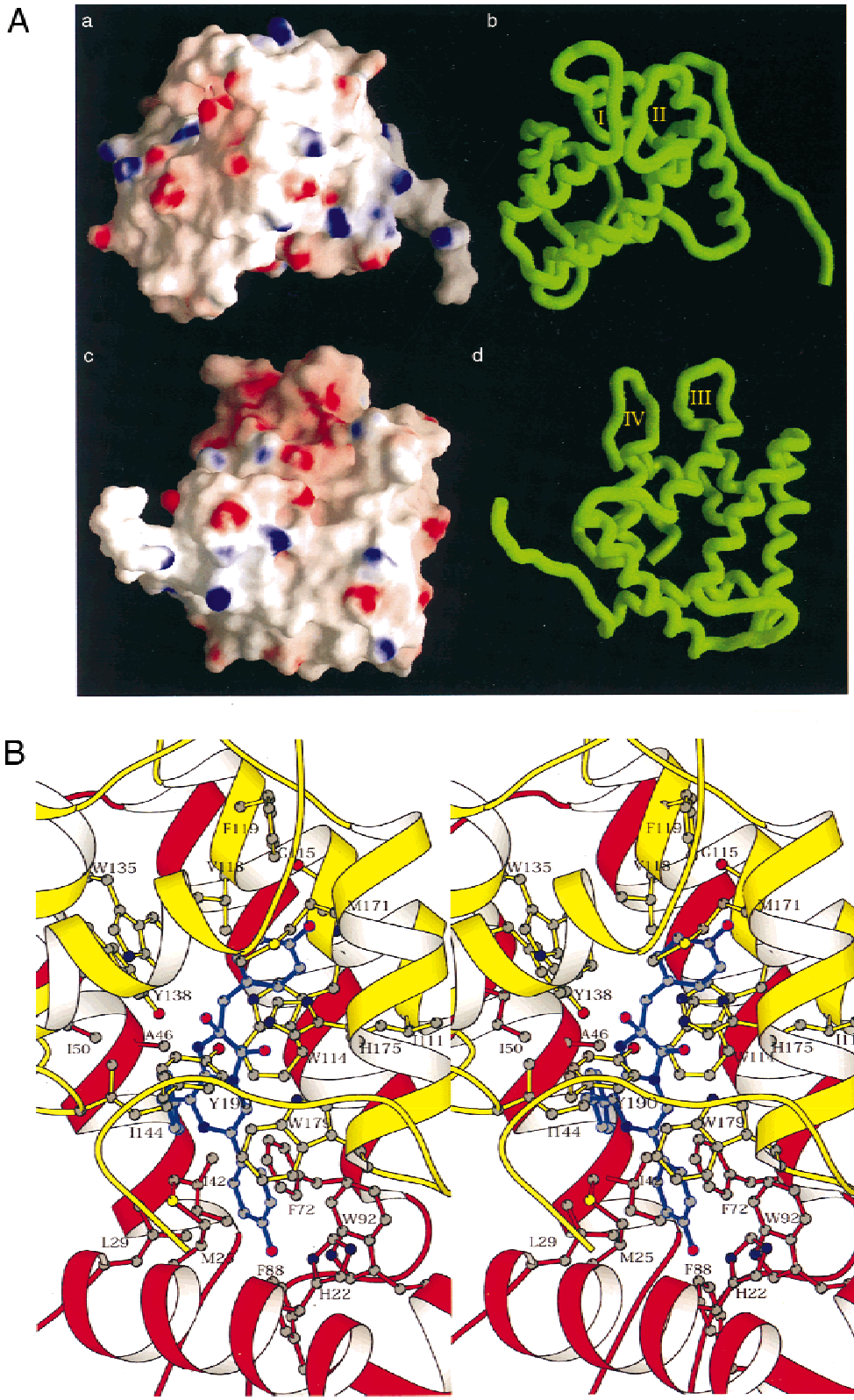


**Fig. 2. A:** (Left) A view of the solvent-accessible surface of obelin colored by electrostatic potential (negative is red, positive is blue) (Nicholls et al., 1991). **a** and **c** represent the molecular surface of the N- and C-terminal domains respectively, showing HTH motifs I and II (**a**) and III and IV (**c**). (Right) The corresponding worm drawings of the N- and C-terminal domains of obelin having the same orientation as shown in **a** and **c.** Roman numbers I to IV designate the loops of HTH motifs. **B:** A stereoview of the coelenterazine-oxygen binding pocket showing residues within 4 Å of the coelenterazine-oxygen molecule. Atoms are colored by element type (gray, C; red, O; blue, N; yellow, S), with bonds colored by residue location (yellow for residues located in HTH motifs I or II; red for residues located in HTH motifs III or IV).

