

SHINING THE LIGHT: STRUCTURE AND FUNCTION RELATIONSHIP OF CALCIUM-REGULATED PHOTOPROTEINS

by

LU DENG

(Under the Direction of Bi-Cheng Wang)

ABSTRACT

Bioluminescence is a widespread natural phenomenon in which visible light is emitted by a living organism. Ca^{2+} -regulated photoproteins are found and responsible for the light emission in a variety of bioluminescent marine organisms, mostly in coelenterates. They are members of the broad EF-hand calcium binding protein family. Upon binding calcium, Ca^{2+} -regulated photoprotein undergoes conformational changes, converting itself into a luciferase that catalyzes the oxidation of coelenterazine by the bound molecular oxygen, yielding visible blue light, carbon dioxide and Ca^{2+} -discharged photoprotein. X-ray diffraction experiments are capable of capturing snap shots of molecular conformations trapped in the crystal and revealing structural details at near atomic resolution. The structural studies of photoproteins from the jellyfish *Aequorea* (aequorin) and the hydroid *Obelia* (obelin) demonstrate that conformations of apo-photoproteins are controlled by binding various ligands such as Ca^{2+} , coelenterazine, and coelenteramide which are necessary components for photoprotein function and unambiguously conclude the hydroperoxy-coelenterazine binding state of the luciferin. A proton relay mechanism is proposed for the Ca^{2+} -triggered bioluminescence reaction of photoproteins. New insights are obtained into how the protein environment could induce different ionic states of the

bound coelenteramide whose excited states are responsible for the different bioluminescence and fluorescence spectra of photoproteins. A new approach of structure determination using diffraction signals of weak anomalous scatterers, sulfur and calcium, pushes X-ray crystallographic methodology to a higher level and could be of general interest.

INDEX WORDS: Photoprotein, obelin, bioluminescence, calcium binding, EF-hand, aequorin, fluorescence

SHINING THE LIGHT: STRUCTURE AND FUNCTION RELATIONSHIP OF CALCIUM-
REGULATED PHOTOPROTEINS

by

LU DENG

B.S., Sichuan University, People's Republic of China, 1993

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

© 2004

LU DENG

All Rights Reserved

SHINING THE LIGHT: STRUCTURE AND FUNCTION RELATIONSHIP OF CALCIUM-
REGULATED PHOTOPROTEINS

by

LU DENG

Major Professor:	Bi-Cheng Wang
Committee:	John Lee Robert S. Phillips James H. Prestegard Donald M. Kurtz, Jr.

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

DEDICATION

To my dear grand-parents and parents

ACKNOWLEDGMENTS

It has been four and half years since I started my endeavor for my Ph. D. degree. I was fortunate to have been associated with a wonderful group of people helping every step along the way. I am most grateful to my major advisor and mentor, Prof. Bi-Cheng Wang, for his generous support, wisdom, encouragement, guidance and patience during the course of my graduate study. I appreciate deeply the scientific environment, the freedom and opportunities that he fostered; even more so was his love and passion for science that have given me courage and persistence to face challenges and tough times in my research, preparing me for a new level of my academic career.

I would also like to thank Drs. John Lee, Robert S. Phillips, James H. Prestegard and Donald M. Kurtz, Jr. for their serving on my advisory committee. They were always there whenever I was in needs of help and advice. I have my special thanks to Dr. John Lee, who is also a collaborator of my research project, for his patience and help in guiding me through the project; his enthusiasm in science, extensive knowledge, creative ideas, are inspirations that will be with me for many years ahead of me.

Many heartfelt thanks are sent to Drs. John Rose and James Liu for their training and teaching in crystallography. My deepest thanks also extend to Drs. Eugene S. Vysotski and Svetlana V. Markova for their warmth, wisdom, providing me the greatest protein samples. They all have contributed to the growth of my knowledge. Their experience, involvements and imagination were very important to my progress.

The best five years of my life has been with the Department of Chemistry, UGA, leaving me with great memories. I am indebted to my co-workers who have each enriched my life in their own ways.

Last but not least, I am especially grateful to my parents, my grand-parents, my brother and my friend. Without their love, their unconditional sacrifices and support, my achievements would not have been possible.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	x
LIST OF FIGURES AND SCHEMES	xi
LIST OF ABBREVIATIONS	xiii
 CHAPTER	
1 Introduction	1
1.1 Components in a bioluminescence system	1
1.2 A class of photoproteins requires calcium for function	2
1.3 Ca^{2+} -regulated photoproteins may form complex in vivo with GFP	5
1.4 Essential features of the Ca^{2+} -regulated photoproteins	5
1.5 Spectral properties of Ca^{2+} -regulated photoproteins	8
1.6 Previous studies on bioluminescence emitter identification	8
1.7 A simple model of coelenterazine chemiluminescence mechanism	12
1.8 Mutational studies of photoproteins	14
1.9 Early structural investigations of Ca^{2+} -regulated photoproteins	15
1.10 Questions remain to be addressed	19
1.11 Potential applications of our research	20
1.12 Significance of this work	21

2	Experimental Procedures and Methods	24
2.1	Preparation of photoprotein samples	24
2.2	Crystallization of photoproteins	26
2.3	Calcium soaking experiment	31
2.4	Data collection and processing of photoproteins	31
2.5	Phasing of photoproteins	37
2.6	Structure refinement of photoproteins.....	42
3	Structure-function Relationship	48
3.1	Crystal structure of OG-obelin	50
3.2	Coelenterazine binding pocket of OG-obelin.....	55
3.3	Structure comparisons of Ca^{2+} -regulated photoproteins	60
3.4	Coelenterazine-oxygen binding state	68
3.5	Identification of the metal ions in the atomic resolution OL-obelin structures.....	74
3.6	Crystal structure of Ca^{2+} -discharged W92F-obelin.....	79
3.7	Coelenteramide binding site	81
3.8	Ca^{2+} -triggered bioluminescence reaction mechanism of photoproteins.....	84
3.9	Bioluminescence emitter	90
3.10	Structures of Ca^{2+} -loaded apo-aequorin and Ca^{2+} -loaded apo-obelin.....	94
	REFERENCES	102
	APPENDICES	
A	Crystallization and Preliminary X-ray Analysis of GlcNAc α 1,4Gal-releasing Endo- β -galactosidase from <i>Clostridium perfringens</i>	108

B	Structure Determination of Fibrillarin from the Hyperthermophilic Archaeon <i>Pyrococcus furiosu</i>	111
C	Structural Basis for the Emission of Violet Bioluminescence from a W92F Obelin Mutant	119
D	Preparation and X-ray Crystallographic Analysis of the Ca ²⁺ -discharged Photoprotein Obelin	125

LIST OF TABLES

	Page
Table 2.1: Data collection and processing statistics	32
Table 2.2: X-ray anomalous scattering factors for S and Ca atoms.....	35
Table 2.3: Phase searching results of Ca ²⁺ -loaded apo-aequorin	39
Table 2.4: Refined positions of the calcium and sulfur atom sites used in phasing	40
Table 2.5: Refinement statistics and PDB deposition accession codes	43

LIST OF FIGURES AND SCHEMES

	Page
Figure 1.1: Crystal luminescent images of photoprotein	3
Figure 1.2: Bioluminescence and regeneration of photoproteins	4
Figure 1.3: Amino acid sequence alignment of calmodulin and apo-photoproteins	7
Figure 1.4: Bioluminescence and fluorescence spectra of aequorin and OL-obelin	9
Scheme 1.1: Coelenterazine and coelenteramide species	11
Scheme 1.2: MaCapra and Chang's model	13
Figure 1.5: Crystal structures of aequorin and OL-obelin	16
Figure 1.6: Coelenterazine-oxygen binding pocket of OL-obelin	17
Figure 1.7: Fit of the coelenterazine-oxygen molecules to electron density maps	18
Figure 2.1: Crystal pictures of photoproteins	27
Figure 2.2: Crystal fluorescence of Ca^{2+} -discharged W92F-obelin	30
Figure 2.3: The very first experimental electron density map of Ca^{2+} -loaded apo-aequorin	41
Figure 3.1: Photoprotein conformation change pathway	49
Figure 3.2: Crystal structure of OG-obelin	51
Figure 3.3: Sequence alignment of aequorin, OL-obelin and OG-obelin	54
Figure 3.4: Electron density maps fitted with hydroperoxy-coelenterazine molecules	56
Figure 3.5: Interactions between hydroperoxy-coelenterazine and OG-obelin	59
Figure 3.6: Bioluminescence and fluorescence spectra of OL-obelin and h-obelin	61
Figure 3.7: Structural superimpositions of photoproteins	63

Figure 3.8: Hydroproxy-coelenterazine and EF-hand motif III and IV of OG-obelin	64
Figure 3.9: Hydrogen-bonding interactions at the binding site of coelenterazine	66
Figure 3.10: Electron density maps fitted with coelenterazine of OL-obelin (C2)	71
Figure 3.11: B-factors of peroxy-coelenterazine atoms and key atoms.....	73
Figure 3.12: Anomalous difference Fourier maps of OL-obelin (C2).....	75
Figure 3.13: Positions of cobalt ions and potassium ion in OL-obelin (C2)	76
Figure 3.14: Ca^{2+} and K^{+} coordinate systems in OL-obelin (C2).....	78
Figure 3.15: Structure of Ca^{2+} -discharged W92F-obelin.....	80
Figure 3.16: Coelenteramide binding site and its electron density map	82
Figure 3.17: 2-D drawing of the hydrogen bond network of coelenteramide binding site.....	83
Scheme 3.3: Proposed proton relay mechanism	85
Figure 3.18: Fluorescence comparison of Ca^{2+} -discharged W92F-obelin	89
Figure 3.19: Crystal structures of Ca^{2+} -loaded apo-aequorin and Ca^{2+} -loaded apo-obelin	97
Figure 3.20: Ca^{2+} -binding site at the EF-hand motif of Ca^{2+} -loaded apo-obelin.....	98
Figure 3.21: Superimpositions of Ca^{2+} -loaded apoproteins and with photoprotein	100

LIST OF ABBREVIATIONS

ADP: anisotropic displacement parameters

APS: the Advanced Photon Source

BRET: bioluminescence resonance energy transfer

DMSO: dimethyl sulfoxide

EDTA: ethylenediaminetetraacetic acid

FOM: figure-of-merit

GFP: green fluorescent protein

HLH: helix-loop-helix

h-obelin: apo-obelin from *O. longissima* charged with an analog substrate, coelenterazine h

IMCA-CAT: Industrial Crystallography Association Collaborative Access team

MAD: multiple-wavelength anomalous diffraction

MIR: multiple isomorphous replacement

MR: molecular replacement

OG-obelin: obelin from *Obelia geniculata*

OL-obelin: obelin from *Obelia longissima*

RMSD: root-mean-square deviation

SAS: single-wavelength anomalous scattering

SER-CAT: the South East Regional Collaborative Access team

SIR: single isomorphous replacement

W92F-obelin: W92F mutant of obelin from *Obelia longissima*

CHAPTER 1

Introduction

Bioluminescence is a widespread natural phenomenon in which visible light is emitted by an organism, for example, in summer we see fireflies emitting beautiful yellow light in the dark. In the deep ocean, about 97% of species are bioluminescent. Bioluminescence is defined as light emission generated with high efficiency from a biological system, which provides behavioral and survival advantages to the organism. Its biological function can be summarized by several aspects: communication, illumination, predation, determent and mating, depending on organisms; a given organism may utilize luminescence in more than one way (1).

1.1 Components in a bioluminescence system

The biochemical reaction of bioluminescence involves oxidation of a substrate termed luciferin by an enzyme named luciferase, generally in the presence of molecular oxygen directly or indirectly. Light can also be produced from luciferin by oxidation with chemical reagents independent of luciferases, a phenomenon known as chemiluminescence. Basically, bioluminescence results from a chemiluminescence reaction in a protein environment generated with a high quantum yield. The high level of interest in this phenomenon has led to detailed studies of its chemistry and molecular biology. The most well-known and understood luciferin-luciferase system is that found in the firefly (2).

The term "photoprotein" is a general designation for certain bioluminescent proteins that do not fit the classical concept according to which an enzyme (luciferase) catalyzes the oxidation of a smaller organic substrate molecule (luciferin) with the creation of an excited state and the

emission of light (3). It was defined that a photoprotein directly participates in a light emitting reaction of a living organism capable of emitting light in proportion to the amount of protein (rather than in proportion to the amount of substrate such as a luciferin), and that is not the unstable, transient intermediate of an enzyme-substrate reaction (3). The great majority of photoproteins now known to exist are stimulated to give luminescence by calcium or certain other inorganic ions.

1.2 One class of photoproteins requires calcium for function

One of the photoproteins, which is found and responsible for the light emission in a variety of bioluminescent marine organisms, is called the Ca^{2+} -regulated photoprotein. The first Ca^{2+} -regulated photoprotein was discovered and isolated from the jellyfish *Aequorea*, named aequorin, by Shimomura et al. in 1962 (4). In 1971, a homologous protein called obelin from the hydroid *Obelia* was identified by Morin and Hastings (5). Similar photoproteins have been identified later on in other organisms, mostly in coelenterates (6).

Ca^{2+} -regulated photoproteins consist of a single polypeptide chain of relatively small size, approximately 22 kDa, and a non-covalently bound chromophore, generated by a combination of molecular oxygen and an organic molecule called coelenterazine, an imidazole pyrazine derivative. Upon binding calcium, a Ca^{2+} -regulated photoprotein undergoes conformational changes, converting itself into a luciferase that catalyzes the oxidation of coelenterazine by the bound molecular oxygen, yielding visible blue light (Fig. 1.1), a molecule of carbon dioxide and a Ca^{2+} -discharged photoprotein. The last is composed of coelenteramide, an oxidation product of coelenterazine, bound non-covalently to apoprotein (Fig. 1.2) (7).

Because the energy of light emission is derived from the oxidative degradation of the chromophore, each molecule of photoprotein can react only once. The reaction product of

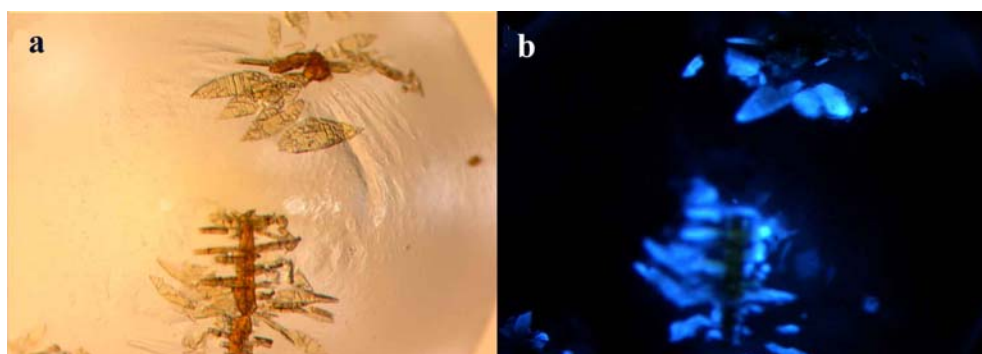


Fig. 1.1 Crystal luminescent images of Ca^{2+} -regulated photoprotein obelin from *Obelia geniculata* after crystals were soaked with CaCl_2 . Low quality crystals were transferred to a droplet containing only the crystallization mother liquor. CaCl_2 solution was inserted from top of the crystals. Luminescence from the crystals was observed in the dark as bright spots which lasted more than 1 min. During light emission the crystals deteriorated mechanically as periodically examined with a stereomicroscope. (a) Before soaking. (b) 10 seconds after soaking.

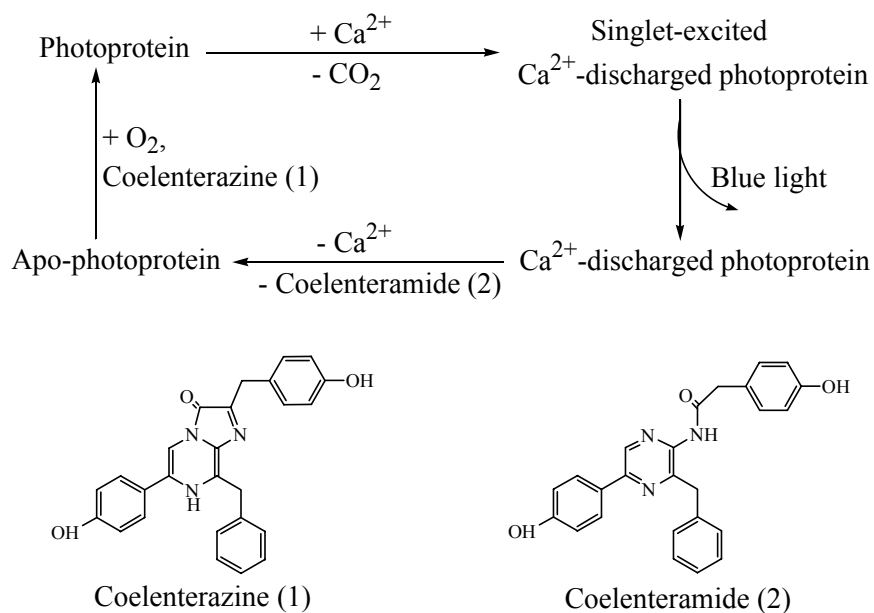


Fig. 1.2 Bioluminescence and regeneration of Ca^{2+} -regulated photoproteins. The photoprotein is a complex of coelenterazine (compound 1), apo-photoprotein and molecular oxygen. Binding of calcium triggers the bioluminescence reaction, producing singlet-excited coelenteramide (compound 2). The complex of coelenteramide and apo-photoprotein is called Ca^{2+} -discharged photoprotein. Light is emitted during decay of the excited coelenteramide to the ground state. Coelenteramide can be removed from Ca^{2+} -discharged aequorin by gel filtration in the presence of EDTA, regenerating the apo-photoprotein. Adapted from reference (8).

coelenterazine, coelenteramide, is more easily dissociated from the apo-aequorin than the coelenterazine. This is one of the reasons why apo-aequorin can be regenerated by gel filtration in the presence of ethylenediaminetetraacetic acid (EDTA) (Fig. 1.2). Recombinant apo-photoproteins can be recharged *in vitro* by incubation with synthetic coelenterazine under calcium free conditions in the presence of molecular oxygen and a sulfhydryl-reducing agent, such as dithiothreitol or β -mercaptoethanol (7). Since the aequorin could be reacted and recharged, Ca^{2+} -regulated photoprotein is classified as an enzyme, and resembles enzymes trapped in an intermediate state. It is not clear how active aequorin or obelin is regenerated *in vivo*.

1.3 Ca^{2+} - regulated photoproteins may form complex in vivo with GFP

Although aequorin and obelin emit blue light, the jellyfish *Aequorea* and hydroid *Obelia* have intense green bioluminescence. This is because the jellyfish *Aequorea* and hydroid *Obelia* do not give bioluminescence by the classical luciferin-luciferase mechanism; instead, the observed green light of the creatures arises from another protein called the green fluorescent protein (GFP) which is believed to associate with photoproteins (5). Photoproteins are contained in specialized cells called photocytes. The organism regulates its bioluminescence reaction by nerve-controlled calcium release and sequestration in the photocyte. When calcium ions enter and bind to the protein, followed by the bioluminescence reaction of the photoprotein with subsequent energy transfer to GFP to produce excited GFP which emits green light as it relaxes to ground state.