transfer leaving the phenolate anion. The Nϵ atom of His175 lies near the carbonyl oxygen of coelenterazine (3.20 Å), and it is probable that this residue participates in some steps of the bioluminescent reaction. For instance, it could assist the formation of the transient anionic dioxetane in the course of the luminescent reaction.

Among other residues surrounding the coelenterazine-oxygen molecule in obelin are Tyr138 and Tyr190. The hydroxyl oxygen atom of Tyr138 lies 2.75 Å from the N1 atom of coelenterazine, implying a strong hydrogen bond between them. The hydroxyl group of Tyr190 lies 3.16 Å from the oxygen at C2 and only 2.6 Å from the Nϵ atom of His175.

One other important residue for the bioluminescence of photoproteins is a Pro residue at or near the C-terminal (Nomura et al., 1991). Deletion or replacement of this Pro residue destroys the luminescent capacity of photoproteins. Analysis of the obelin structure reveals that this residue can probably support the active conformation of the photoprotein through interaction between its oxygen atom and the Nϵ atom of Arg21, localized in the A-helix. As a result the N- and C-terminal domains form a closed conformation, inside of which the coelenterazine-oxygen substrate is contained.

### *Structure of the bound coelenterazine-oxygen*

Shimomura and Johnson (1978) suggested that in the photoprotein aequorin, the coelenterazine was bound to the protein through a peroxidic linkage substituted at the C2-position of the imidazopyrazinone ring. The peroxide structure was based on early experiments where attempts to remove molecular oxygen from the reacting solutions failed to ever change the kinetics of the bioluminescence, implying that the oxygen must be bound with the coelenterazine on charging the apo-aequorin. The C2-position of substitution was supported by NMR observation of a chemical shift of the $^{13}$C2-enriched coelenterazine on charging apo-aequorin (Musicki et al., 1986).

In the obelin structure shown here, however, the electron density distribution (Fig. 3B) near the C2-position of the coelenterazine molecule clearly corresponds to only a single oxygen, not peroxy,

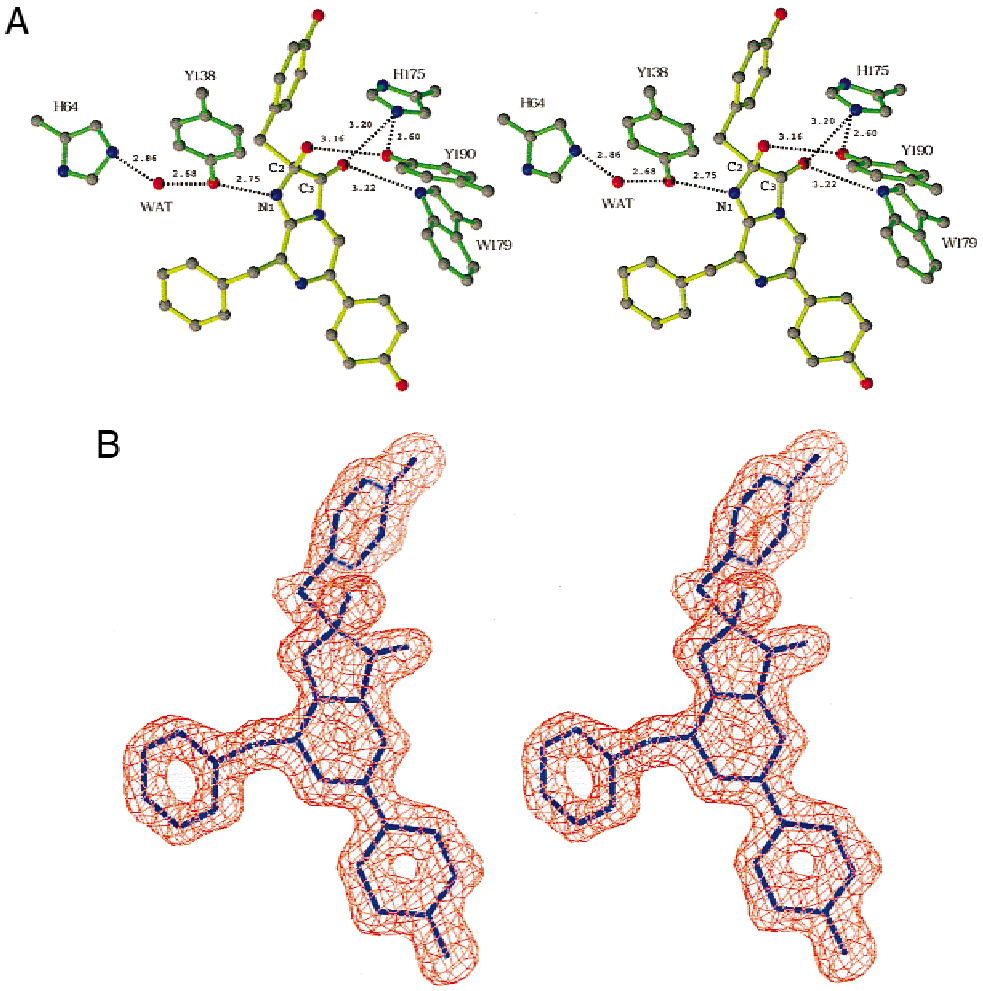


**Fig. 3. A:** The interaction of the coelenterazine-oxygen molecule with the Tyr-His-Trp triad (residues Y190, H175, W179) and residue Y138. **B:** A stereoscopic view of the fit of the coelenterazine-oxygen molecule to the 1.73 Å simulated annealing $2F_o - F_c$ omit map contoured at $1.5\sigma$. Note that the electron density next to the C2 position can only be fitted by a single oxygen atom, not by a peroxide group.

with no covalent binding to the protein. In addition, a hydrogen bond is evident between the C2-O and the hydroxyl oxygen of Tyr190 (3.16 Å) (Fig. 3A) leaving no space for a second oxygen atom.

*Crystal bioluminescence*

This Figure 3A structure is competent for the bioluminescence reaction. A droplet of $CaCl_2$ (100 mM) was placed on a microscope slide and viewed under a low power microscope attached to a monitor. A crystal that had been irradiated for the structural study was inserted into the droplet and an easily visible bright bioluminescence emission resulted. On the monitor, the bioluminescence emission was observed to be confined within the dimensions of the crystal, with a slow rise and decay of intensity over about 1 min apparently limited by the diffusion of the $Ca^{2+}$ into the crystal interior. In contrast, in free solution the bioluminescence reaction is finished within a few seconds. At the same time the crystal disintegrates and loses its yellow color, thus it is not possible to exclude solubilization of the obelin prior to the bioluminescence reaction. A fresh crystal, unexposed to X-rays, produced bioluminescence of the same intensity. Because the crystal size is small, it is not feasible to make quantitative statements about the relative activity of the obelin before and after irradiation. We conclude nevertheless that the crystals are made up of obelin molecules that are fully competent for light emission regardless of the irradiation by X-rays.

*Comparison of structure between obelin and aequorin*

After submission of this work described here, an independent study of the related photoprotein aequorin at 2.3 Å resolution appeared in publication (Head et al., 2000). A brief comparison of these two structures will be appended here to be expanded in a following work.

As expected from the homology of their primary sequence (Fig. 1B), both aequorin and obelin have the same tertiary structure, a compact globule with four HTH motifs, and the second EF-hand defective for $Ca^{++}$ binding. The root-mean-square deviation (RMSD) between the C$\alpha$ position of the two proteins is 1.48 Å. The same "cup" analogy was also used in describing the aequorin structure (Head et al., 2000). The coelenterazine-oxygen substrate in both photoproteins lies in a corresponding position and, interestingly, proximate to a Tyr-His-Trp triad (for example, Fig. 3A). This supports the notion that the triad is involved in the biochemistry of the bioluminescence reaction. Other residues in the two binding sites are also similar. No Cys residues are near the coelenterazine-oxygen molecule in either photoprotein.