1.4 Essential features of the Ca^{2+} -regulated photoproteins

Two unique properties of the Ca^{2+} -regulated photoprotein distinguish it from other luciferase-luciferin bioluminescence systems. One feature was the discovery of calcium

dependency of photoproteins in order to produce light, which means that it has direct Ca^{2+} -dependent enzymatic activity. Calcium ion regulates but, is not essential for the function of these proteins (9). Photoprotein alone gives off a very low level of light emission called the Ca^{2+} -independent luminescence, but the light intensity is less than a millionth of light emission triggered by calcium binding (10). The amino acid sequence of aequorin determined 20 years after its first discovery revealed three canonical Ca^{2+} -binding EF-hands (11, 12) that are now known to be a common sequence feature in photoproteins from other genera (13, 14) (Fig. 1.3), indicating that they are members of the broad family of EF-hand calcium binding proteins (15). Sequence comparison of photoproteins aequorin (11, 12), clytin (16), mitrocomin (17), and obelin (13) with that of bovine calmodulin (18) shows that there is an additional EF-hand motif but lacking characteristic Ca^{2+} binding residues. Also the sequence distance between the second and third Ca^{2+} -binding sites in the photoproteins is found to be identical to the distance between the third and fourth Ca^{2+} -binding sites in calmodulin, suggesting that these proteins have a common evolutionary origin (19) (Fig. 1.3).

The other feature is that the bioluminescence intensity of photoproteins does not depend on oxygen concentration, in contrast to all bioluminescence reactions that we know of in which oxygen is involved. To explain the apparent oxygen independence of Ca^{2+} -regulated photoprotein bioluminescence, Shimomura and Johnson proposed in 1978 (20) that coelenterazine is able to react with dissolved oxygen to produce the hydroperoxy-coelenterazine complex as shown in Scheme 1.1, where a peroxy group is substituted at the C2 position of coelenterazine. So instead of using coelenterazine, Ca^{2+} -regulated photoproteins bind coelenterazine's peroxy derivative for bioluminescence reaction hence carry its own oxidizing agent.

apoAQ	VKLTSDFDNPRWIGRHKHMFNFDVNHNGKISLDEMVKASDIVINN	(48)
apoCL	VKLRPNFDNPKWVNRHKFMFNFLDINGDGKITLDEIVSKASDDICAKL	(48)
apoMI	VKLTTDFDNPKWIARHKHMFNFLDINSNGQINLNEMVHKASNIICKKL	(48)
apoOL	VKLTPDFDNPRWIKRHKHMFDFLDINGNGKITLDEIVSKASDDICAKL	(54)
Cal.	ADQLTEEQ IAEFKEAFSLFDKDGDTITTKELGTVM RSL	(39)
apoAQ	GATPEQAKRHKDAVEAFFGGA GMKYGVETDWPAYIEGWKKLATDELEK	(96)
apoCL	GATPEQTKRHQDAVEAFFKKI GMDYGKEVEFPFVDGWKELANYDLKL	(96)
apoMI	GATEEQTKRHQKCVEDFFGGA GLEYDKDTTWPEYIEGWKRLAKTELER	(96)
apoOL	EATPEQTKRHQVCVEAFFRGC GMEYGKEIAEPQFLDGWKQLATSELKK	(102)
Cal.	GQNPTEAE LQDMINEVDADGNG TIDFPEFLTMMAR KMKD	(78)
apoAQ	YSKNQITLIRLWGDALFDIIDKDQNGAISLDEWKAYTKSDGIIQSSD	(146)
apoCL	WSQNKSLIRDWGEAVFDIFDKDGSISLDEWKAYGRISGICSSDED	(146)
apoMI	HSKNQVTLIRLWGDALFDIIDKDRNGSVSLDEWIIQYTHCAGIQQSRGQ	(146)
apoOL	WARNEPTLIREWGDVAFDIFDKDGSITLDEWKAYGKISGISPSQED	(152)
Cal.	TDS EEE IR EA FRVFDKDGNGYISAAELRHVMTNLGEKLTDEE	(110)
apoAQ	CEETFRVCDIDESGQLDVDEMTRQHLGFWYTMDPACEKLYGGAVP	(189)
apoCL	AEKTFKHCDLNSGKLDVDEMTRQHLGFWYTLDPNADGLYGNFVP	(189)
apoMI	CEATFAHCDLDGDKLDVDEMTRQHLGFWYSVDPTCEGLYGGAVPY	(189)
apoOL	CEATFRHCDLNSGDLVDEMTRQHLGFWYTLDPNADGLYGNFVP	(195)
Cal.	VDEMIREADIDGDKVNYEEFVQMMTAK	(148)

Fig. 1.3 Amino acid sequence alignment of bovine calmodulin (Cal.) and apo-photoproteins aequorin (AQ), clytin (CL), mitrocomin (MI) and obelin (OL). Amino acid residues that are identical in the four apo-photoproteins are highlighted in blue. A red letter indicates an amino acid that is essential for bioluminescence activity. The Ca^{2+} -binding sites (EF-hand structures) are highlighted with yellow boxes. Adapted from reference (8).

1.5 Spectral properties of Ca^{2+} -regulated photoproteins

Following the bioluminescence emitted by Ca^{2+} -regulated photoproteins, the product, Ca^{2+} -discharged photoproteins give bright fluorescence upon excitation by near UV. Note that there is a difference between bioluminescence and fluorescence. The energy of the former is obtained from a chemical reaction whereas that of the latter is from the absorption of UV.

The bioluminescence spectral distributions of Ca^{2+} -regulated photoproteins are broad in the range 400 - 495 nm with maxima depending on the types. Photoproteins themselves are hardly fluorescent but the Ca^{2+} -discharged photoproteins have strong fluorescence with type-dependent maxima in the range 465-520 nm. Although aequorin and obelin are very similar to each other in primary structures with over 60% sequence identity, there are some obvious differences. Both photoproteins emit blue bioluminescence but the emission from obelin ($\lambda_{\text{max}} = 485 \text{ nm}$) is shifted to a longer wavelength in comparison with that of aequorin bioluminescence ($\lambda_{\text{max}} = 465 \text{ nm}$). Also a shoulder at 400 nm that is clearly revealed in obelin bioluminescence spectrum is not observed in the aequorin spectrum (Fig. 1.4) (14, 21, 22). Ca^{2+} -discharged photoproteins will emit fluorescence upon excitation by near UV. Although both proteins have approximately the same excitation maximum ($\sim 350 \text{ nm}$), the Ca^{2+} -discharged aequorin displays a blue fluorescence with $\lambda_{\text{max}} = 465 \text{ nm}$ that matches its bioluminescence spectrum whereas the Ca^{2+} -discharged obelin has a green fluorescence with $\lambda_{\text{max}} = 510 \text{ nm}$ that is shifted from its bioluminescence maximum by 25 nm (14, 21, 22). It is assumed that the coelenteramide binding pockets in these Ca^{2+} -discharged photoproteins differ from each other.

1.6 Previous studies on bioluminescence emitter identification

Identification of the bioluminescence emitter is a key requirement in solving the reaction mechanism. In the past the approach taken has been to study the chemiluminescence reaction of

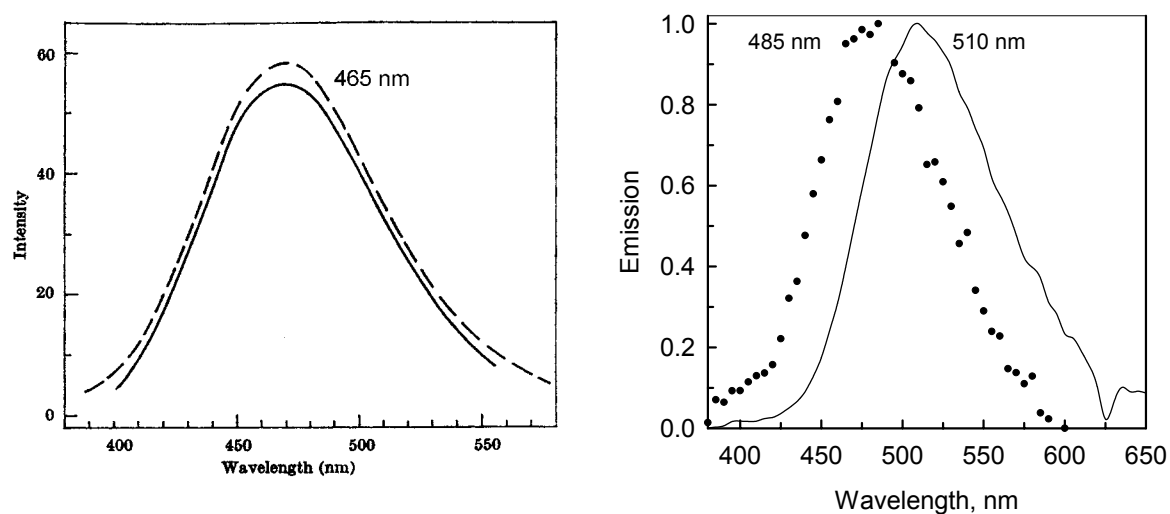
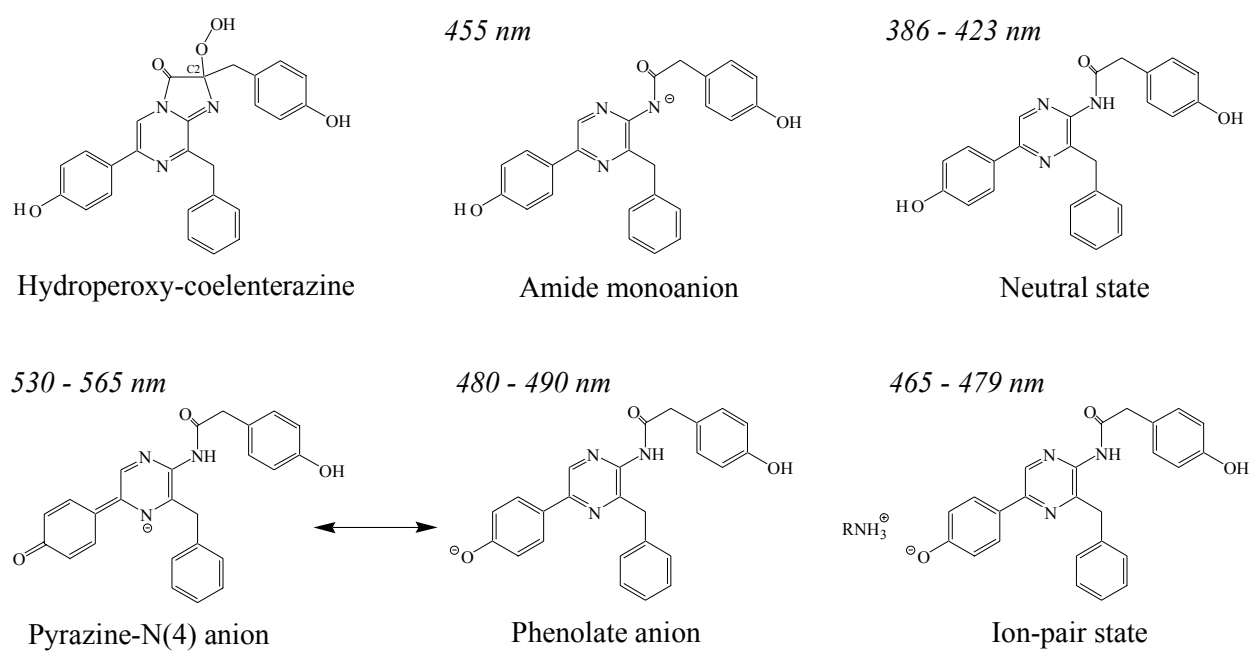


Fig. 1.4 Left panel: bioluminescence spectrum of aequorin (line) and fluorescence emission spectrum of Ca^{2+} -discharged aequorin (dash line). Adapted from (21). Right panel: bioluminescence spectrum of OL-obelin (dots) and fluorescence emission spectrum of Ca^{2+} -discharged OL-obelin (line) (14).

the substrate or close substrate analogues under various solution conditions that might mimic the protein reaction site environment. Fluorescence studies of coelenteramide and analogues have shown that the differences of the emission spectra among photoproteins are due to the different ionic states of coelenteramide caused by the protein environment. Several other ionic species can exist with emission maximum depending on the solvent system used (Scheme 1.1), a neutral species with fluorescence spectral maximum around 400 nm, amide monoanion around 455 nm, a phenolate anion 480-490 nm, and other anionic species with maxima at longer wavelengths (23, 24). Small variations are also introduced by solvent dielectric constant and solvent viscosity changes.

The general consensus is that the amide monoanion is the primary chemical product generated in its excited state from which the bioluminescence emission occurs. Part of the identifying evidence for the amide monoanion as the primary excited state was in the case of aequorin, a close match between its bioluminescence spectrum and the fluorescence of the protein-bound product. However for other photoproteins such as obelin, the product fluorescence is not the same and is at longer wavelength than the bioluminescence (22). The studies by Shimomura and Teranishi (23) and by Imai et al. (24) suggested that the phenolate ion from coelenteramide had a singlet excited state with a charge-transfer character. Consequently the energy level should be very sensitive to solvent polarity. Their interpretation was that the phenolate fluorescence maximum ranged from 592 nm in aprotic polar solvents such as DMSO, to 466 nm in benzene. They concluded that the phenolate was the source of bioluminescence emission from aequorin. It was suggested that the Ca^{2+} -discharged obelin fluorescence is from the phenolate anion, as it is matched by the fluorescence of coelenteramide in certain basic apolar solvents (23).

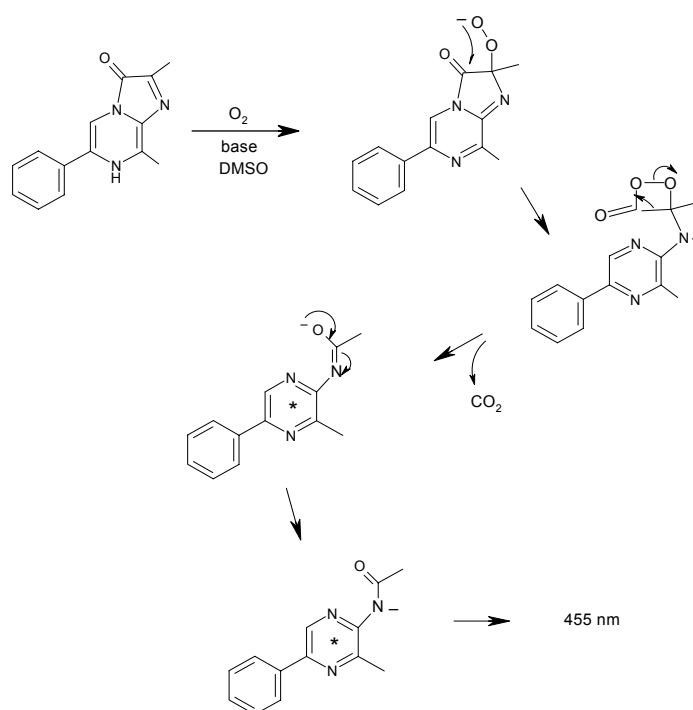
Scheme 1.1

Up to now, all these observations led to the generally agreed conclusion that in the bioluminescence systems involving coelenterazine, i.e., the photoproteins under consideration here and *Renilla* luciferase, as well as the related *Vargula* luciferase using a structurally similar luciferin, the amide anion is the primary chemical product generated in its excited state from which the bioluminescence emission occurs. The range of bioluminescence spectral maxima recorded, 462 nm for the *Vargula* luciferin-luciferase reaction to 495 nm for the Ca^{2+} -triggered bioluminescence of obelin from *O. geniculata* (14), can be attributed to perturbation of the singlet excited-state energy level of the amide anion by the properties of the binding site environment, such as its effective dielectric constant. However, the more recent results suggest that the amide anion, the phenolate, and the neutral forms of coelenteramide are all contenders for the emitting state in the several bioluminescent reactions utilizing coelenterazine or its analogues such as the luciferin in the *Vargula* bioluminescence system.

1.7 A simple model of coelenterazine chemiluminescence mechanism

Photoproteins presumably generate bioluminescence by a common mechanism as coelenterazine does in the chemiluminescence reaction (8). The chemical mechanism of this reaction has been solved partly by analogy with that of the firefly which has been thoroughly studied and well established, and by using chemiluminescence model reactions. McCapra and Chang (25) investigated the chemiluminescence of an imidazolopyrazine derivative analogous to coelenterazine. They put forward and proved this reaction mechanism as shown in Scheme 1.2 as a model for the bioluminescence systems that utilize structurally related substrates. In an aprotic polar solvent such as dimethyl sulfoxide (DMSO) and in the presence of a strong base (such as potassium t-butoxide) and oxygen, this analogue reacts with oxygen at this C2 position, producing peroxide anion, which cyclizes to give highly a unstable dioxetanone intermediate.

Scheme 1.2



The key step to produce a light-emitting species is the retro [2+2] cyclization reaction of the dioxetanone, generating carbon dioxide and a singlet-excited state of amide anion, which decays with the emission of the blue luminescence.

1.8 Mutational studies of photoproteins

Ca^{2+} -regulated photoprotein is distinctive in having an amino acid sequence with an unusually high content of cysteine, histidine and tryptophan residues (OL-obelin contains 5 cysteine, 5 histidine, and 6 tryptophan residues, for example), which are not commonly found in Ca^{2+} -binding proteins (15, 26). Site-directed mutagenesis and direct chemical modifications of photoproteins showed the importance of a proline residue at or near the C-terminus. Deletion or replacement of this proline residue destroys the luminescent capacity of photoproteins (27).

The mutations of cysteine residues suggest that cysteine plays an important role in the regeneration of aequorin but not in its catalytic activity (28). Site-directed mutagenesis of the five histidine residues has shown that substitution of His169 in aequorin (corresponding to His175 in obelin) to Ala, Phe, or Trp, leads to almost complete loss of activity, whereas modification of the remaining four histidine residues yielded mutant aequorins with varying bioluminescence activities (29). The tryptophan residues of both aequorin and obelin were replaced by the close analogue, phenylalanine. One of the mutants of aequorin, W86F, showed a striking change with an additional band at 400 nm in its bioluminescence emission maximum (30). The corresponding mutant of obelin, W92F (W92F-obelin), produced an even larger relative intensity of the 400 nm band similar to that of the major band (485 nm) of wild-type obelin, resulting in a violet color of bioluminescence (31, 32). In spite of the difference in bioluminescence spectra between W92F-obelin and wild-type obelin, the fluorescence spectrum of Ca^{2+} -discharged W92F obelin is green ($\lambda_{\text{max}} = 510 \text{ nm}$) the same as that of Ca^{2+} -discharged

wild-type obelin (31, 32), suggesting that the environment of coelenteramide bound in these two Ca^{2+} -discharged proteins is the same.