The striking difference between the two studies concerns the oxygen atoms attached to the C-2 position of the coelenterazine. In the obelin electron density, we were only able to fit a single oxygen at this position and this gave us some pause before preparing this work for publication because previous studies, particularly on the photoprotein aequorin, strongly indicated that a peroxy-coelenterazine molecule was the bound state. Head et al. (2000) conclude that in aequorin, the electron density near the C-2 position is consistent with the coelenterazine bound with a peroxide. They fit two oxygen atoms between C-2 and the equivalent Tyr hydroxy atom (cf. Y190 in Fig. 3A). They point out, however, that the electron density is weaker here than for the rest of the coelenterazine molecule and that radiation damage could be responsible for the "loss of peroxide" density. In our structure, however, we have no evidence for a peroxide.

*Sulfur SAS phasing*

We would like to briefly describe the solvent flattening phasing procedure used in this work and how it differs from other anomalous phasing techniques such as multi-wavelength anomalous dispersion (MAD) (Hendrickson et al., 1990) and resolved anomalous scattering (RAS) (Hendrickson & Teeter, 1981) methods. The most popular MAD method requires the substitution of methionine residues by selenomethionine to introduce selenium as an anomalous phasing probe. MAD requires synchrotron X-ray data that are collected at three or four different wavelengths close to the selenium X-ray absorption edge (0.97 Å). Despite the fact that the Se-MAD method has been very successful for X-ray analysis, the preparation of selenomethionine substituted protein is not always straightforward.

Earlier, Hendrickson and Teeter (1981) used what they termed the RAS method in the structure determination of the small (4.72 kDa) protein crambin using the single-wavelength anomalous scattering (SAS) signal of the six sulfur atoms present in the asymmetric unit. Here, data were recorded to 1.5 Å resolution on a Cu sealed tube diffractometer system. The RAS method used in the crambin structure determination requires a large contribution (1.4% of the total scattering power) of the sulfur substructure to the Bijvoet differences, which is greater than that observed for most proteins. This is because phase estimate is based on sulfur phases that are assumed in the RAS method to be close to the protein phases. Thus, this assumption is valid only when the sulfur substructure dominates the protein structure, which is not the case for most proteins.

The procedure used here is called the iterative single-wavelength anomalous scattering (ISAS) method (Wang, 1985). It uses anomalous scattering signal of sulfur to first estimate the image of the protein (not the image of the sulfur substructure), which is in turn enhanced by solvent flattening. Thus, the protein phases are based on the initial protein image, not on the relative phases of the sulfur substructure as in the RAS method. Through simulation studies, Wang (1985) showed that the structure of a 12.5 kDa protein (Rhe) containing only a single disulfide could be solved using only the sulfur SAS data at 3.0 Å resolution. Wang's results demonstrated that if the signal could be measured accurately, solvent flattening could be an effective means of estimating protein phases based on only the protein's sulfur substructure, even when the sulfur content in the protein is relatively low. A recent study on egg-white lysozyme (Dauter et al., 1999) confirmed Wang's idea that sulfur SAS data combined with solvent flattening could be a useful tool for protein crystal structure determination.

The obelin structure described here represents the first de novo structure that has been determined using solvent flattening from sulfur SAS data. Figure 4 shows a representative section of the SAS electron density map at 3 Å resolution illustrating the quality of the sulfur phasing results. This approach, which will be described in detail elsewhere, appears not to be limited by (1) data resolution, (2) the requirements of selenium substitution, (3) a large sulfur substructure, or (4) data collected at multiple wavelengths as required by other anomalous scattering methods and should therefore be generally applicable to most structure determinations.