1.9 Early structural investigations of Ca^{2+} -regulated photoproteins

Analysis of structures of obelin from *O. longissima* (OL-obelin, space group $P6_2$) (33) and aequorin from *A. aequorea* (34), which were solved to resolutions 1.73 Å and 2.3 Å, respectively, reveals that photoprotein molecules are almost entirely helical, highly compact and globular (Fig. 1.5). The overall folding of the molecule is characterized by four sets of helix-loop-helix (HLH) structural motifs wrapping around to form a hydrophobic cavity that accommodates the ligand coelenterazine-oxygen complex. As with other Ca^{2+} -binding proteins, the four HLH motifs are arranged in pairs. Motifs I, III and IV are organized to form typical calcium ion binding sites, which are referred as EF-hand motifs. The loop region of motif II does not have the canonical sequence for calcium binding and is therefore not functional. The ligand-binding pocket (Fig. 1.6) is formed by residues originating from the helices of the four HLH motifs. Almost all residues forming the coelenterazine binding pocket are conserved among all Ca^{2+} -regulated photoproteins.

The peroxide substitution on the coelenterazine (Fig. 1.7) was claimed by Head et al., 2000 (34) in the aequorin structure, where weak electron density consistent with the presence of two oxygen atoms was found at the C2 position of coelenterazine, in accordance with the prediction from the earlier biochemical and chemical model studies (20). They also suggested that radiation damage could be responsible for the weak electron density of the peroxide substitution. A puzzle, however, was that in OL-obelin the electron density at the C2 position of coelenterazine can only fit a single oxygen atom at the 1.73 Å structure reported in Liu et al., 2000 (Fig. 1.7) (33). Both the peroxide in aequorin and the mono-oxygen in obelin are stabilized by hydrogen bonding to

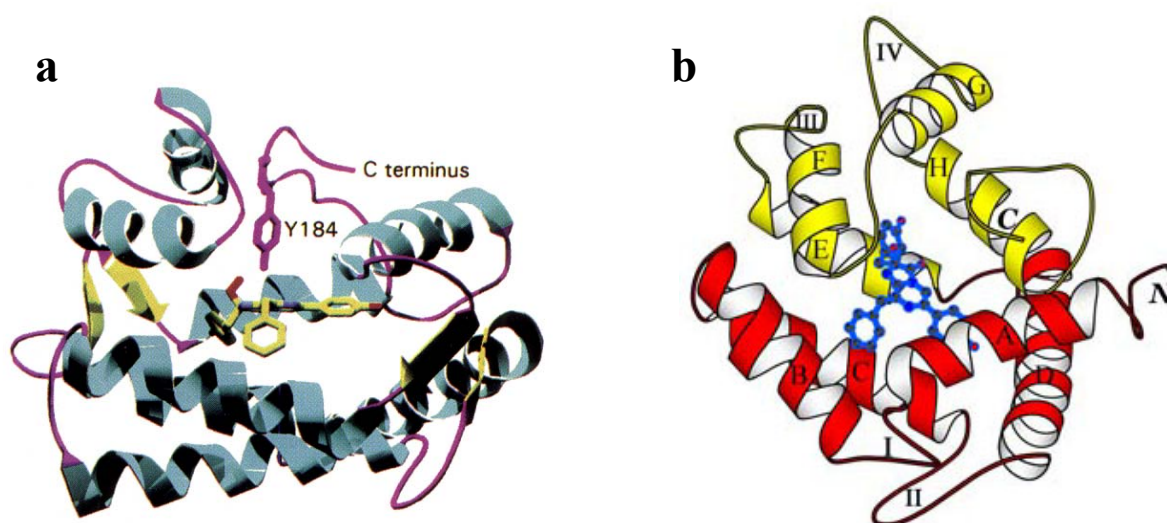


Fig. 1.5 Ribbon representations showing the crystal structures of aequorin (a) (34) and OL-obelin (b) (adapted from reference (33)). The coelenterazine-oxygen molecule is colored in blue. Both proteins are globular molecules containing a hydrophobic core cavity that accommodates the ligand. Two domains, N-terminal and C-terminal domains are well defined.

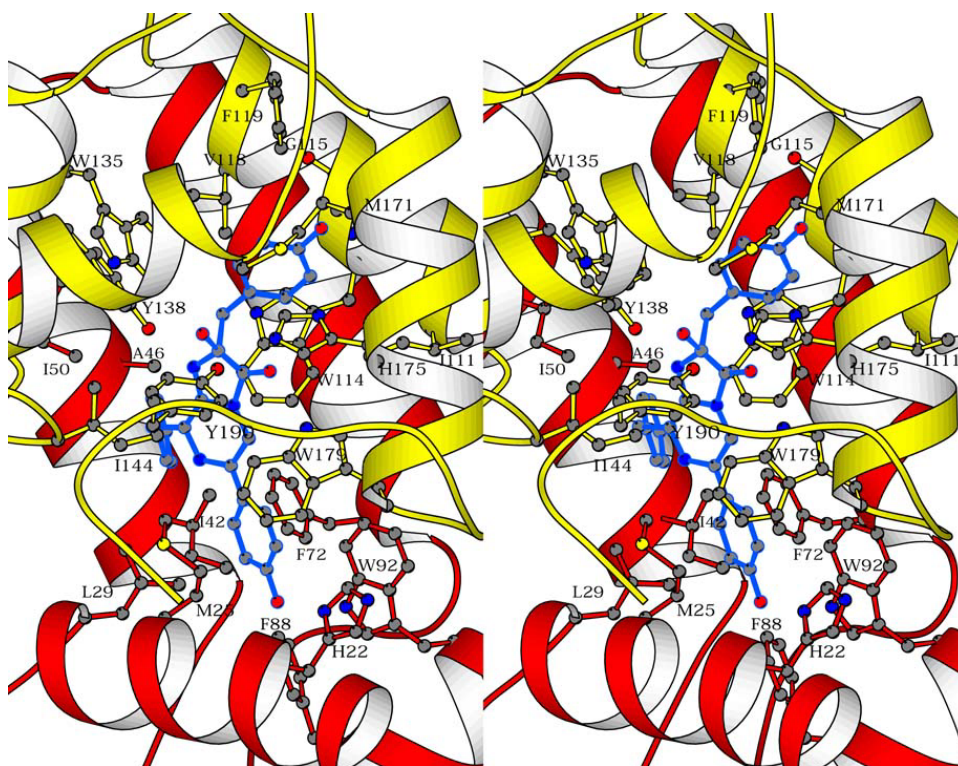


Fig. 1.6 A stereoview of the coelenterazine-oxygen binding pocket of OL-obelin showing residues within 4 Å of the coelenterazine-oxygen molecule. Atoms are colored by element type (gray, C; red, O; blue, N; yellow, S), with bonds colored by residue location (yellow for residues located in HTH motifs I or II; red for residues located in HTH motifs III or IV). Adapted from reference (33).

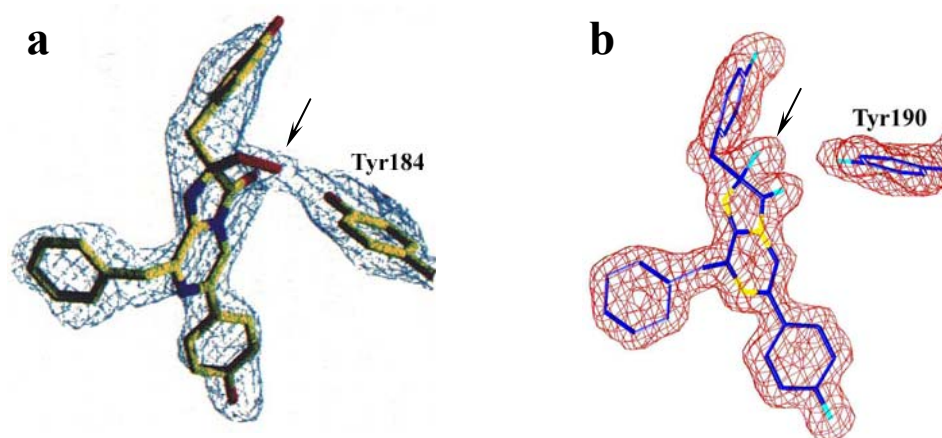


Fig. 1.7 Fit of the coelenterazine-oxygen molecules to the $2F_o - F_c$ electron density maps. (a): peroxidized coelenterazine (arrow) is indicated in the electron density map of aequorin contoured at 1.0σ . Adapted from reference (34). (b): In OL-obelin, the electron density next to the C2 position can only be fitted by a single oxygen atom (arrow), not by a peroxide group (33). The map is contoured at 1.0σ .

the phenolic oxygen of a tyrosine residue (Tyr184 in aequorin, Tyr190 in OL-obelin) which is itself hydrogen bonded to a histidine residue (His169 in aequorin, His175 in OL-obelin) (33, 34).

Attention is directed to the locations of cysteine, histidine, tryptophan and the C-terminal proline in order to explain the results from the mutagenesis study. His22, Trp92, Trp114, Tyr138, His175 and Trp179 in obelin, for example, (Fig. 1.6) make close hydrogen-bond contact with the coelenterazine-oxygen complex (33). Among these residues, His175 probably plays an essential role by assisting the formation of the transient anionic dioxetane in the course of the luminescence reaction. No cysteine residues are observed near the coelenterazine-oxygen complex. There are no disulfide bonds as well. The C-terminal proline appears to help to form a closed conformation, inside of which the coelenterazine-oxygen complex is contained.

1.10 Questions remain to be addressed

Progress has been made in understanding the mechanism of function of Ca^{2+} -regulated photoproteins since its first discovery in 1962, in which various approaches have been used involving biochemical, biophysical, and genetic techniques. However, questions and puzzles remain to be addressed in order to fully appreciate at atomic detail events following the reaction pathway. Furthermore, the mechanism needs to comply with a large body of existing mutational data, and more importantly, our understanding of the bioluminescence mechanism should predict events that are experimentally verifiable. A current hypothesis suggests three different conformations representing apo-, hydroperoxy-coelenterazine bound, and Ca^{2+} -discharged protein, respectively. Are there any other conformations existing along the reaction pathway? Are there structural differences between proteins with and without bound ligand, and with different ligands such as coelenterazine, coelenterazine analogues, coelenteramide and calcium ions? What are differences in ligand binding for different species of photoproteins? How is the

bound coelenterazine stabilized and what forces are involved? How do the protein structures provide an explanation that would account for spectral differences among different photoproteins? What role does Ca^{2+} play in triggering emission of bioluminescence? How is the protein involved in the bioluminescence reaction pathway? ...

1.11 Potential applications of our research

Both aequorin and obelin have found their way into scientific research as a calcium ion sensor in biological systems because of their light emission dependence upon calcium binding. They have now been used successfully inside a great many different types of living cells both to estimate intracellular calcium concentration and to study the role of calcium transients in the regulation of cellular function. Properties that make them suitable for scientific research are listed here: ease of signal detection, high sensitivity to free calcium ion concentration, relatively specific for calcium ion, tolerance for intracellular environment, free of toxicity. Among many similar photoproteins, only aequorin has been used widely, perhaps partly because it was the first photoprotein to be discovered but more significantly because of its more general availability. However, aequorin has a number of shortcomings, such as its slow response to follow some rapid intracellular calcium transients and its high sensitivity to physical concentrations of magnesium, all of which limit its utility. Obelin, on the other hand, responds considerably faster and is relatively insensitive to physical concentrations of magnesium, which motivated us to focus on studying obelin primarily.

Ca^{2+} -regulated photoproteins together with other fluorescent proteins such as GFP can be used as bioluminescent reporters for continuous monitoring and for noninvasive, real-time imaging of cellular events. This work will lead to: (a) the development of synchronous multiwavelength detection of expression of several genes in the cell; that can be very important

for studies of cell regulation pathways and for screening new drugs; (b) the measurement of transient bursts of calcium concentrations in various cell compartments simultaneously through expression of DNAs encoding photoprotein color mutants with different Ca^{2+} response time and appropriate target sequences directing these to the desired cell compartments; this is important for understanding the role of calcium as a second messenger in the regulation of cell events; (c) the development of multiwavelength Bioluminescence Resonance Energy Transfer (BRET) systems (a natural phenomenon observed in marine bioluminescent organisms in which an energy transfer occurs between a bioluminescent reaction donor and fluorescent acceptor proteins) for synchronous detection of several protein-protein interactions inside living cells; this is important in order to image an overall picture of cell events following activation of cell receptors; (d) the design of bioluminescence sensors capable of selectively detecting metal ions other than calcium by protein engineering.

1.12 Significance of this work

Our lab develops novel methods for crystallographic phasing, which are widely used for elucidating protein structures providing their functional implications. X-ray diffraction experiments are capable of capturing snap-shots of molecular conformations trapped in a crystal and revealing structural details at near atomic resolution. We follow the approach of crystallizing proteins trapped in different conformations that are relevant to its functional pathway and identifying structural features unique to particular conformations. Such an approach could not only prove or disprove the existence of reaction intermediates that are proposed by an existing hypothesis but also potentially identify new intermediates and therefore modify or change an existing hypothesis.

The following structures have been determined in the course of my studies: Ca^{2+} -regulated obelin from *O. geniculata* (OG-obelin) at 1.82 Å, apo-obelin from *O. longissima* charged with an analog substrate, coelenterazine h as the bound ligand (h-obelin) at 1.14 Å resolution, the W92F mutant of obelin from *O. longissima* (W92F-obelin) at 1.8 Å, the Ca^{2+} -discharged W92F-obelin at 1.96 Å, the Ca^{2+} -loaded apo-proteins for both aequorin and obelin at 1.7 Å and 2.2 Å, respectively. I also obtained the structure of OL-obelin in a different crystal form (space group C2) that has significantly higher diffraction limit (1.0 Å). The new form of OL-obelin crystal was exposed to a trace of Ca^{2+} and the resolution limit was further extended to 0.97 Å.

The X-ray crystallographic studies of photoproteins supports the HSQC-NMR spectroscopy study of OL-obelin which demonstrates five possible conformations (35), three of which have been confirmed by our work. Our work demonstrates that conformations of photoproteins are controlled by binding of various ligands such as Ca^{2+} , coelenterazine, and coelenteramide. Crystallographic work demonstrates unambiguously that the hydroperoxy group does exist in OL-obelin, as well as in other species of photoproteins, and molecular forces that stabilized the hydroperoxy moiety have been elucidated.

New insights are obtained into how the protein environment could induce different ionic states of the bound coelenteramide whose excited states are responsible for the different spectra of photoproteins. Structural work suggests that the bioluminescence originates from the coelenteramide phenolate ion-pair excited state. The differences in the bioluminescence and fluorescence spectra among different species of Ca^{2+} -regulated photoproteins could be a result of different polarity and by tuning the effective dielectric constant of the chromophore binding site. A proton relay mechanism of the calcium triggered coelenterazine decarboxylation of photoproteins is also proposed, in which conformational adjustment in the loops of EF-hand

motifs induced by calcium binding can be expected to influence the hydrogen bond donor-acceptor relationship around the coelenterazine to produce the peroxy anion for the initiation of bioluminescence.

The structure of Ca^{2+} -loaded apo-aequorin was solved by the single-wavelength anomalous scattering (SAS) method using diffraction signals of weak anomalous scatterers, sulfur and calcium. We overcame the challenge of collecting diffraction data with very long wavelength of chromium radiation and successfully determined the structure with only one complete data set from a native crystal, which greatly simplifies the crystal preparation, data collection procedures. This new approach of structure determination pushes X-ray crystallographic methodology to a higher level and could be of general interest.

CHAPTER 2

Experimental Procedures and Methods

2.1 Preparation of photoprotein samples

2.1.1 Expression, purification of Ca^{2+} -regulated photoproteins

High-purity recombinant photoproteins: OL-obelin (from *Obelia longissima*, 8.44 mg/ml), OG-obelin (from *Obelia geniculata*, 7.86 mg/ml), W92F-obelin (mutant of obelin from *Obelia longissima*, 8.92 mg/ml), h-obelin (apo-obelin from *Obelia longissima* charged with an analog substrate, coelenterazine h as the ligand, 7.93 mg/ml), W179F-obelin (mutant of obelin from *Obelia longissima*, 8.49 mg/ml) and aequorin (from *Aequorea aequorea*, 7.97 mg/ml) in the buffer of 1 mM EDTA, 10 mM potassium/sodium phosphate, pH 7.3, were provided by Dr. Eugene S. Vysotski from our collaborator Dr. John Lee's laboratory. Site directed mutagenesis was used to delete six amino acid residues from the N-terminus of apo-aequorin, and to construct the site-directed variants, W92F-obelin and W179F-obelin. The protocols used for protein production can be found in the reference papers of Illarionov et al. (36); Markova et al. (14), Deng et al. (31); Vysotski et al. (32).

2.1.2 Preparation of Ca^{2+} -discharged photoproteins

Since photoproteins were prepared in 10 mM potassium/sodium phosphate buffer, in order to prevent the precipitate formation of calcium phosphate during the discharging process, the protein solutions were first diluted ten times with 10 mM Bis-Tris solution pH 7.0. Then pre-prepared 1 M CaCl_2 solution in water was gradually added into the diluted protein solution with stirring at room temperature. The final concentration of calcium ions was 1mM. During this

procedure a bright blue bioluminescence (violet bioluminescence for W92F-obelin) was observed. After the bioluminescence emission ceased, the yellow protein solutions had turned colorless indicating that coelenterazine was converted into coelenteramide. To test the presence of bound coelenteramide, the final products were excited with near UV. The obelin product solutions displayed green fluorescence; the aequorin product solution displayed blue fluorescence.

The product solutions were transferred to Millipore filters (Biomax-5K NMWL membrane, Bedford, MA). Centra[®]-MP4R centrifuge from the International Equipment Company was pre-cooled to 4 °C and used to concentrate the protein. The product solutions were concentrated to their original volume with the speed of 4500 rpm. For changing the buffer from potassium/sodium phosphate to Bis-Tris buffer purpose, the product solutions were diluted ten fold with pre-mixed 1 mM calcium chloride in 10 mM Bis-Tris pH 7.0 solution; and then concentrated to its original volume again. This dilution-concentration procedure was repeated for four times in order to guarantee that the original buffer solution has been totally replaced. Finally the product solutions were filtered with 0.22 µm cellulose acetate Costar[®] centrifuge tube filter.

The product protein concentrations were measured by UV-visible Spectrophotometer (BioMate) using the Bradford method with chicken albumin in 1 mM CaCl₂, 10 mM bis-Tris pH 7.0 as a standard. The protein concentrations used in the final optimized crystallization conditions are: OL-obelin (21.04 mg/ml), OG-obelin (13.4 mg/ml), W92F-obelin (28.29 mg/ml), h-obelin (24.04 mg/ml), W179F-obelin (20.49 mg/ml) and aequorin (40.65 mg/ml).