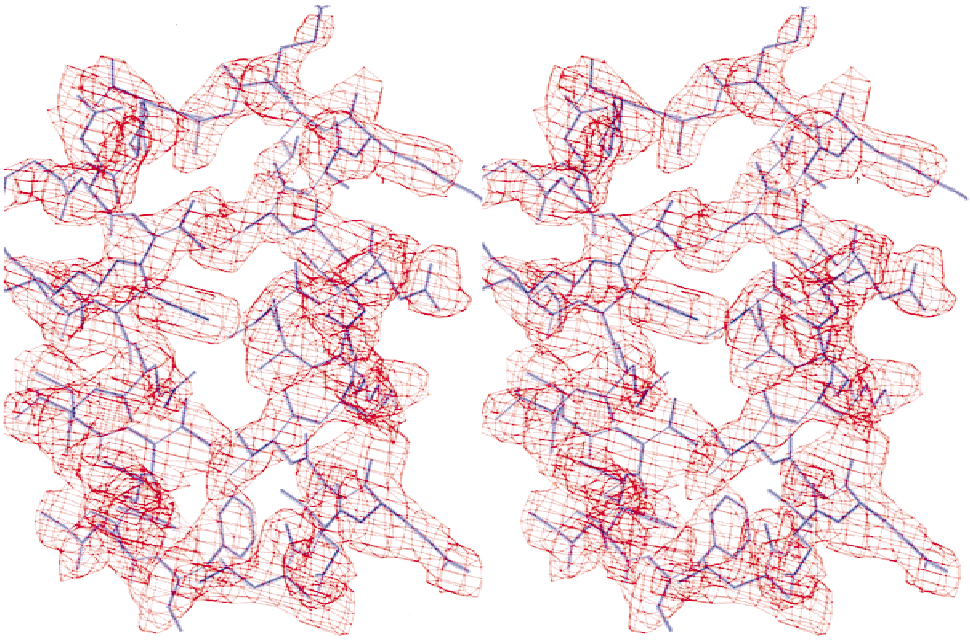

**Fig. 4.** An experimental ISAS electron density map at 3 Å resolution overlapped with refined coordinates of helices B and C to illustrate the quality of the sulfur phasing results by the current approach.

## Materials and methods

### *Protein expression, purification, and crystallization*

Recombinant obelin was expressed in *Escherichia coli* and purified as previously reported (Illarionov et al., 2000) with small modifications. Apoobelin was converted to obelin with synthetic coelenterazine (Prolume Ltd., Pittsburgh, Pennsylvania). The protein was homogeneous according to LC-Electrospray Ionization Mass Spectrometry. As shown by N-terminal amino acid sequence of recombinant obelin, Met1 is digested on producing recombinant apoobelin in *E. coli* cells. Thus, Ser2 corresponds to the first amino acid residue of the recombinant protein (Illarionov et al., 1995). The obelin mass obtained from electrospray ionization mass spectrometry measurements was in agreement with the amino acid sequence result. Crystals of obelin were grown as described previously (Vysotski et al., 1999). The crystals belong to space group $P6_2$ with cell dimensions $a = b = 81.55$ Å, $c = 86.95$ Å, and there is one molecule per asymmetric unit.

### *Data collection and data processing*

The sulfur anomalous scattering signal ($\Delta f''$) is comparatively weak, between 0.24 to 0.89 electrons in the tunable X-ray range (0.8–2.0 Å) of most synchrotrons. For comparison, selenium has a $\Delta f''$ of 3.85 electrons at the Se X-ray absorption edge (0.9797 Å). Thus, special care must be taken to enhance the anomalous scattering signal as well as to minimize errors (noise) in the experiment. In addition, one must continually check for the possible disappearance of sulfur anomalous scattering signal in the data due to radiation damage. To maximize the sulfur anomalous signal, longer wavelength (1.5–2.0 Å) X-rays are preferred. However, using longer wavelength X-rays can introduce crystal absorption errors into the data and when high intensity synchrotron radiation is used it can cause radiation damage to the sulfur substructure resulting in the loss of the sulfur SAS signal. We have found that a wavelength between 1.5–1.8 Å gives a reasonable compromise between maximizing the sulfur SAS signal and minimizing problems associated with using longer wavelength X-rays. In addition, we have concentrated our attention on obtaining the best possible anomalous scattering data to 3 Å resolution since the phasing procedure does not require high resolution data. This allowed us to place the detector as far as possible from the crystal to reduce the background noise level and increase the signal to noise level in the data for phasing purposes.

Data for phasing were collected to 2.5 Å resolution on two obelin crystals, flash-cooled to 100 K, at beamline ID-17 (IMCA-CAT), Advanced Photon Source (APS), Argonne National Laboratory using a MarResearch 165 mm CCD detector and 1.74 Å X-rays. The longer wavelength is used for the reasons discussed above. HKL2000 (Otwinowski & Minor, 1997) was used to determine a data collection strategy (98% completion), and data were collected using the "Freidel-flip" method (Derewenda et al., 1989). Data processing was carried out using HKL2000 (Otwinowski & Minor, 1997). The processed data from the two data sets were merged to increase redundancy. Data used for the high resolution refinement of the protein were collected to 1.7 Å resolution on a single obelin crystal, flash-cooled to 100 K, using the experimental setup described above and 0.94 Å X-rays. The shorter wavelength used in this study was chosen to decrease radiation damage to the crystal, reduce errors due to absorption, and to obtain higher resolution data. Data collection and processing statistics are given in Table 1.

### *Phasing and model building*

Low-resolution data (4.5 Å) was used to locate the sulfur atom sites. A total of 1,949 Bijvoet reflection pairs $>2\sigma$, between 20.0

**Table 1.** *X-ray crystallography data statistics*

| | |
|---|---|
| **A. Data collection and processing statistics** | |
| Data set used for phasing | |
| Wavelength (Å) | 1.74 |
| Resolution[a] (Å) | 2.5 |
| Completeness (%) | 99.9 |
| Bijvoet redundancy | 5.8 |
| $R_{merge}$[b] (%) | 3.7 (8.0) |
| Data set used for refinement | |
| Wavelength (Å) | 0.94 |
| Resolution[a] (Å) | 1.73 |
| Completeness (%) | 97.5 |
| $R_{merge}$[b] (%) | 3.8 (21.6) |
| **B. Refinement statistics** | |
| $R$ value $2\sigma_F$ cutoff | 0.189 |
| $R$ value all data | 0.192 |
| Free $R$ value | 0.207 |
| Free $R$ value test set | 2,665 reflections |
| RMSDs from ideality | |
| Bond lengths (Å) | 0.005 |
| Bond angles | 1.1° |
| Dihedral angles | 19.6° |
| Improper angles | 0.70° |
| Wilson $B$ value (Å$^2$) | 17.5 |
| Mean $B$ value (Å$^2$) | 18.8 |
| Coordinate error (Å)[c] | 0.21 |
| Resolution range (Å) | 20.0 − 1.73 |
| Data cutoff | $2.0\sigma_F$ |
| Number of reflections | 33,058 reflections |
| Completeness for range | 97.5% |
| Number of protein atoms | 1,557 |
| Number of SUBSTRATE atoms | 33 |
| Number of solvent atoms | 269 |

[a]Highest resolution of data set followed by resolution range in highest bin for compiling statistics.
[b]$R_{merge} = \sum\sum_j |F - \langle F \rangle| / \sum \langle F \rangle$.
[c]Estimated coordinate error from the Luzzati (1952) plot.

and 4.5 Å resolution (Bijvoet difference $R$-factor was 1.5%), were used to determine six of the eight sulfur atom positions (SOLVE V1.16; Terwilliger, 1997). PHASES (Furey & Swaminathan, 1997) was used to refine the sulfur positions and to locate the remaining sulfur sites by Bijvoet difference Fourier analysis. Three additional sites were identified, giving a total of nine sites, suggesting that one of the sulfur sites was most probably an anomalous solute such as $Cl^-$. All anomalous sites were treated as sulfur for the phase calculations.

The nine sulfur sites were used to estimate the protein phases at 3.0 Å resolution using ISAS (Wang, 1985). After three filters and two cycles of phase extension, the final average figure-of-merit and map-inversion $R$-value were 0.74 and 0.294, respectively. A preliminary chain trace was done using the 3.0 Å map. Sequence assignment was initially guided by the eight sulfur atom positions. The model, 99% complete, was built (O) (Jones et al., 1991) using the 3.0 Å ISAS map. The N-terminal serine was not observed in the electron density maps and was assumed to be disordered. The coelenterazine-oxygen molecule was then modeled into the electron density using O.

*Refinement*

The initial refinement was carried out using CNS 0.9 (Brünger et al., 1998) and included all data to 2.5 Å resolution. The resolution of the structure was then extended to 1.73 Å during the final stages of refinement using data collected at an X-ray wavelength of 0.94 Å. Using the 1.73 Å data, the structure was refined against all data between 20.0 and 1.73 Å by simulated annealing and positional refinement using the Engh and Huber force field (Engh & Huber, 1991) as incorporated in CNS. Individual atoms were assigned isotropic $B$-factors, which were refined. A total of 268 solvent molecules with peak heights above $4\sigma$ and good hydrogen bonding geometry were identified by difference Fourier analysis using CNS. In addition, one of the initial sulfur sites identified by SOLVE was located in solvent and modeled as $Cl^-$.