2.2 Crystallization of photoproteins

2.2.1 Crystallization of obelins

The crystals of OL-obelin (space group C2), OG-obelin, W92F-obelin and h-obelin were provided by Dr. Eugene S. Vysotski. The hanging-drop vapor diffusion technique was used. All crystallization conditions contain 0.1 M hexaminecobaltic chloride as additive. The droplet included 9 μ l of protein, 5 μ l precipitant solution, and 1 μ l of additive. Obelin crystals grew as secondary light yellow cubic crystals on the surface of brown salt crystals and had a size about $0.5 \times 0.5 \times 0.45$ mm after 5-10 days at 4°C. Crystal of OG-obelin is shown in Fig. 2.1 as an example of obelin crystals. The crystallization conditions are:

OL-obelin (space group C2): 23% PEG 8000, 50 mM KH_2PO_4 pH 5.8;

OG-obelin: 25% PEG 8000, 50 mM KH_2PO_4 , pH 5.9;

W92F-obelin: 18% PEG 8000, 50 mM KH_2PO_4 , pH 6.0;

H-obelin: 16% PEG 8000, 50 mM KH_2PO_4 , pH 6.0.

2.2.2 Crystallization of Ca^{2+} -discharged W92F-obelin, Ca^{2+} -loaded apo-aequorin and Ca^{2+} -loaded apo-obelin

Preliminary screening of crystallization conditions was performed for all Ca^{2+} -discharged photoproteins using the sparse-matrix sampling method (37) with commercially prepared reagents from both Hampton Research (Laguna Niguel, CA) (Crystal screen, Crystal screen 2, and MembFac crystallization screening kits) and Emerald Biostructures (Bainbridge Island, WA) (Wizard I and Wizard II crystallization screening kits). The modified microbatch method (38, 39) was carried out using an ORYX 1-6 protein crystallization robot from Douglas Instruments Ltd (East Garston, UK) for screening initial crystallization conditions. Equal volumes (0.5 μ l) of the protein and the crystallization solutions were mixed in the wells of a Nunc HLA plate. The

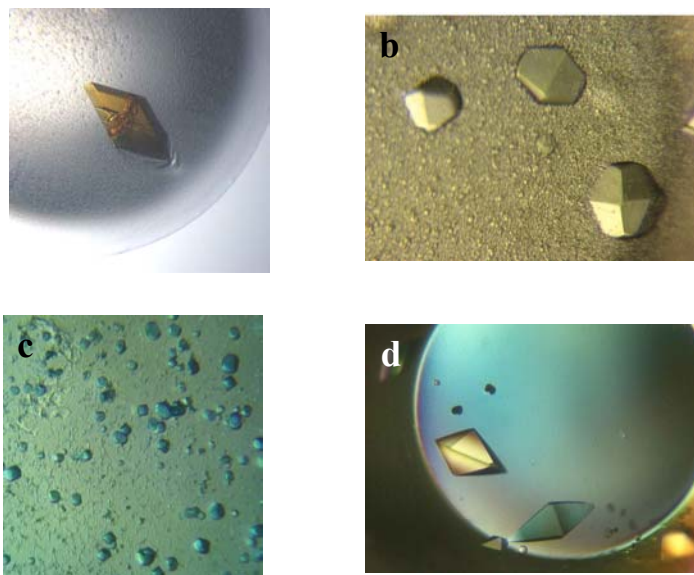


Fig. 2.1 Crystal pictures of photoproteins. (a) Crystal of OG-obelin from *O. geniculata*. The crystal grew as a secondary light yellow cubic crystal on the surface of brown salt crystals. (b) Crystals of Ca^{2+} -discharged W92F-obelin from *O. longissima*. The yellow color of the crystal and droplet is because of high concentration of the protein. (c) Crystals of Ca^{2+} -loaded apo-obelin from *O. longissima*. The crystals were stained with Izit crystal dye (Hampton Research, Laguna Niguel, CA). (d) Crystals of Ca^{2+} -loaded apo-aequorin from *A. aequorea*. The color is caused by polarized lens.

mixed solutions were covered with 10 μ l of paraffin oil. After all the crystallization droplets have been set up, the HLA plate was sealed with 4 ml of a combination of silicon and paraffin oil (7:3). Two incubation temperatures, 4 °C and 18 °C have been tried for each screening experiment. Low protein concentrations of Ca^{2+} -discharged obelins and aequorin (8 – 10 mg/ml) were used for the first try without success. Then the protein concentrations were doubled. 240 different conditions have been screened in total for each concentration of each protein sample. The results are: two hits for Ca^{2+} -discharged W92F-obelin, two hits for Ca^{2+} -discharged h-obelin and no hits for Ca^{2+} -discharged OL-obelin, Ca^{2+} -discharged OG-obelin, and Ca^{2+} -discharged W179F-obelin.

However, it took one month for the Ca^{2+} -discharged W92F-obelin and Ca^{2+} -discharged h-obelin crystals to show up with a relatively small size in a precipitation droplet. The success in obtaining crystals was also variable under the same condition and by the time the crystals appeared, the droplet was almost dried out; all crystals were covered by a heavy skin. Crystallization optimization for improving the crystals' size and quality was carried out manually for each hit of both Ca^{2+} -discharged obelins, including precipitant concentration screening, protein concentration screening, additive screening and pH screening. The best condition for Ca^{2+} -discharged W92F-obelin (Fig. 2.1) is 1.5 M tri-sodium citrate in the buffer of 0.1 M HEPES-sodium pH 7.5 at a temperature of 4 °C. The best condition for Ca^{2+} -discharged h-obelin (Fig. 2.1) is 20% w/v polyethylene glycol 10,000 in the buffer of 0.1 M HEPES pH 7.5 at a temperature of 18 °C. The crystals showed up within two weeks, no skin cover, with a maximum size of about $0.05 \times 0.1 \times 0.2$ mm. Then different crystallization methods, such as hanging-drop vapor diffusion, sitting-drop vapor diffusion, micro-seeding, macro-seeding and capillary

diffusion method were tried to increase the crystal size. Only sitting-drop vapor diffusion method succeeded in increasing the crystal size to 0.4 x 0.4 x 0.3 mm.

After the failure of the first crystallization try with low protein concentration of Ca^{2+} -discharged aequorin, high concentrations up to 40.65 mg/ml protein were used for the second try, which gave beautiful crystals (Fig. 2.1) under one week or less of incubation at 4°C. The condition is 0.02 M calcium chloride, 30% v/v 2-methyl-2,4-pentanediol and 0.1 M sodium acetate pH 4.6. The maximum size is about 0.35 x 0.3 x 0.25 mm. Crystals are very easily reproduce.

To be sure that coelenteramide is still bound in the protein crystal, the crystal fluorescence of the Ca^{2+} -discharged W92F-obelin, Ca^{2+} -discharged h-obelin and Ca^{2+} -discharged aequorin, were examined. A single crystal was transferred to a droplet containing only the crystallization mother liquor. The green fluorescence from the Ca^{2+} -discharged W92F-obelin crystal observed under a stereomicroscope with excitation at 350 nm (Fig. 2.2) clearly indicates that coelenteramide is bound within the protein. However, no fluorescence was observed for Ca^{2+} -discharged h-obelin and Ca^{2+} -discharged aequorin, indicating that the crystals do not contain coelenteramide. The analysis of the crystallization conditions shows that the precipitants are polyethylene glycol 10,000 and 2-methyl-2, 4-pentanediol respectively; both can serve as organic solvent to extract coelenteramide out of the protein. It is known that coelenteramide does fluorescence in organic solvent but not with the presence of water. Since both the protein solution and the precipitants in the protein crystallization droplets contain water, no fluorescence should be observed for the mixture of protein solution and the precipitants. In order to verify this, crystallization solutions were added to Ca^{2+} -discharged h-obelin and Ca^{2+} -discharged aequorin protein solutions respectively while the protein solutions were being stimulated by near UV light, gradually the

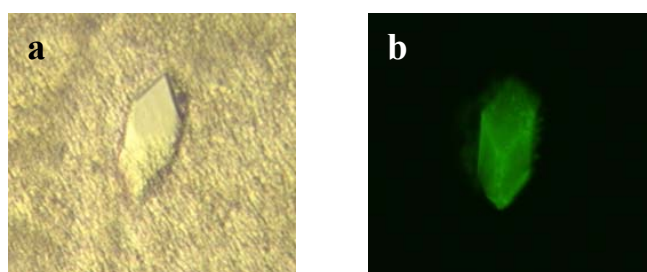


Fig. 2.2 (a) Crystal of Ca^{2+} -discharged W92F-obelin; (b) green fluorescence of the crystal on excitation by near UV.

fluorescence color from both protein solutions disappeared. Hence these two crystals should contain only calcium and apo-proteins which were confirmed by later structural studies. We suggest calling them Ca^{2+} -loaded apo-aequorin and Ca^{2+} -loaded apo-obelin.

2.3 Calcium soaking experiment

The soaking of the OL-obelin (C2 space group) crystals was carried out by exposing them to traces of Ca^{2+} insufficient to cause visible bioluminescence. The hanging-drop vapor diffusion method was used to grow the obelin crystals. The exposure of some crystals to a trace of Ca^{2+} was done by placing the cover slip with a droplet containing crystals onto a stage for monitoring under a stereomicroscope, and carefully inserting 1 μl of a 50 mM CaCl_2 solution into one side of the droplet. Some crystals at the point of entry of the CaCl_2 glowed. The cover slip was then put back over its original reservoir in a Linbo box and allowed to stand overnight at 4 °C. On reexamination, crystals located at the point of entry of the CaCl_2 were dissolved but those on the far side of the droplet were intact and retained a yellow color indicating that the coelenterazine substrate was unreacted. Then the yellow crystal was scooped up and flash-frozen in liquid nitrogen for later data collection. The crystal retains a yellow color after soaking and during later data collection.

2.4 Data collection and processing of photoproteins

The information on data collection and processing statistics, space group, unit cell dimensions of all the crystals is listed in Table 2.1.

2.4.1 Data collection of obelins

30% glycerol was used as cryoprotectant to soak all obelin crystals for several seconds before crystals were mounted to a fiber loop (40) and flash-frozen to 100 K. Complete data sets for OG-obelin and W92F-obelin were collected at the beamline 17-1D in the facilities of the Industrial

Table 2.1 Data collection and processing statistics

	OG-obelin	W92F-obelin	h-obelin	OL-obelin	Soaked OL-obelin	Ca ²⁺ -discharged W92F-obelin	Ca ²⁺ -loaded apo-aequorin* (Cr-X-rays)	Ca ²⁺ -loaded apo-aequorin [♦] (synchrotron)	Ca ²⁺ -loaded apo-obelin
Wavelength (Å)	0.94	0.94	0.97	0.97	0.97	1.5418	2.2909	0.97	0.97
Space group	P2 ₁ 2 ₁ 2 ₁	P4 ₁ 2 ₁ 2	C2	C2	C2	P4 ₁ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₃ 2 ₁ 2	P4 ₁ 2 ₁ 2
Unit cell									
a (Å)	33.5	53.4	83.0	83.1	83.3	53.4	54.4	54.4	58.6
b (Å)	67.8	53.4	54.3	54.1	54.6	53.4	54.4	54.4	58.6
c (Å)	69.2	144.5	52.4	52.5	52.7	144.0	135.1	135.1	110.4
β (°)			112.2	112.1	112.3				
Resolution	20-1.82/	30-1.7/	20-1.17/	50-1.03/	50-0.97/	20-1.96/	50-2.5/	50-1.7/	50-2.2/
Overall/outer (Å)	1.86-1.82	1.76-1.7	1.21-1.17	1.07-1.03	1.02-0.97	2.05-1.96	2.59-2.5	1.76-1.7	2.28-2.2
No. of reflections observations/unique	168180/14015	266085/21633	155308/61453	206351/106356	229006/127376	235255/15775	98825/7537	189216/22954	110264/18694
Completeness Overall/outer (%)	95.1/93.9	93.5/87.7	84.7/65.7	98.7/93.8	93.0/84.4	99.5/99.2	99.5/95.2	98.8/91.7	99.8/99.4
R _{merge} [#] Overall/outer (%)	3.5/21.9	7.6/35.0	4.5/24.9	4.6/16.2	3.0/27.0	6.0/18.8	5.3/11.2	3.8/27.0	7.4/25.1
I/σ _I Overall/outer	28.2/3.8	18.1/3.0	20.1/3.4	25.8/5.2	25.1/2.7	9.37/2.86	40.5/13.7	45.9/4.7	20.4/4.7

* Listed here is the data set from which the phasing result was chosen for later model building.

♦ This data set was used in the final structure refinement.

[#] $R_{\text{merge}} = \sum_{hkl} [\sum_i (|I_{hkl,i} - \langle I_{hkl} \rangle|)] / \sum_{hkl,i} \langle I_{hkl} \rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with the Miller indices h , k and l and $\langle I_{hkl} \rangle$ is the mean intensity of that reflection.

Crystallography Association Collaborative Access team (IMCA-CAT) at the Advanced Photon Source (APS) with resolution of 1.82 Å and 1.72 Å respectively. The atomic resolution data sets for OL-obelin, soaked OL-obelin and h-obelin were collected at the beamline 5.0.3 at the Advanced Light Source in the Ernest Orlando Lawrence Berkeley National Laboratory with resolution of 1.03 Å, 0.97 Å and 1.17 Å respectively. The wavelength was 1.0 Å. The anomalous data sets of soaked OL-obelin at the wavelength of 1.74 Å and 1.59 Å respectively were collected at beamline 17-1D in the facilities of IMCA-CAT at APS. Data were processed with program HKL2000 (41).

2.4.2 Data collection of Ca²⁺-loaded apo-aequorn

As is well known, the stronger the anomalous signal is, the greater the probability of success is in locating the anomalous scatterers and in producing interpretable electron density maps. The simplest way to increase the anomalous signal of a protein is to introduce atoms having a large anomalous signal ($\Delta f''$) into the lattice either by soaking in heavy atom salts containing platinum, mercury or by chemical modification such as iodination or the use of selenomethionine derivatives. The degree of success of this approach highly depends on how well the heavy atom can incorporate in the crystal and on how isomorphic the heavy atom derivative and the native crystals are. The preparation of the protein samples or crystals is often the limitation of SAS method and other methods such as multiple-wavelength anomalous diffraction (MAD), and multiple isomorphous replacement (MIR), which also use heavy atom position to estimate protein phases.

Most proteins contain sulfur atoms and many proteins contain calcium ions, thus it would be very convenient and efficient that the sulfur and calcium signals could be used alone to solve protein structures (42). 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. The fact that nowadays, advances in X-ray optics and detector technology can measure diffraction data with much higher precision makes it possible to obtain anomalous signals of weak anomalous scatterers, such as sulfur and calcium. To take advantages of calcium ions and sulfur atoms contained in the Ca^{2+} -loaded apo-aequorin crystal, the single-wavelength anomalous scattering (SAS) method was used in solving protein phases.

The main problem in collecting weak anomalous scattering data is the signal-to-noise level of data. This problem arises because the measured anomalous signal, the difference between Bijvoet pairs ($|F_+|$ and $|F_-|$), is generally about one order of magnitude smaller than the difference between $|F_H(hkl)|$ and $|F_P(hkl)|$. Therefore, the signal-to-noise level in the data plays a critical role in the success of the SAS phasing. As we all know, synchrotron X-ray diffraction data has much better quality in a lot of respects, but the availability of synchrotron X-ray source is always limited plus the cost and the risk of crystal transportation, so collecting data in-house is a good alternative. The most common in-house X-ray source is copper radiation. However, the anomalous scattering signals ($\Delta f''$) of the weak anomalous scatterers, sulfur (0.56 electrons) and calcium (1.29 electrons) are comparatively weak at the copper X-ray wavelength (1.5418 Å) (Table 2.2). Thus special care must be taken to enhance the anomalous scattering signal as well as to minimize errors in the experiment. In order to maximize sulfur and calcium anomalous signals and accurately measure them, a chromium X-ray source with a longer wavelength (2.2909 Å) is used in our lab. The anomalous signals of sulfur (1.14 electrons) and calcium (2.53 electrons) are obviously increased (Table 2.2). Nevertheless, using longer wavelength X-rays

Table 2.2 X-ray anomalous scattering factors for S and Ca atoms

Atom	Atomic number	Cu-X-rays ($\lambda = 1.5418 \text{ \AA}$)		Cr-X-rays ($\lambda = 2.2909 \text{ \AA}$)	
		f'	f''	f'	f''
S	16	0.33	0.56	0.39	1.14
Ca	20	0.37	1.29	-0.21	2.53

introduces crystal absorption errors. Flushing the incident beam path with helium dramatically minimizes the problems associated with using longer wavelength X-rays.

All the crystals were directly mounted to a fiber loop without any cryoprotectant and flash-frozen to 100 K before the data collection process. Special care had to be taken since the crystals of Ca^{2+} -loaded apo-aequorin are so fragile that a small disturbance caused by scooping the crystal out of the droplet would increase its mosaicity dramatically. Five complete data sets from five different crystals with 720 frames of 0.5° oscillation were collected for *ab initio* single-wavelength anomalous scattering (SAS) phasing using an in-house Rigaku rotating chromium-anode X-ray generator operated at 50 kV and 90 mA with an Rigaku R-AXIS IV image plate detector. The chromium X-rays were aligned/focused by Rigaku/MSB confocal optics CMF15-50Cr8 and a cone-shaped chamber filled with helium gas was attached to the front of the image plate to reduce air absorption at the longer wavelength. The maximum resolution the crystals can get is 2.5 Å because of the instrument's limitation. Data used for refinement were collected to 1.7 Å resolution at the beamline 22-ID in the facilities of the South East Regional Collaborative Access team (SER-CAT) at APS with the wavelength of 0.97 Å. The shorter wavelength was chosen here in order to decrease the radiation damage to the crystal, to reduce errors due to absorption and to obtain higher resolution data. Data were processed separately with program HKL2000 (41).

2.4.3 Data collection of Ca^{2+} -loaded apo-obelin and Ca^{2+} -discharged W92F-obelin

Both crystals were directly mounted to a fiber loop without any cryoprotectant and flash-frozen to 100 K before the data collection process. One data set of Ca^{2+} -loaded apo-obelin was collected to 2.2 Å at the beamline 22-ID in the facilities of SER-CAT at APS with the wavelength of 0.97 Å. Data were processed with program HKL2000 (41). A 360° data set of

Ca²⁺-discharged W92F-obelin was collected at the in-house copper X-ray source with Bruker Smart 6000 CCD detector. An 8° of 2θ was used for the purpose of extending the data resolution to 1.96 Å. The data were processed and scaled by Bruker program Proteum.

2.5 Phasing of photoproteins

There are several different methods to determine protein structure phases. The most general methods in macromolecular phasing are multiple-wavelength anomalous dispersion (MAD), single-wavelength anomalous scattering (SAS), molecular replacement (MR), single isomorphous replacement (SIR) and multiple isomorphous replacement (MIR). Different approaches have different requirements on crystal quality, crystal contents, data collection techniques and data processing procedures. Based on the quality and characteristics that photoprotein crystals have and the specialty of our laboratory, SAS using weak anomalous scatterers, and MR methods were chosen to obtain the initial phase information.

2.5.1 Phasing of obelins

The molecular replacement method seeks to solve the crystallographic phase problem by exploring molecular redundancy within the same crystal or between different crystal forms, or by applying known structural knowledge to an unknown structure. The phase for OG-obelin, W92F-obelin, h-obelin, OL-obelin (C2 space group) and soaked OL-obelin were determined by the molecular replacement method using OL-obelin (P6₂ space group; PDB entry code 1EL4) as a search model. The orientational and positional parameters in the unit cell were obtained using program CNS1.0 (43). The initial models were examined by the program O (44) to inspect the crystal packing and model fitting. A further search model was made by deleting the N-terminal residues because of the unfavorable contacts between symmetrically related neighboring

molecules for all the obelins; truncating all non-identical side chains to alanine was made for the model used in OG-obelin.

2.5.2 Phasing of Ca^{2+} -loaded apo-aequorin

The first try was to use the low resolution (3.0 Å) portion of the SAS data to locate the three calcium ions. However, the calcium anomalous signals alone could not give a solution. The sulfur anomalous signals were added into the calculation. Eleven sites, three calcium ions plus eight sulfur atoms, were requested. Five different data sets were used in the calculation independently with the program SOLVE & RESOLVE (45, 46). The results are summarized in Table 2.3. The one that has the most number of atoms traced was chosen for later model building. Eight sites, three calcium and five sulfur sites, were identified and refined by Bijvoet difference Fourier analysis using the program SOLVE (Table 2.4) (45). The mean figure-of-merit (FOM) is 0.33. These nine sites were used to estimate the protein phases at 2.6 Å resolution using RESOLVE (46). The phased chromium X-ray data were then subjected to density modification also using the program RESOLVE (46). 684 atoms were successfully automatically traced, including some side-chain atoms. The map obtained from the chromium X-ray data after SOLVE and RESOLVE showed clear electron density for most of the chain, including significant side-chain density (Fig. 2.3). On the basis of the electron density map and the information provided from the automatically traced model pieces, the chain was readily built up from residues 11 to 191 using program XTALVIEW (47).

2.5.3 Phasing of Ca^{2+} -loaded apo-obelin and Ca^{2+} -discharged W92F-obelin

The orientation and position parameters of the Ca^{2+} -loaded apo-obelin molecules in the unit cell were determined by the molecular replacement method using CNS1.0 (43). The search model used in the calculation was W92F-obelin with some truncations. The cross-rotation

Table 2.3 Phase searching results of Ca^{2+} -loaded apo-aequorin using in-house chromium X-rays data sets (11 anomalous scatterer sites (3 calcium sites plus 8 sulfur sites) were requested for calculation)

Data sets	Resolution* (Å)	Results from Solve & Resolve			
		Z score	Sites found	Mean FOM [♦]	No. of atoms traced
1	3.0	12.30	5	0.34	548
1 [♥]	2.6	17.96	9	0.33	684
2	3.0	17.20	8	0.35	626
2	2.6	11.99	8	0.33	671
3	3.0	9.80	8	0.33	376
3	2.6	10.50	8	0.32	435
4	3.0	12.99	3	0.30	391
4	2.6	11.30	3	0.27	419
5	3.0	13.20	10	0.31	314
6	2.6	9.60	5	0.14	131

* Data below this resolution were used to locate calcium and sulfur atom positions.

♦ Figure-of-merit (FOM) = $\langle \cos \Delta\alpha_j \rangle$, where $\Delta\alpha_j$ is the phase angle error for the j^{th} phase angle.

♥ This result was chosen for later model building.

Table 2.4 Refined positions (in fractional coordinates) of the calcium and sulfur atom sites used in phasing

Site	Atom	X	Y	Z	Occupancy	B-factor
1	Ca	0.4793	0.6142	0.0433	0.8086	32.6638
2	Ca	0.3041	0.7252	0.0204	0.7288	37.6391
3	Ca	0.5391	0.7399	0.0395	0.2970	22.6039
4	S	0.5735	0.2933	0.0461	1.0238	59.7862
5	S	0.5725	0.0468	0.0421	0.6294	60.0000
6	S	0.8015	0.5143	0.0372	0.5811	60.0000
7	S	0.5534	0.1489	0.0898	0.6244	60.0000
8	S	0.6105	0.1716	0.1111	0.4447	60.0000
9	S	0.1898	0.8126	0.0735	0.2654	60.0000

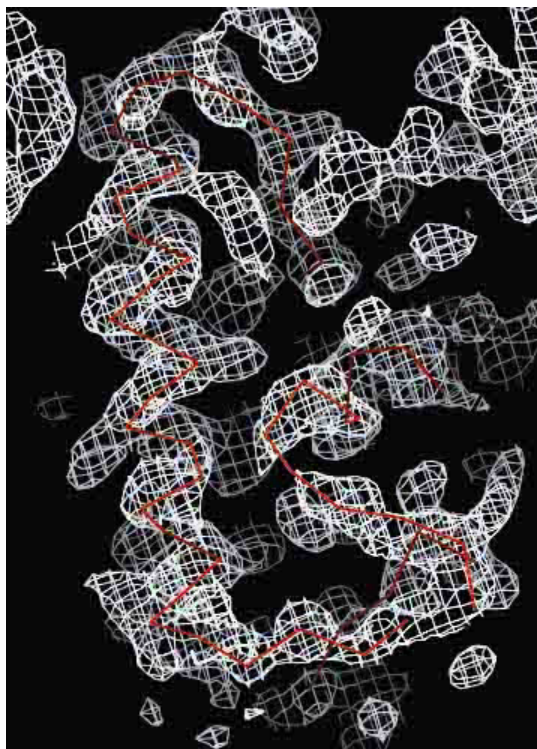


Fig. 2.3 The very first experimental electron density map at 3.0 Å based on the phases obtained from weak anomalous signals of Ca^{2+} and sulfur overlapped with part of peptide backbone (red) of Ca^{2+} -loaded apo-aequorin. The side chain density can be clearly located. The further model building was done readily.

function and translation function search did not give outstanding result. The solution which is slightly better than others was chosen to do the rigid-body refinement using RefMac 5.0 (48) resulted in an R factor of 53.4%, R-free factor of 56.9%. The graphic program XTALVIEW (47) was used to observe the crystal packing and the initial electron density map. The $2F_o - F_c$ map showed that part of the protein density is continuous with some side-chain electron density. However, the model does not match the electron density well.

Since the space group of both Ca^{2+} -discharged W92F-obelin and W92F-obelin is $P4_12_12$, direct rigid body refinement was tried using Refmac5 (48), resulting in R factor of 30.3%, R-free factor of 30.7%, indicating that there is no big overall conformational change between W92F-obelin and Ca^{2+} -discharged W92F-obelin. The electron density of coelenteramide was clearly shown in the center of the protein molecule. Based on the model of coelenterazine, the stereo model of coelenteramide was built and gradually fitted into the electron density. The graphic work was done using the program XTALVIEW (47).

2.6 Structure refinement of photoproteins

X-ray crystallographic refinement is both a mathematical and an empirical procedure that attempts to minimize errors in an atomic model and therefore reduces the difference between calculated and observed structure factors. Small observation to parameter ratio and less accurate diffraction data make the macromolecular structure refinement difficult to do. Constraints and restraints obtained from stereochemical knowledge are used to increase the observation to parameter ratio, hence improve the refinement. Refinement also involves manual interventions. The final refinement statistics and PDB deposition accession codes of all protein structures are shown in Table 2.5.

Table 2.5 Refinement statistics and PDB deposition accession codes

	OG-obelin	W92F-obelin	h-obelin	OL-obelin	Soaked OL-obelin	Ca ²⁺ - discharged W92F- obelin	Ca ²⁺ -loaded apo-aequorin	Ca ²⁺ - loaded apo- obelin
Resolution range (Å)	20.0 – 1.82	20.0 – 1.72	10.0 – 1.17	10.0 – 1.03	10.0 – 0.97	50.0 – 1.96	50.0 – 1.7	50.0 – 2.2
R _{work} value *	0.21	0.21	0.15	0.15	0.14	0.22	0.22	0.19
R _{free} value	0.26	0.24	0.19	0.17	0.17	0.26	0.24	0.24
Free R value test set (%)	8	8	8	8	8	5	5	5
RMSDs from ideality								
Bond lengths(Å)	0.006	0.005	0.014	0.015	0.016	0.012	0.019	0.015
Bond angles (°)	1.1	1.1		2.2	2.1	1.2	1.6	1.3
Mean B-factor (Å ²)	24.1	22.2	15.2	12.1	10.3	16.4	30.4	27.2
No. of protein atoms	1636	1452	1518	1553	1553	1494	1460	1402
No. of solvent atoms	111	212	211	268	227	172	159	128
No. of substrate atoms	34	34	34	34	34	31	0	0
No. of heterogen atoms	0	0	0	9	10	8	3	3
PDB accession code	1JF0	1JF2	1SL9	1QV0	1QV1	1S36	1SL8	1SL7

$$* R = \sum |F_{\text{obs}}| - k |F_{\text{calc}}| / \sum |F_{\text{obs}}|$$

2.6.1 Refinement of obelins

2.6.1.1 Refinement of OG-obelin and W92F-obelin

The refinements of OG-obelin and W92F-obelin were carried out using CNS1.0 (43) against 1.82 Å and 1.72 Å resolution data sets respectively. 8% of the reflections were randomly selected for cross-validation analysis by calculating the free R-factor to follow the progress of refinement (49). The first step was based on rigid-body refinement. Refinement of the atomic positions was performed by minimization refinement and simulated-annealing. The composite omit maps were used to assist in building the structure. During the refinement procedure, the side chains were adjusted stepwise, and the truncated side chains were rebuilt. $F_o - F_c$ maps were calculated to aid the rebuilding of the truncated N-terminal. Residue 1 to 3 in OG-obelin and residues 1 to 4 in W92F-obelin were disordered. Several cycles of minimization and annealing refinement, followed by water pick-up and B-factor refinement, resulted in a model with R-factor and free R-factor of 21.2 % and 26.2 % for OG-obelin, 21.3 % and 24.2 % for W92F-obelin. Analysis of the Ramachandran plot (PROCHECK) (50) of both models showed no residues in disallowed or generally allowed regions.

2.6.1.2 Refinement of atomic resolution structures of h-obelin, OL-obelin(C2), soaked OL-obelin

The models of OL-obelin (C2 space group), soaked OL-obelin and h-obelin were initially refined against a 1.2 Å data set using CNS1.0 (43). After rigid-body refinement, minimization and simulated annealing were performed to refine the atomic positions, water pick-up and isotropic B-factor refinements were carried out, leading to a model with R-factor and free R-factor of 18.32 % and 21.02 % for OL-obelin (C2 space group), 18.07 % and 19.00 % for soaked OL-obelin and 18.95% and 19.31% for h-obelin, respectively. The composite omit maps were

calculated to assist in building accurate structures using the program O (44). Residues 2 to 4, 124 to 127 were disordered for these three obelin structures. Anomalous Fourier maps were calculated based on the diffraction data collected at the wavelengths of 1.74 Å and 1.59 Å respectively. Two cobalt ions and one potassium ion were found in both structures of OL-obelin (C2) and soaked OL-obelin. One extra calcium ion was found in the soaked OL-obelin structure. The identification of metal ions is described in detail in section 3.5.

All subsequent refinement was continued with the program SHELXL-97 (51) based on diffraction intensities rather than structure-factor amplitude. Before moving to SHELX, all the water molecules were deleted, B-factors were reset to 15.0 Å² for all the atoms. The atomic positions and restrained individual isotropic temperature factors were refined in the initial steps. The refinement led to the highest resolution for each obelin. The $2F_o - F_c$ and $F_o - F_c$ electron density maps were examined after each refinement step. The models were adjusted manually where necessary using XTALVIEW (47). Alternative conformations were modeled for residues 69, 75, 77, 83, 159 and 167 of both obelin structures, residues 67, 98 and 140 of OL-obelin only and residues 42, 44, 53, 63, 81 and 130 of soaked OL-obelin only. Residues 53, 69, 75, 77, 159 and 176 were modeled as alternative conformations for h-obelin. The relative occupancies of the multiple side chain conformers were refined keeping the sum constrained to unity. Occupancy factors were also refined for the metal ions and O33, O34 of the ligand coelenterazine. Anisotropic displacement parameters (ADPs) were refined for all non-hydrogen atoms. ADPs of solvent molecules were restrained to be approximately isotropic. In the final steps, hydrogen atoms were added to the models according to geometrical criteria, yielding free R-factors of 17.25%, 16.60% and 18.79% for the OL-obelin, soaked OL-obelin and h-obelin. When convergence of the conjugate-gradient least-squares minimization was achieved, all data

including the cross-validation data set were used in the refinement, leading to final R-factors of 14.56%, 14.25% and 15.73% respectively. After all the refinement was complete, one additional cycle was run using full-matrix least-squares minimization to estimate the standard deviations for the atomic positions and temperature factors. The quality of all three models was checked using PROCHECK (50) and MolProbity (52). In addition to the water molecules observed as solvent, one glycerol molecule introduced during the cryoprotectant soaking procedure was also found in three structures.

2.6.2 Refinement of Ca^{2+} -loaded apo-aequorin

The initial model of Ca^{2+} -loaded apo-aequorin was refined against the chromium X-ray data using CNS 1.0 (43). Each cycle of refinement such as minimization, simulated annealing, temperature factor refinement, was followed by a manual rebuilding. The resulting model was subsequently refined against the high resolution data collected at the synchrotron with phases gradually extended to 1.7 Å using maximum likelihood refinement with the program RefMac 5.0 (48). Several cycles of refinement and model adjustment resulted in a final R factor of 21.6%, a free R-factor of 23.6%. SIGMAA weighted phases were used to calculate $2F_o - F_c$ and $F_o - F_c$ maps simultaneously with the program XTALVIEW (47) to aid the model building. The first three residues at the N-terminus were not observed in the electron density maps and were assumed to be disordered. The final stereochemical parameters of the structure were evaluated with the programs PROCHECK (50) and MolProbity (52).

2.6.3 Refinement of Ca^{2+} -loaded apo-obelin and Ca^{2+} -discharged W92F-obelin

The refinement of Ca^{2+} -loaded apo-obelin was carried out using RefMac 5.0 (48). The model was gradually built up through the phase improvement by manual model adjustment and missing atom rebuilding after each cycle of refinement using the graphic program XTALVIEW (47).

Trying to build the model into the density was very difficult at the beginning since the electron density was not very good. The first ten residues at the N-terminus and the last twelve residues at the C-terminus were not observed in the electron density maps and assumed to be disordered. The final stereochemical parameters of the structure were evaluated with the programs MolProbity (52).

The further refinement of Ca^{2+} -discharged W92F-obelin was extended to 1.96 Å using RefMac 5.0 (48). Each cycle of refinement was followed by manual model adjustment using XTALVIEW (47). The final stereochemical parameters of the structure were evaluated with the programs PROCHECK (50) and MolProbity (52).

CHAPTER 3

Structure-function Relationship

Calcium-binding proteins are probably one of the most extensively studied protein families. The main reason for such attention is because these proteins regulate numerous vital intracellular events in living cells and organisms. Ca^{2+} -regulated photoproteins are members of the EF-hand calcium binding protein family. It has been generally considered that, as with other calcium-binding proteins, a structural change induced by Ca^{2+} is responsible for initiating the full bioluminescence activity. Indeed, from HSQC-NMR experiments there appear to be five distinct conformation states controlled by the binding of the various ligands, Ca^{2+} , coelenterazine and coelenteramide (Fig. 3.1) (35): the first one is the apoprotein, the second is when calcium is included with the apoprotein (Ca^{2+} -loaded apoprotein), the third on charging the apoprotein (in the absence of calcium) with coelenterazine to produce the holoprotein (Ca^{2+} -regulated photoprotein), then on the addition of calcium to cause the bioluminescence reaction (Ca^{2+} -loaded Ca^{2+} -discharged photoprotein), and finally on removal of calcium from the product (Ca^{2+} -discharged photoprotein). In order to completely elucidate the mechanism of this bioluminescent reaction and the structural transients accompanying each step of the mechanism, the three-dimensional structures of each conformational state need to be solved. Although the structure of conformation state II has been solved for aequorin and obelin (33, 34), questions and puzzles remain to be addressed in order to fully appreciate at atomic detail events following a reaction pathway, such as what the coelenterazine-oxygen binding state(s) really is (are) in different photoproteins, how the binding of Ca^{2+} might trigger the bioluminescence reaction, how different

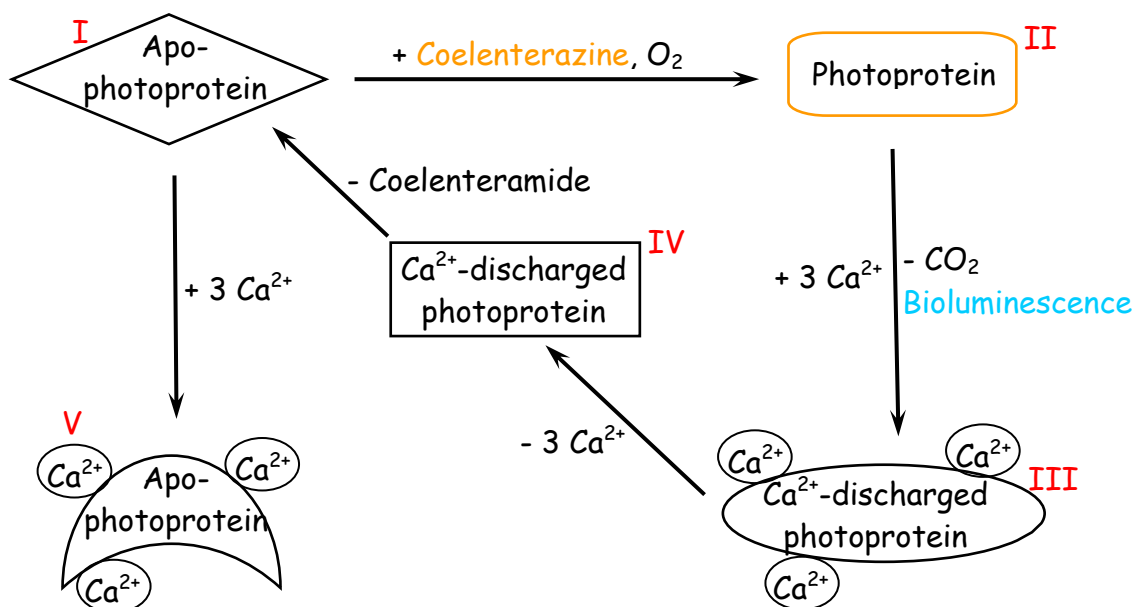


Fig. 3.1 Photoprotein conformation change pathway revealed by HSQC-NMR spectroscopy study of obelin (35): **I**. the apo-protein, **II**. on charging the apo-protein with coelenterazine and oxygen in the absence of Ca²⁺, **III**. then on the addition of Ca²⁺ to cause the bioluminescence reaction, **IV**. on removal of calcium ions from the product **V**. Ca²⁺ bound to the apo-protein. The structure reported here represents state IV.

excited states could result, etc.. Clearly for revealing intimate details of the processes, the structures of several of these conformation states would be highly desirable to have in hand.