The final model includes 194 of 195 amino acids, one coelenterazine-oxygen molecule, one $Cl^-$ ion and 268 water molecules. The $R$-factor is 18.9% for all data between 20.0 and 1.73 Å resolution. Using the 8% reflection test set the free $R$-factor value is 20.7%. The final model contains a total of 1,557 protein atoms, 32 coelenterazine atoms, one oxygen atom, 268 water molecules, and 1 solvent molecule assigned as a $Cl^-$ ion. The model has good stereochemistry with RMSDs in bond length and angles of 0.005 Å and 1.1°, respectively. Analysis of the Ramachandran plot (PROCHECK) (Laskowski et al., 1993) showed no residues in disallowed regions. Atomic coordinates have been deposited with the Protein Data Bank (Bernstein et al., 1977), accession code 1EL4.

## Acknowledgments

We wish to thank Andrew Howard, John Chrzas, Andrzej Joachimiak, Gerd Rosenbaum, and Rongguang Zhang, for help in data collection, Arthur Robbins and Bayer Corporation for donation of beam time, Gary Newton and Franz Müller for helpful discussions, Bruce Bryan (Prolume Ltd.) and David Smith for gifts of coelenterazine. This work was supported by ONR grant N 00014-99-1-0414, Georgia Research Alliance, and in part by grant 99-04-48452 from the Fundamental Research Foundation of the Russian Academy of Sciences.

Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Energy Research, under Contract No. W-31-109-Eng-38. For the IMCA beamline: "Data were collected at beamline 17-ID in the facilities of the Industrial Crystallography Association Collaborative Access Team (IMCA-CAT) at the Advanced Photon Source. These facilities are supported by the companies of the Industrial Macromolecular Crystallography Association through a contract with Illinois Institute of Technology (IIT), executed through the IIT's Center for Synchrotron Radiation Research and Instrumentation."

## References

Bernstein FC, Koetzle TF, Williams GJ, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem 80*:319–324.

Bondar VS, Frank LA, Malikova NP, Inzhevatkin EV, Illarionova A, Vysotski ES. 1999. Bioluminescent activity of the recombinant obelin after chemical modification of histidine and cysteine residues. In: Roda A, Pazzagli M, Kricka LJ, Stanley PE, eds. *Bioluminescence and chemiluminescence: Perspectives for the 21st century*. Chichester, UK: John Wiley & Sons. pp 400–403.

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr 54*:905–921.

Charbonneau HW, Walsh KA, McCann RO, Prendergast FG, Cormier MJ, Vanaman TC. 1985. Amino acid sequence of the calcium-dependent photoprotein aequorin. *Biochemistry 24*:6762–6771.

Conti E, Franks NP, Brick P. 1996. Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure 4*:287–298.

Dauter Z, Dauter M, de La Fortelle E, Bricogne G, Sheldrick GM. 1999. Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J Mol Biol 289*:83–92.

Derewenda Z, Yariv J, Helliwell JR, Kalb AJ, Dodson EJ, Papiz MZ, Wan T, Campbell J. 1989. The structure of the saccharide-binding site of concanavalin A. *EMBO J 8*:2189–2193.

Engh RA, Huber R. 1991. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr A47*:392–400.

Esnouf RM. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph Model 15*:112–113, 132–134.

Fagan TF, Ohmiya Y, Blinks JR, Inouye S, Tsuji FI. 1993. Cloning, expression and sequence analysis of cDNA for the $Ca^{2+}$-binding photoprotein, mitrocomin. *FEBS Lett 333*:301–305.

Fisher AJ, Raushel FM, Baldwin TO, Rayment I. 1995. Three-dimensional structure of bacterial luciferase from Vibrio harveyi at 2.4 Δ resolution. *Biochemistry 34*:6581–6586.

Furey W, Swaminathan S. 1997. PHASES-95: A program package for the processing and analysis of diffraction data from macromolecules. *Methods Enzymol 277*:590–620.

Head JF, Inouye S, Teranishi K, Shimomura O. 2000. The crystal structure of the photoprotein aequorin at 2.3 Å resolution. *Nature 405*:372–376.

Hendrickson WA, Horton JR, LeMaster DM. 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure. *EMBO J 9*:1665–1672.

Hendrickson WA, Teeter MM. 1981. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature 290*:107–112.

Hermann A, Cox JA. 1995. Sarcoplasmic calcium-binding protein. *Comp Biochem Physiol B Biochem Mol Biol 111*:337–345.