3.1 Crystal structure of OG-obelin

The recombinant Ca^{2+} -regulated photoprotein obelin originating from marine bioluminescent hydroid *O. geniculata*, is a single subunit protein of relatively low molecular weight (22.2 kDa) consisting of 195 amino acid residues (14). The crystal structure of the OG-obelin molecule was determined and refined to 1.82 Å resolution. The overall structure, as shown in Fig. 3.2, is highly compact and globular with high helix content and a radius of ~25 Å. Although it is in the calcium-free state, the whole globular molecule is formed in a manner rather analogous to calmodulin in the absence of calcium. The final model includes 192 out of 195 amino acid residues (1636 atoms), hydroperoxy-coelenterazine (34 atoms) and 111 solvent molecules. The first three residues at the N-terminus are disordered. The protein scaffold is well characterized by two domains, N-terminal and C-terminal domain, held together through the linkage of a loop (residue 106 – 109). Each domain contains one pair of helix-loop-helix (HLH) structural motifs I, II, III and IV, three of which (motifs I, III and IV) are classical EF-hand calcium-binding domains. The ligand peroxy-coelenterazine resides in a hydrophobic core cavity in the center of the protein molecule.

The four HLH motifs consist of two sets of four helices designated: A (16 - 29), B (39 - 54), C (58 - 74), 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. 3.2 & Fig. 3.3). The loop regions of the four HLH motifs are all exposed to solvent. The A- and B- helices are joined by a 9-residue loop, which together with 3 residues of the B-helix forms the first expected calcium-binding site. The C- and D-helices are linked by a 10-residue loop (75CGLEYGKETK84) to form the second

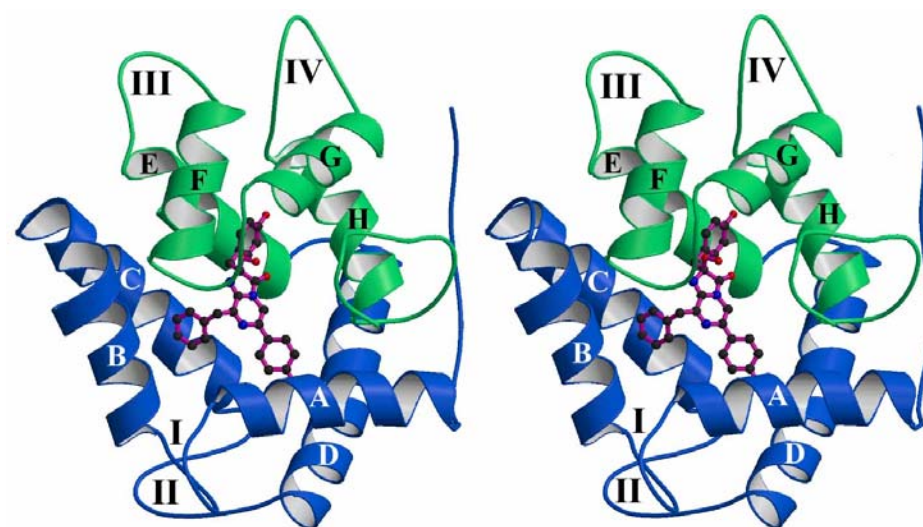


Fig. 3.2 A stereoview of the crystal structure of OG-obelin. The N- and C-terminal domains of the molecule are colored in blue and green respectively. The coelenterazine-oxygen molecule is colored in pink. The helices are marked by capital letters A through H. Roman numbers I through IV designate the loops of HLH motifs.

HLH motif. There seems to be few special features distinguishing it from the other domains. The E- and F-helices are connected through a 9-residue loop, which together with 3 residues of the F-helix forms HLH motif III (the second expected calcium-binding site). The G and H helices are linked via a 10-residue loop, plus 2 residues of the adjacent H helix forms HLH motif IV (the third calcium-binding site). The amino acid composition and structural organization of HLH motifs I, III and IV are typical for canonic Ca^{2+} -binding sites, similar to the calcium-binding sites of other calcium-binding proteins (53) and consequently HLH motif I from N-terminal domain and both HLH motifs of the C-terminal domain can be expected to bind calcium ions.

The amino acid residues forming the loop of HLH motif II are known to be unable to coordinate calcium because of the neutral charge character in this loop region. This was also observed for cardiac troponin C (54) and troponin C from crayfish (55). Although this HLH motif II is not functional for calcium binding, it seems to function for positioning the coelenterazine by pointing the side-chains of its amino acid residues into the binding pocket and for forming a stable scaffold against which the rest of the molecule moves when calcium is bound. Despite these, there is no obvious evidence of this region having a role in the enzymatic function such as involved in the coelenterazine oxidation process.

The N-terminal domain is formed by the interactions between HLH motifs I and II while the C-terminal domain is formed by the interactions between HLH motifs III and IV. Inter-motif interactions occur via the loops and the helices. Two loops of the adjacent HTH motifs interact via hydrogen bonding in the form of anti-parallel β -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

Leu166(O) (2.88 Å), Ile130(O) and Leu166(N) (2.99 Å). Similar inter-motif interactions have also been observed in other calcium-binding proteins (56).

Site-directed mutagenesis and direct chemical modifications of photoproteins have shown certain amino acid residues to be important for photoprotein bioluminescence. These are tryptophan, histidine, cysteine, and the C-terminal proline. OG-obelin contains 6 tryptophan, 3 cysteine, and 5 histidine residues, for example (Fig. 3.3). Most of these tryptophan and histidine residues are located in the coelenterazine binding pocket which will be discussed in the later sections. It was shown that one of cysteine residues is important for the bioluminescence function of photoproteins and that cysteine might play an important role in the regeneration of holoprotein but not in its catalytic activity (28, 57). However, in the current structure no cysteine residues are observed near the coelenterazine-oxygen molecule. Also there are no disulfide bonds. Probably the cysteine 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 (57).

One other important residue for the bioluminescence of photoproteins is a proline at or near the C-terminus (27). It has been shown that any immediate preceding insertions, or deletions, or replacement, or following extensions of this C-terminal Pro, will destroy the luminescent capacity of photoproteins, suggesting the absolute requirement for this proline. 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 Ne atom of Arg21, localized in the A-helix. The C-terminus is formed by an extended loop consisting residues 183 - 195 lying in a space between the first helix of HLH motif I and the first helix of HLH motif IV. Numerous


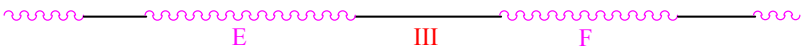
Aequorin	MTSKQYSVKLTSDFDNPRWIGR HKH MFNF LDVNHNGKISLDE MVYKASDIV (51)	
Obelin OL	MSSKYAVKLKTDFDNPRWIKR HKH MFDF LDINGNGKITLDE IVSKASDDI (50)	
Obelin OG	MASKYAVKLQTDFDNPKWIKR HKF MF LDINGNGQITLDE IVSKASDDI (50)	
		
Aequorin	INNLGATPEQAKRHKDAVEAF FGGAGMKYGVETD WPAYIEGWKKLATDEL (111)	
Obelin OL	CAKLEATPEQTKRHQVCVEAF FRGCGMEYGKEIA FPQFLDGWKQLATSEL (110)	
Obelin OG	CKNLGATPAQTQRHQDCVEAF FRGCGLEYGKETK FPEFLEGWKNLANADL (110)	
		
Aequorin	EKYAKNEPTL IRIW GDAL FDIVDKDQNGAITLDE WKAYTKAAGIIQSSSED (171)	
Obelin OL	KKWARNEPTL IREW GDA VDIFDKDGS GTITLDEWKAYGKISGISPSQED (170)	
Obelin OG	AKWARNEPTL IREW GDA VDIFDKDGS GTITLDEWKAYGRISGISPSEED (170)	
		
Aequorin	CEETFRCV CDIDESGQLD VDEMTRQHLGF WYTMD PACEKLYGGAVP (196)	
Obelin OL	CEATFRHC DLDNSGDL DVDEMTRQHLGF WYTLD PEADGLYGNGVP (195)	
Obelin OG	CEKTFQHC DLDNSGEL DVDEMTRQHLGF WYTLD PEADGLYGNGVP (195)	
		

Fig. 3.3 Sequence alignment of aequorin (11) (12), OL-obelin (13) and OG-obelin (14) showing the relative locations of residues forming the hydroperoxy-coelenterazine binding pocket with respect to the secondary structure elements. The loop regions of four HLH motifs are designated by I, II, III and IV. *Blue*, Strictly conserved residues in the binding pocket; *green*, other residues in the binding pocket; *red*, predicted Ca^{2+} -binding loops; *brown*, non- Ca^{2+} binding loop of HLH motif; *pink curve*, α -helices A – H; *black line*, loops.

other hydrogen bonds are formed in this region anchoring the C-terminal chain into its conformation. As a result the N-terminal and C-terminal domains form a closed conformation, inside of which the coelenterazine-oxygen substrate is contained.

3.2 Coelenterazine binding pocket of OG-obelin

The coelenterazine binding pocket is buried in the center of the globular OG-obelin molecule; hence, there is no solvent access to the site from the surface in this conformation, which structurally supports the bioluminescence function of this molecule. The binding pocket is formed mostly by hydrophobic residues and a few hydrophilic residues originating from the eight helices. These residues are helix A (His22, 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 and Tyr138), Ile144 from the loop linking helices F and G, and helix H (Met171, His175 and Trp179). In addition, Tyr190 located near the C-terminus of the protein is directed into the pocket as well.

Based on early experiments where attempts to remove molecular oxygen from the reacting solutions failed to change the kinetics of the bioluminescence, in marked contrast to the clear oxygen dependence of other bioluminescence reactions under study at the time, the firefly luciferase and the bacterial luciferase reactions, it was proposed that a peroxy group is substituted at the C2 position of coelenterazine, producing a peroxy-coelenterazine complex. Instead of using coelenterazine, photoproteins bind their peroxy derivative for bioluminescence reaction (58). In the center of OG-obelin crystal structure, the electron density is well resolved and is consistent with coelenterazine coupled to a hydroperoxide at the C2 position as shown in Fig. 3.4. The hydroperoxide appears to be stabilized by hydrogen bonding to the phenolic oxygen of Tyr190 (2.39 Å) which is itself hydrogen bonded to the Nε2 of His175 (2.55 Å). The

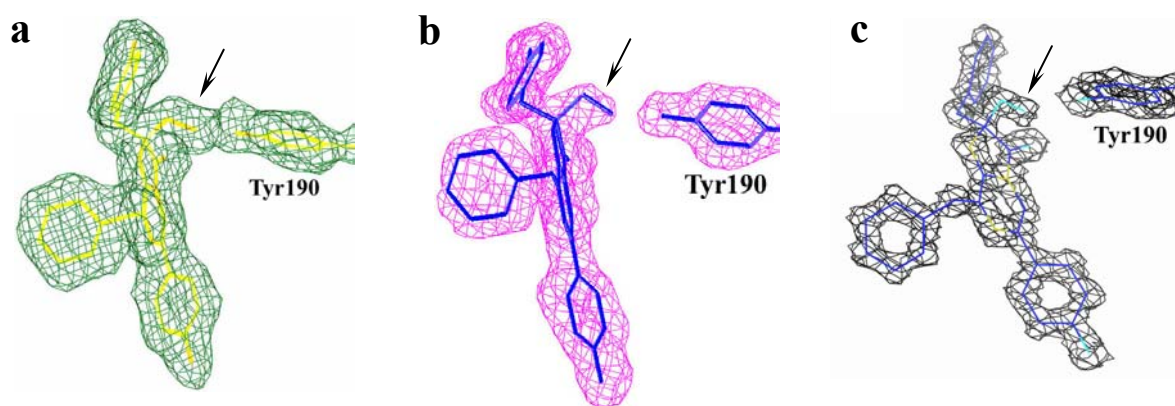


Fig. 3.4 Electron density maps contoured at 1.0 σ of hydroperoxy-coelenterazine hydrogen bonded with residue Tyr190 of OG-obelin (a), W92F-obelin (b) and h-obelin (c). Note that the electron density next to the C2- position (arrow) can easily be fitted by two oxygen atoms.

imidazole of His175, in turn, is situated close to the carbonyl oxygen (3.36 Å) of C3 on the ligand and to the indole of Trp179. Besides that these interactions may position the coelenterazine in the binding site; it is possible that these residues participate in some steps of the bioluminescence reaction. For instance, they could assist the formation of the transient anionic dioxetane in the reaction pathway. The importance of these interactions is supported by mutagenesis studies of obelin and the corresponding residue in aequorin; it has been shown that substitution of His169 in aequorin or His175 in OL-obelin to Ala, Phe, or Trp leads to almost complete loss of activity (29, 57); W173F (Trp179 in obelin) mutant has very little luminescence activity (30).

The Ca^{2+} -regulated photoproteins are distinctive from other Ca^{2+} -binding proteins in having an amino acid sequence with many tryptophan, cysteine, and histidine residues, which are not commonly found in Ca^{2+} -binding proteins (15, 26). Four tryptophans (Trp92, Trp114, Trp135, Trp179) and two histidines (His22, His175) are located in the hydroperoxy-coelenterazine binding pocket. Fig. 3.5 shows the coelenterazine-oxygen substrate and its surrounding residues which form the hydrogen-bound network of the active center. The histidines and tryptophans are organized in three pairs, two of them closely associated with a tyrosine. The three pairs of histidines (His22, His64, His175) and tryptophans (Trp92, Trp114, Trp179) together with two tyrosines (Tyr138, Tyr190) are located close to C2-C3, O25, and N1 of coelenterazine respectively which are the key atoms either for the oxidation of coelenterazine or the formation of the product excited states.

Ohmiya et al. (30) made all the Trp to Phe mutants of aequorin and found that one of them, W86F, together with the usual bioluminescence emission with $\lambda_{\text{max}} = 465 \text{ nm}$, showed an additional band with $\lambda_{\text{max}} = 400 \text{ nm}$. In an attempt to understand why the substitution of this one

amino acid favors the formation of the 400 nm emission band, the OL-obelin mutant, W92F was produced. Trp92 in OL-obelin corresponds to the Trp86 in aequorin. The bioluminescence of W92F-obelin is a violet color as a result of the addition of a new band with $\lambda_{\text{max}} = 405$ nm and an intensity similar to the major band of wild type OL-obelin at the longer wavelength (485 nm). The substitution of Trp92 in OL-obelin results in a larger relative intensity of the shorter wavelength emission than in the case of aequorin. It was suggested that Trp86 in aequorin may be involved in the generation of the product excited state during photoprotein luminescence (30). The structures of aequorin and OL-obelin (33, 34), together with OG-obelin support this suggestion, since the side chains of Trp92 and Trp179 in OG-obelin, for example, “sandwich” the 6-p-hydroxyphenyl ring of coelenterazine. The planes of the phenyl ring of coelenterazine and the Trp92 indole are almost parallel. The side chains of Trp114 and Trp135 are localized near the 2-p-hydroxybenzyl group of coelenterazine.

In OG-obelin, Trp92 and His22 are in the vicinity of the 6-p-hydroxyphenyl group of the bound coelenterazine (Fig. 3.5). Trp92 not only forms a hydrogen bond (3.27 Å) between its Nε1 nitrogen atom and the 6-p-hydroxyphenyl group of coelenterazine, but also has interaction of its side chain atoms with side chain atoms of His22. The Nε atom of His22 is very close (2.94 Å) to the oxygen atom of the 6-p-hydroxyphenyl ring of coelenterazine implying a strong hydrogen bond or even favoring complete proton transfer leaving the phenolate anion. These interactions would influence the charge status of the phenol oxygen during the excited state of product coelenteramide.

The phenolic OH of Tyr138 is positioned laterally to the plane of the imidazopyrazinone ring of the ligand and hydrogen bonds to N1 (Fig. 3.9). It is also linked through hydrogen bonds by a water molecule to His64. This histidine imidazole is adjacent to the indole ring of Trp114 which

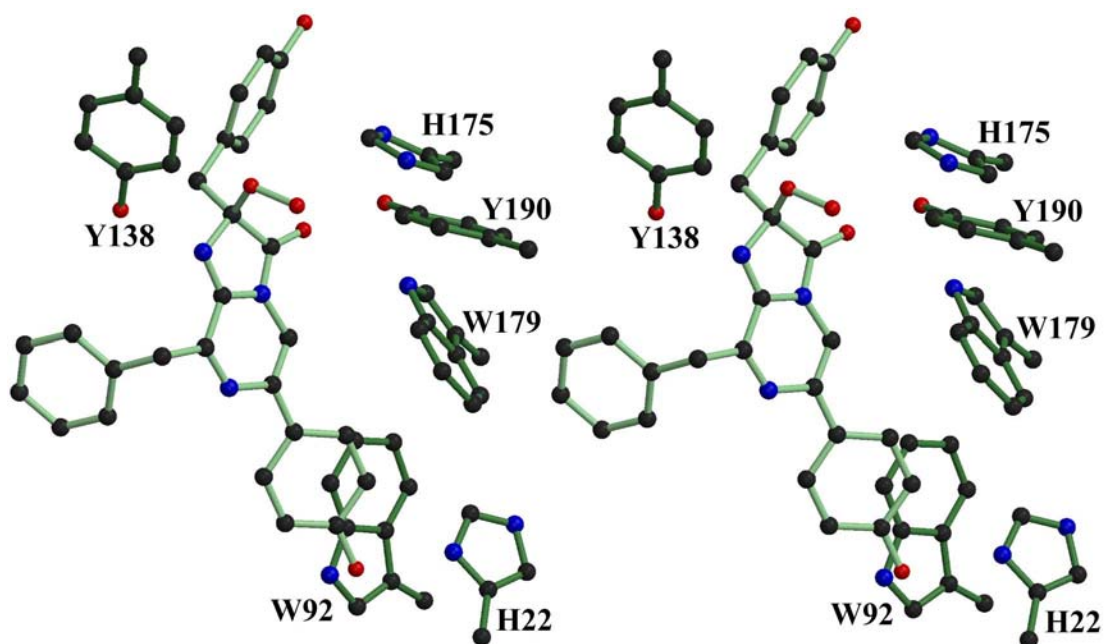


Fig. 3.5 A stereoview of the interactions between hydroperoxy-coelenterazine molecule and the OG-obelin protein residues directly hydrogen-bonded with the substrate.

partly overlays the imidazopyrazinone ring system. The p-OH group on the benzyl substitution at C2 of the ligand is hydrogen bonded to a water molecule which itself interacts with the carbonyl oxygen of Ile111 and the side chain oxygen of Thr172. This OH group does not appear to be essential to the function of photoproteins, as replacing it with hydrogen or halogens does not significantly affect light-emitting capabilities (59). In the study of apo-obelin bound with coelenterazine h (an analogue of coelenterazine with a benzyl group substituted at the C2 position instead of a hydroxybenzyl group as in coelenterazine) (h-obelin), no significant bioluminescence and fluorescence spectral differences were observed (Fig. 3.6).