Illarionov BA, Bondar VS, Illarionova VA, Vysotski ES. 1995. Sequence of the cDNA encoding the $Ca^{2+}$-activated photoprotein obelin from the hydroid polyp *Obelia longissima*. *Gene 153*:273–274.

Illarionov BA, Frank LA, Illarionova VA, Bondar VS, Vysotski S, Blinks JR. 2000. Recombinant obelin: Cloning, and expression of cDNA, purification, and characterization as a calcium indicator. *Methods Enzymol 305*:223–249.

Inouye S, Noguchi M, Sakaki Y, Takagi Y, Miyata T, Iwanaga S, Tsuji FI. 1985. Cloning and sequence analysis of cDNA for the luminescent protein aequorin. *Proc Natl Acad Sci USA 82*:3154–3158.

Inouye S, Tsuji FI. 1993. Cloning and sequence analysis of cDNA for the $Ca^{2+}$-activated photoprotein, clytin. *FEBS Lett 315*:343–346.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A 47*:110–119.

Kretsinger RH, Nakayama S. 1993. Evolution of EF-hand calcium-modulated proteins. IV. Exon shuffling did not determine the domain compositions of EF-hand proteins. *J Mol Evol 36*:477–488.

Kurose K, Inouye S, Sakaki Y, Tsuji FI. 1989. Bioluminescence of the $Ca^{2+}$-binding photoprotein aequorin after cysteine modification. *Proc Natl Acad Sci USA 86*:80–84.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Luzzati V. 1952. Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallogr 5*:802–810.

Moncrief ND, Kretsinger RH, Goodman M. 1990. Evolution of EF-hand calcium-modulated proteins. I. Relationships based on amino acid sequences. *J Mol Evol 30*:522–562.

Morin JG. 1974. Coelenterate bioluminescence. In: Muscatine L, Lenhoff HM, eds. *Coelenterate biology: Reviews and new perspectives*. New York: Academic Press. pp 397–438.

Musicki B, Kishi Y, Shimomura O. 1986. Structure of the functional part of photoprotein aequorin. *J Chem Soc Chem Comm 21*:1566–1568.

Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins 11*:281–296.

Nomura M, Inouye S, Ohmiya Y, Tsuji FI. 1991. A C-terminal proline is required for bioluminescence of the $Ca^{2+}$-binding photoprotein, aequorin. *FEBS Lett 295*:63–66.

Ohmiya Y, Ohashi M, Tsuji FI. 1992. Two excited states in aequorin bioluminescence induced by tryptophan modification. *FEBS Lett 301*:197–201.

Ohmiya Y, Tsuji FI. 1993. Bioluminescence of the $Ca^{2+}$-binding photoprotein, aequorin, after histidine modification. *FEBS Lett 320*:267–270.

Otwinowski Z, Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol 276*:307–326.

Shimomura O, Johnson FH. 1978. Peroxidized coelenterazine, the active group in the photoprotein aequorin. *Proc Natl Acad Sci USA 75*:2611–2615.

Strynadka NC, James MNG. 1989. Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu Rev Biochem 58*:951–998.

Terwilliger TC. 1997. Multiwavelength anomalous diffraction phasing of macromolecular structures: Analysis of MAD data as single isomorphous replacement with anomalous scattering data using the MADMRG Program. *Methods Enzymol 276*:530–537.

Tsuji FI, Ohmiya Y, Fagan TF, Toh H, Inouye S. 1995. Molecular evolution of the $Ca^{2+}$-binding photoproteins of the Hydrozoa. *Photochem Photobiol 62*:657–661.

Van Eerd JP, Takahshi K. 1976. Determination of the complete amino acid sequence of bovine cardiac troponin C. *Biochemistry 15*:1171–1180.

Vysotski ES, Liu ZJ, Rose J, Wang BC, Lee J. 1999. Preparation and preliminary study of crystals of the recombinant calcium-regulated photoprotein obelin from the bioluminescent hydroid *Obelia longissima*. *Acta Crystallogr D Biol Crystallogr 55*:1965–1966.

Wang BC. 1985. Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol 115*:90–112.

Wnuk W, Schoechlin M, Stein EA. 1984. Regulation of actomyosin ATPase by a single calcium-binding site on troponin C from crayfish. *J Biol Chem 259*:9017–9023.