3.3 Structure comparisons of Ca^{2+} -regulated photoproteins

Although Ca^{2+} -regulated photoproteins have been identified in a variety of marine organisms, mostly coelenterates (6), only six of them have so far been isolated and purified. These are aequorin (4), halistaurin (mitrocomin) (60) and phialidin (clytin) (61) from the jellyfish *Aequorea*, *Halistaura* (*Mitrocoma*), and *Phialidium* (*Clytia*) respectively; obelin from the marine hydroids *O. geniculata* (62, 63) and *O. longissima* (64); and mnemiopsin and berovin (65, 66) from the ctenophores *Mnemiopsis* and *Beroe*. During the past fifteen years, cloning and sequence analysis has been achieved for the cDNAs coding for four Ca^{2+} -regulated photoproteins: aequorin (11, 67, 68), phialidin (clytin) (16), halistaurin (mitrocomin) (17), and obelins from *O. longissima* and *O. geniculata* (13, 14). All photoproteins show high sequence homology. For example, OG-obelin shows a high identity to other calcium-regulated photoproteins: OG-obelin to OL-obelin = 86%; OG-obelin to aequorin = 64%; OG-obelin to clytin = 76%; and OG-obelin to mitrocomin = 64%.

In spite of the high homology, photoproteins are different in some biochemical characteristics. For example, OL-obelin and OG-obelin are considerably faster in

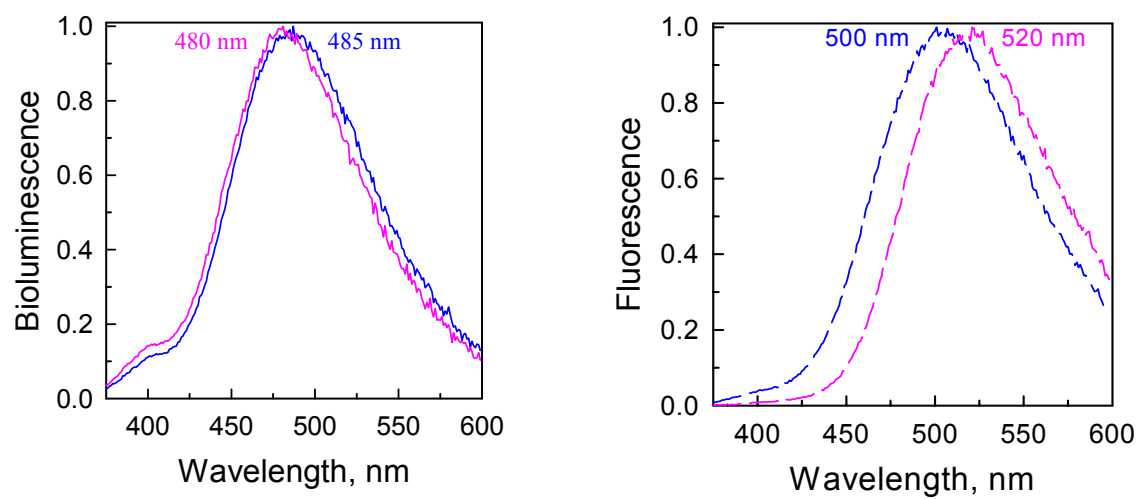


Fig. 3.6 Left: bioluminescence spectrum of OL-obelin (pink) and h-obelin (blue). Right: fluorescence spectrum of Ca²⁺-discharged OL-obelin (pink) and Ca²⁺-discharged h-obelin (blue).

bioluminescence response than aequorin (14, 36, 63), and obelins (14, 36) and halistaurin (69) are relatively insensitive to physiological concentrations of Mg^{2+} . Phialidin is reported to be less sensitive to calcium than aequorin (70). There are even differences in bioluminescence properties (bioluminescence spectrum, kinetics, etc.) for the two obelins from *Obelia* species which have 86% sequence homology (14). Obviously, the variations of photoprotein properties would be expected to originate from the differences in their primary and tertiary structures.

There are five photoprotein structures available: OG-obelin (1.82 Å), OL-obelin (1.73 Å) (33), aequorin (2.3 Å) (34), h-obelin (1.14 Å), W92F-obelin (1.72 Å). The three wild type photoprotein structures are very similar (Fig. 3.7) consistent with their amino acid sequence homology; the same fold is observed in h-obelin with a different substrate and in W92F-obelin with a mutation site as well. The root-mean-square deviations (RMSDs) of the C α -atomic positions of OG-obelin against OL-obelin, aequorin, h-obelin and W92F-obelin are 0.46, 1.23, 0.65 and 0.52, respectively (first ten solvent-exposed residues from the N-termini are excluded from the calculation), showing the well conserved structural features.

However, obvious deviations are observed at the loop regions of HLH motifs III and IV (Fig. 3.7) which are calcium binding sites of the EF-hand motifs. The structure of EF-hand motif I is essentially the same among all available photoproteins. In photoproteins the binding of calcium to the loops of EF hands initiates the bioluminescence reaction. Certain residues, such as Tyr138, His175 and Trp179 in obelins, which have critical proximity to the substrate in the reactive center, are located within the helices of EF-hand motif III and IV (Fig. 3.8). Any deviations at loop regions of EF-hands III and IV would cause a different response to the calcium binding which requires strict ligation geometry. With the caution that some of the differences in the loop regions may arise from crystal packing contacts, there remains sufficient evidence to suggest that

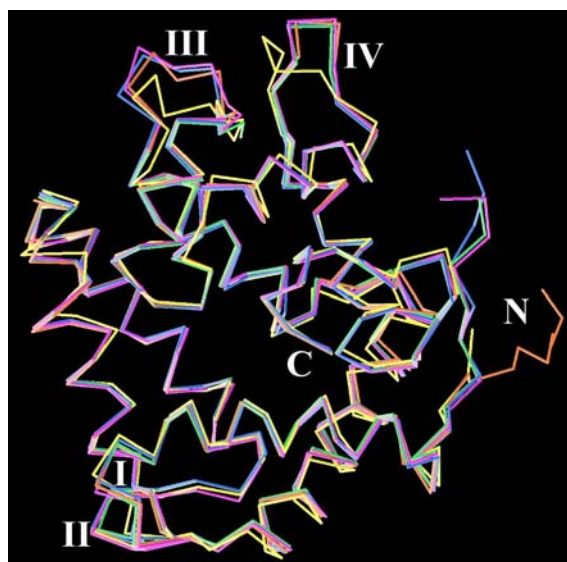


Fig. 3.7 Superimpositions of Ca^{2+} -regulated photoproteins, OG-obelin (blue), W92F-obelin (magenta), h-obelin (green), OL-obelin (brown), and aequorin (yellow), show well conserved overall scaffolding in this protein family when apo-proteins bound with hydroperoxy-coelenterazine. However, obvious deviations are observed in solvent exposed regions, N terminus and loops of HLH motifs III and IV. The RMSD values of the $\text{C}\alpha$ -atomic positions of OG-obelin against OL-obelin, aequorin, h-obelin and W92F-obelin are 0.46, 1.23, 0.65 and 0.52, respectively (first ten solvent-exposed residues from the N-terminuses are excluded from the calculation).

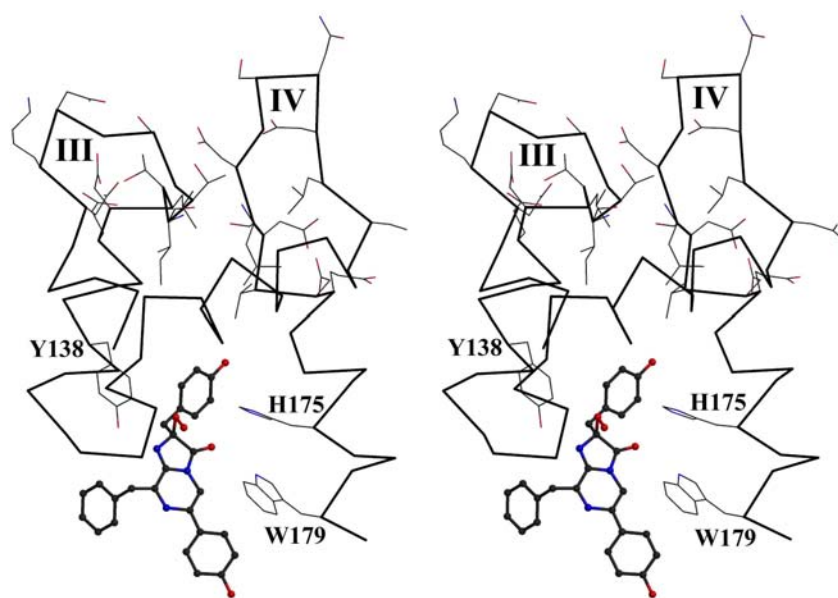
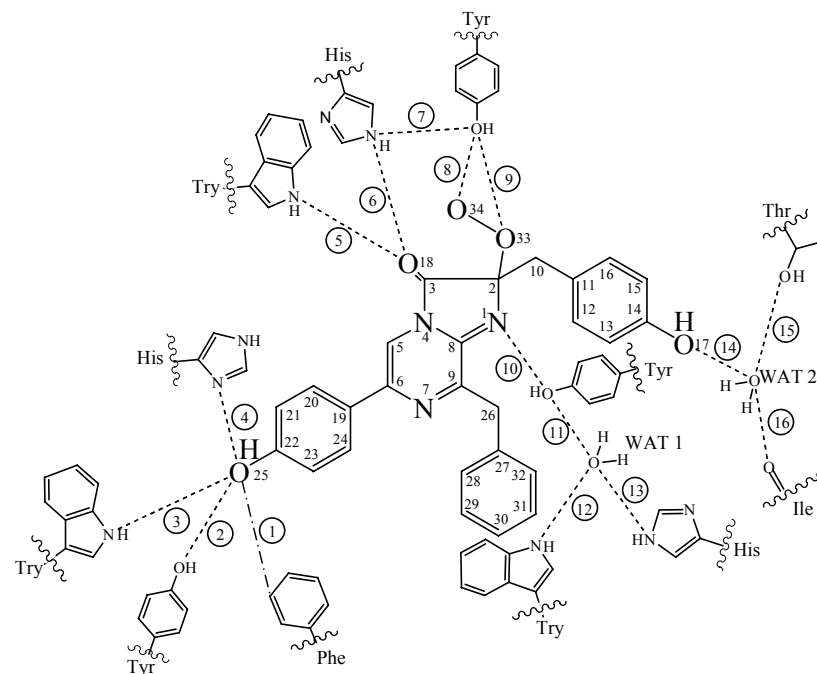


Fig. 3.8 A stereoview of the hydroproxy-coelenterazine binding site of OG-obelin and nearby residues from EF-hand motif III and IV.

the differences in the kinetics of the bioluminescence response of photoproteins could result from the mode of calcium binding to the protein, which is partially determined by the orientations of the loop regions of the EF-hand motifs.

The binding cavity of the coelenterazine is the structure of most interest in the Ca^{2+} -regulated photoproteins. Almost all the residues forming the coelenterazine binding pocket are conserved among all Ca^{2+} -regulated photoproteins (Fig. 3.3) and consequently the structure of the coelenterazine-oxygen binding sites should be well conserved for all Ca^{2+} -regulated photoproteins. The five available photoproteins have similar dimensions in the coelenterazine binding sites with a radius approximately 25 Å. As shown in Fig. 3.9, similar hydrogen bonding networks of the coelenterazine with the key residues at the active sites are also observed. An obvious difference between obelin and aequorin is that Tyr82 in aequorin which is hydrogen bonded to the oxygen atom of the 6-p-hydroxy-phenyl group of coelenterazine is replaced by Phe88 in obelin. As a result, aequorin has three Tyr-His-Trp triads whereas obelins only have two. The one less hydrogen bonding interaction in obelins at the active site might have a critical influence on the bioluminescence and fluorescence emission spectrum differences between these two species and will be discussed in detail in section 3.7.

There are two “buried” water molecules consistently observed at the active sites of coelenterazine among the photoproteins (Fig. 3.9). In many proteins, internal waters are an essential part of the three-dimensional structure and their positions are typical for a whole family of proteins. In the case of OG-obelin, one water molecule forms hydrogen bonds with the hydroxyl oxygen (2.64 Å) of the p-OH-benzyl group at the C2-position, and with the side-chain oxygen (2.75 Å) of Thr172 and the main-chain oxygen (2.81 Å) of Ile111. The other is bound with the hydroxyl oxygen atom (2.63 Å) of Tyr138 and the NE2 atom (2.81 Å) of His64.



	1*	2*	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Aequorin [†]	-----	2.59	3.14	2.70	3.75	3.33	2.57	2.52	3.84	2.57	2.87	3.30	2.97	2.92	2.82	3.06
OL-obelin	3.16	-----	3.17	2.87	3.32	3.24	2.60	-----	3.17	2.75	2.63	3.64	2.92	2.68	2.72	2.86
OG-obelin	3.06	-----	3.27	2.94	3.32	3.36	2.55	2.39	3.59	2.70	2.63	3.26	2.81	2.64	2.75	2.81
W92F-obelin	3.17	-----	----- [‡]	2.84	3.26	3.37	2.64	2.74	3.54	2.68	2.68	3.54	2.83	2.85	2.64	2.72
h-obelin	3.27	-----	3.26	2.73	3.28	3.31	2.60	2.55	3.22	2.73	2.77	3.60	2.92	-----	2.84	2.99

* In aequorin, there is a Tyr82, corresponding to Phe88 in obelins.

[†] Aequorin was crystallized as a dimer (34); distances shown here are the averages of both structures.

[‡] The mutation point; F92 is around 3.52 Å away from O25 of coelenterazine.

Fig. 3.9 The top sketch is a two dimensional diagram of the hydrogen-bonding (dotted line) interactions at the binding site of the coelenterazine-oxygen complex. WAT indicates a water molecule. The table is the hydrogen-bonding distances in aequorin, OL-obelin, OG-obelin, W92F-obelin and h-obelin. Distances are in Å.

Besides, this water molecule interacts (3.26 Å) with the NE1 atom of Trp114 forming a weak hydrogen bond. Internal water molecules practically in the same positions were found in OL-obelin, aequorin, and W92F-obelin. Even in the structure of h-obelin with the benzyl group at the C2 position, a water molecule is also observed at the corresponding position of WAT1 in OG-obelin after the hydrogen bounding hydroxyl has been removed. Consequently it is concluded that these two water molecules are common in the Ca^{2+} -regulated photoprotein family.

In proteins internal water not only fills structural cavities but also can be necessary to stabilize the three-dimensional fold of the molecule, to perform a function of “lubricant” for conformational change, or to participate in some catalytic steps of the enzyme reaction. The role of “buried” water molecules for calcium-regulated photoproteins is unclear but some speculation having a solid basis arising from studies of the mechanism of the bioluminescence reaction catalyzed by *Renilla* luciferase (71), could be offered. It was shown using isotope-labeled coelenterazine ([3- ^{14}C] and [3- ^{18}O]), H^{16}OH , and $^{17}\text{O}_2$ that the coelenterazine carbonyl oxygen is exchanged rapidly on the luciferase with oxygen from water prior to luminescence. This exchange occurs at some intermediate stage during reaction, before the final concerted bond cleavage to produce CO_2 and light via the “dioxetanone” pathway. The mechanism of the bioluminescence reaction of Ca^{2+} -regulated photoproteins is likely the same as *Renilla* bioluminescence. Then the internal water molecule (one or both) observed in the photoprotein binding cavity, may be involved in the bioluminescence reaction performing the same exchange as in the case of *Renilla* luciferase. The water molecule might just be needed for protein structural folding.

The first two crystal structures of Ca^{2+} -regulated photoproteins are aequorin (34) and OL-obelin at 1.73 Å resolution (space group of P6_2) (33). The peroxide substitution on the

coelenterazine was suggested in the aequorin structure, where weak electron density consistent with the presence of two oxygens, was found at the C2 position of the coelenterazine (Fig. 1.7), in accordance with the prediction from the many earlier biochemical and chemical model studies. However, in the OL-obelin structure at 1.73 Å resolution, the electron density distribution (Fig. 1.7) near the C2 position of the coelenterazine molecule clearly corresponds to only a single oxygen, not peroxy, with no covalent binding to the protein. In addition, a hydrogen bond is evident between the C2-O33 and the hydroxyl oxygen of Tyr190 (3.16 Å). In the later solved photoprotein structures of OG-obelin, W92F-obelin and h-obelin, peroxy oxygen atoms attached at the C2 position of the coelenterazine were observed as in the case of aequorin, confirming the hydroperoxy-coelenterazine binding state (Fig. 3.4). Both the single oxygen coelenterazine and two oxygen coelenterazine appear to be stabilized by hydrogen bonding to the phenolic oxygen of a Tyrosine residue. So does this different observation mean that coelenterazine-oxygen has two different active binding states, hydroperoxy-coelenterazine and hydroxy-coelenterazine?

3.4 Coelenterazine-oxygen binding state

In the report on the structure of aequorin, Head et al. (34) noted that the electron density corresponding to the peroxy substituent was weaker than that of the other atoms of the coelenterazine. They suggested that radiation damage could be responsible. It has been shown that the sulfur electron density of the S-S links in the electron density map of the proteins *Torpedo californica* acetylcholinesterase and hen egg white lysozyme would decrease in level and disappear completely on increasing time in the synchrotron beam (72). However we concluded that radiation damage was unlikely in the case of obelin since the synchrotron irradiated crystal appeared to be intact, retained the yellow color, and produced bright bioluminescence on addition of Ca^{2+} (33, Vysotski, 2001 #1157). Furthermore several

succeeding structures, OG-obelin, h-obelin and W92F-obelin, all showed both peroxy oxygen atoms although the crystals received very similar radiation dosage. In fact, the *O. geniculata* showed electron density for both oxygen atoms of the same level as the other atoms of the coelenterazine ring.

Two puzzles arose: one is that OL-obelin has electron density fitted to one-oxygen bound coelenterazine whereas other photoproteins have electron density corresponding to two-oxygen bound coelenterazine; the other is how the OL-obelin manages to change from one oxygen atom to two oxygen atoms, the required peroxy species for the bioluminescence reaction? Could it be caused by the introduction of calcium binding or by the pH change? In order to address this issue, different crystallization conditions were tried to obtain more crystals. It turned out that obelin is a very cooperative protein in its willingness to form excellent crystals. Higher quality OL-obelin crystals with different crystal form (space group of C2) were obtained and can diffract to atomic resolution of 1.0 Å. After exposing the crystal to a very dilute solution of CaCl₂ so as not to generate any bioluminescence, the resolution was improved to 0.97 Å. The calcium soaking experiment is described in section 2.3 in detail. The OL-obelin not exposed to calcium and the OL-obelin after CaCl₂ soaking will be referred to herein as just OL-obelin and soaked OL-obelin.

The photoprotein obelin provides an excellent model system for biological structure study. Obelin can be regarded in essence as an enzyme with a trapped intermediate, the coelenterazine hydroperoxide located in the substrate binding site. To our knowledge there is no other enzyme bound substrate or intermediate solved to resolution less than 1 Å. The structures of OL-obelin and soaked OL-obelin were solved by molecular replacement using P6₂ space group OL-obelin as the search model. The high crystal quality allowed the two structures to be solved to atomic

resolution of 1.0 Å and 0.97 Å respectively, the highest of any EF-hand calcium-binding protein. Table 2.5 shows the refinement parameters and statistics of the final models of the two obelins. Both models consist of 1553 protein atoms (some side chain atoms were modeled in double conformations with partial occupancies). Residues 124 to 127 are disordered in both protein structures. The trace of the main chains of both obelin structures in this work is very close to the P6₂ space group obelin at 1.73 Å resolution (33). The RMSDs between the C α atoms of the paired structures (residue 11 to 123 and residue 128 to 195) are 0.416 Å and 0.410 Å for the OL-obelin and soaked OL-obelin vs. the P6₂ OL-obelin at 1.73 Å, respectively. A large deviation between two main chains occurs due to the N-terminal residues (5-10) which are exposed to solvent region and not included in the calculation of RMSD. The main chain RMSD between OL-obelin and the soaked OL-obelin is only 0.213 Å.

Diffraction data at 1.0 Å and 0.97 Å resolution allowed full matrix crystallographic refinement in SHELX-97 (51) of atomic positions and anisotropic displacement parameters (ADPs) with minimal reliance on the stereochemical restraints, as well as determination of estimated standard deviations of all parameters. Refinement of obelins at 1.0 Å and 0.97 Å resolution resulted in an electron-density map of excellent quality. For the majority of the molecule, individual atoms have spherical electron densities in the 2F_o – F_c map contoured at 2 σ . The reliability of the refined models was qualified by comparison of their geometry to the restrained target values and also by estimating standard deviations of the refined parameters. Analysis of the stereochemistry of both models of obelin shows that all geometric parameters are within the limits expected for this resolution. The unexpected finding is that a very weak electron density can be detected extending from the first oxygen corresponding to the peroxide group attached to the C2 position of coelenterazine (Fig. 3.10). This is apparently the "missing" second

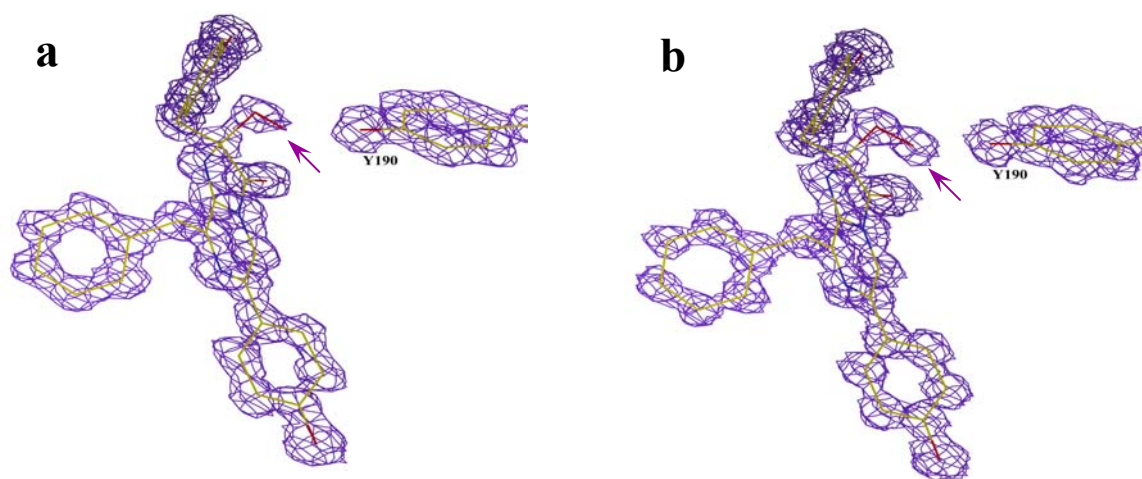


Fig. 3.10 The electron density map contoured at 1.5σ of peroxy-coelenterazine within the OL-obelin (C2) binding site. (a): before soaking with Ca^{2+} ; the electron density (arrow) of the second oxygen atom of the peroxy group at the C2-position is not as strong as for the other atoms in the peroxy-coelenterazine structure and was not observed at all at lower resolution (OL-obelin with P62 space group at 1.73 \AA (33)); (b): after soaking with Ca^{2+} . The augmentation of electron density at the C2-position is clearly observed.

oxygen and leads to the conclusion that the peroxy group does exist in obelin but the question is now raised as to why there is this variability in the electron density now observed in all photoprotein structures, aequorin and the various obelins. Furthermore, in the soaked OL-obelin structure, the electron density of the second peroxy oxygen is strongly enhanced (Fig. 3.10).

Concomitant with this enhancement is a strong increase in rigidity of the constituents of the binding cavity. Fig. 3.11 is a comparison of the isotropic B-factors between obelin and soaked OL-obelin, for the atoms of the coelenterazine molecule and the key atoms of the residues hydrogen bonded to coelenterazine. The average temperature values of both proteins are similar, 12.1 \AA^2 for OL-obelin, 10.3 \AA^2 for the soaked OL-obelin, so it is allowable to compare their B-factors (73). The B-factors of the coelenterazine atoms (except C31 and C32) and the key atoms of the binding site residues in soaked OL-obelin, are consistently lower than those of OL-obelin indicating that the coelenterazine binding pocket becomes more rigid after soaking with calcium. A resulting loss of mobility of a ligand such as the peroxy oxygens, would then account for the electron density enhancement. Notably the temperature factor of the second oxygen O34 in the obelin case is significantly higher than all the other atoms. Its mobility could account for the weak electron density in obelin here and its undetectability in the 1.73 \AA structure. Usually only residues at the terminals or in solvent exposed regions can be difficult to detect in electron density maps but mobility of even buried residues is not without precedent (74). Similar mobility arguments can be proposed to account for the variability in the electron density of these peroxy oxygen atoms in the other photoproteins. There does not seem to be any physiological consequence of this proposed mobility of the peroxy oxygen atoms.

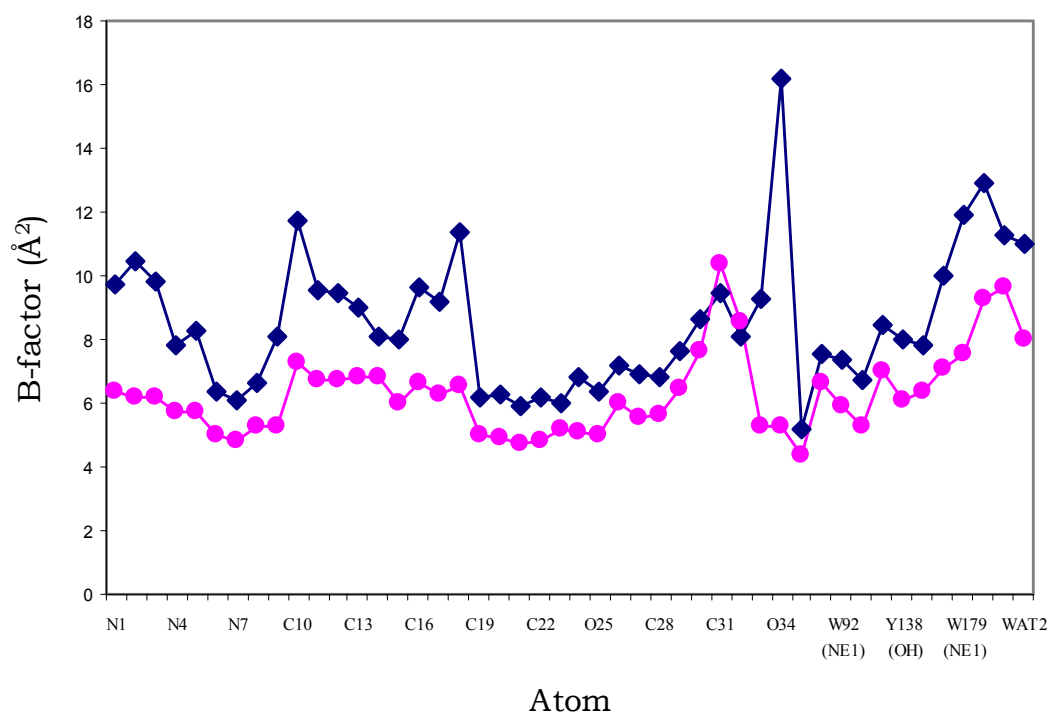


Fig. 3.11 B-factors of peroxy-coelenterazine atoms and key atoms (hydrogen bonded with peroxy-coelenterazine) of the peroxy-coelenterazine binding sites before (navy) and after soaking with calcium (pink). The decrease of the B-factors of key atoms after soaking with calcium indicates that peroxy-coelenterazine binding pocket becomes more rigid resulting in the reduction of flexibility of the second oxygen atom at C2-position.

3.5 Identification of the metal ions in the atomic resolution OL-obelin structures

Anomalous difference Fourier maps were calculated to check if there are any heavy atoms in the crystal (Fig. 3.12). Three heavy atoms were observed in the OL-obelin model (Fig. 3.13). They are all positioned around the IV loop region of the protein molecule. In the soaked OL-obelin there are four heavy atoms, three at the same positions as in obelin and the fourth located in the Ca^{2+} -binding loop I. Tracking back to the protein preparation and crystallization procedure (section 2.1.1 and 2.2.1), the only likely metal ions are calcium, cobalt and potassium.

Elements absorb X-rays as well as emit them, and this absorption drops sharply at the wavelength near their characteristic absorption edge where they exhibit anomalous scattering. In the wavelength range used in crystallography, the anomalous absorption edge of cobalt ions is obvious and is used for cobalt ion identification. Two anomalous data sets were collected at the synchrotron X-ray source for the soaked OL-obelin crystal at wavelengths of 1.74 Å and 1.59 Å which is across the cobalt absorption edge wavelength. Fig. 3.12 shows the anomalous difference Fourier maps at 3.0 Å around the IV loop region at both X-ray wavelengths. The obvious increase of electron density around the green balls from Fig. 3.12a (1.74 Å) to Fig. 3.12b (1.59 Å) identifies them as cobalt ions.

In contrast there is a decrease in the electron density for the third metal (blue ball) near the loop IV from Fig. 3.12a to Fig. 3.12b and for the proposed calcium ion (data not shown). The decrease in electron density agrees with the correlation of anomalous signal strength to changing wavelength, from 1.74 Å to 1.59 Å, for calcium and potassium. The coordinate system of this unknown ion (blue ball) is consistent with it being potassium (75). There are six coordinates with distances between 2.25 Å and 3.17 Å (Fig. 3.14). Three coordinates are from the main chain oxygen atoms, the other three are water molecules.

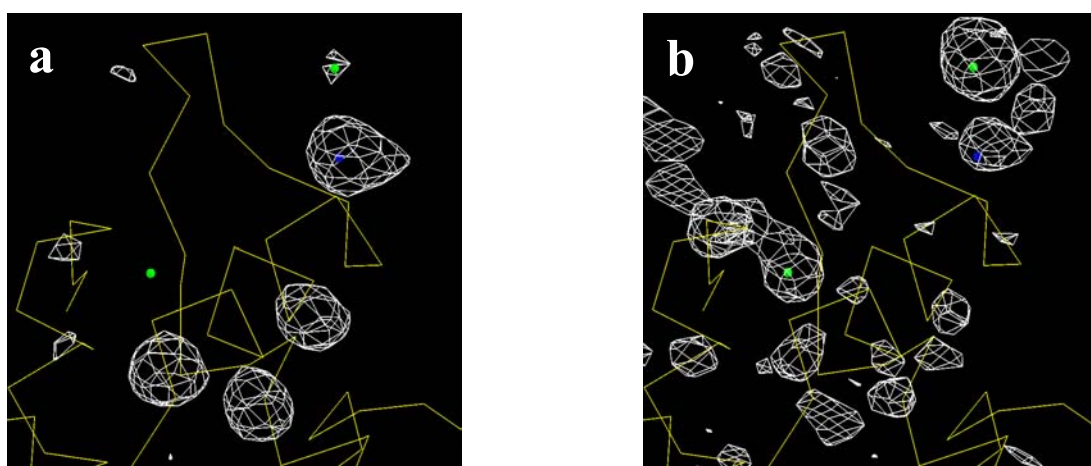


Fig. 3.12 Anomalous difference Fourier map at 3.0 Å using data collected at: (a) 1.74 Å wavelength X-rays; (b) 1.59 Å wavelength X-rays. Yellow curves are protein backbone. The increase of electron density around the green balls from (a) to (b) identifies these as cobalt ions. The electron density around the blue ball decreases. This metal is unidentified but is possibly a potassium ion.

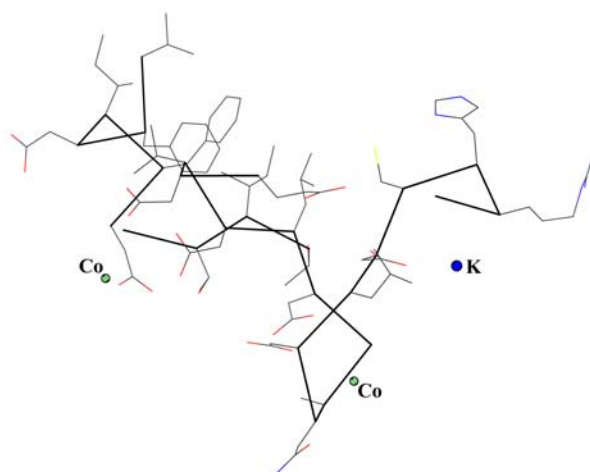


Fig. 3.13 The two cobalt ions (green) and the one assumed potassium ion (blue) are positioned around the fourth loop region of OL-obelin molecule.

The fourth ion at the loop region of the first EF-hand motif of the soaked OL-obelin is tentatively identified as a Ca^{2+} (Fig. 3.14). This position in the loop is known to be a calcium binding site but Ca^{2+} is usually identified by a pentagonal bipyramidal coordinate system with close to a 2.4 Å separation to the central atom (56). Here there are only six coordinates to the metal ion but they are in a bipyramidal configuration and do have a qualifying ~ 2.3 Å separation from the central metal. Four of the ligands are provided by protein residues Asp30, Asn32, Asn34, and the main chain oxygen of Lys36 corresponding to the typical calcium coordinates in positions 1, 3, 5, 7 of the EF-hand loop region. The other two are from water molecules. However this coordination picture is different from that found in most other Ca^{2+} -binding proteins in another respect. The twelfth residue Glu41 does not contribute either of its side-chain oxygen atoms to the calcium ion. In all other known calcium-binding proteins, linkage to the twelfth glutamate is very consistently observed (56). As already mentioned, the coordinate system has only six ligands, not seven ligands as in most other EF-hand proteins. However, six-coordinated EF-hand calcium binding sites are known (76, 77). These unusual characteristics of this calcium-binding site could be a result of the calcium not attaining a final equilibrium binding position within the loop of the EF-hand motif. In other reported calcium-binding protein structures, calcium ions are always co-crystallized in the proteins. The metals and protein have enough time and flexibility to reach the optimum positions for equilibrated binding. In our case, the obelin crystal was soaked into calcium ions so that crystal lattice forces might have impeded some structural adjustment to accommodate the calcium appropriately.

Binding of one calcium to the loop in EF-hand I did not trigger the bioluminescence in the crystal state. The finding of this one Ca^{2+} at the first binding loop implies that the affinity here is much higher than for Ca^{2+} in the other binding loops at the C-terminal. This probably is because

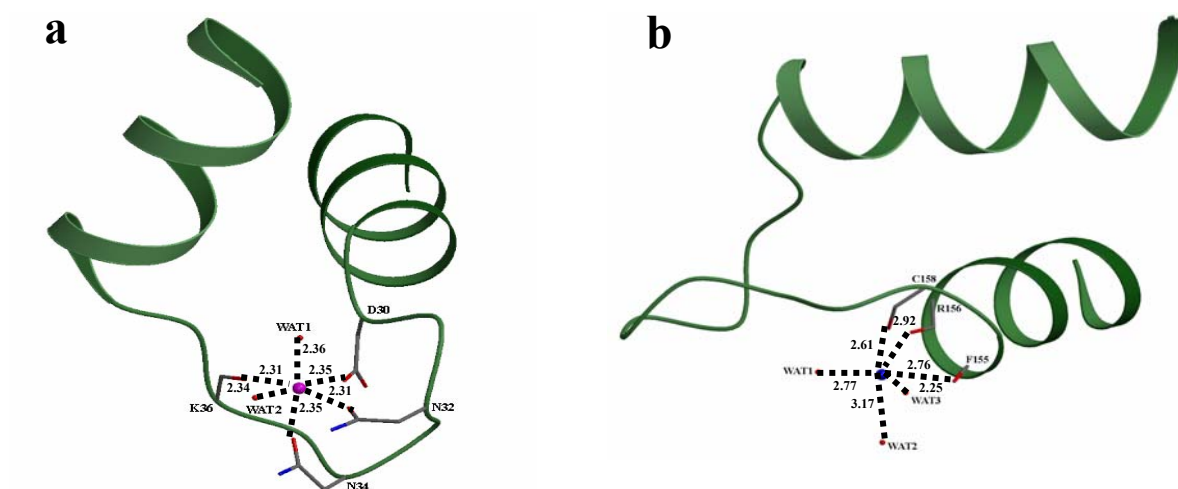


Fig. 3.14 (a) The calcium binding loop region of EF-hand motif I after soaking OL-obelin crystal with trace of calcium. The calcium ion (pink) has six coordinates, forming a bipyramidal calcium coordinate system. (b) The assumed potassium (blue) coordinate system near EF-hand motif IV of OL-obelin.

the geometry of the expected calcium-binding ligand amino acids at EF-hand motif I position are more optimized than those of EF-hand motifs III and IV. What is apparent is, to stretch the analogy, with this Ca^{2+} bound to loop I, the trigger is "cocked", mobility of atoms particularly the peroxy oxygens is restricted, and the photoprotein is readied for "action".

3.6 Crystal structure of Ca^{2+} -discharged W92F-obelin

The breakthrough in obtaining crystals of Ca^{2+} -discharged photoproteins came when screening a large number of site-directed mutants used previously for structure studies and characterization of Ca^{2+} -regulated photoproteins with over 250 different crystallization conditions for each concentration of each protein sample. Crystals of Ca^{2+} -discharged W92F-obelin were grown from a very concentrated droplet. The structure of Ca^{2+} -discharged W92F-obelin bound with the product of the bioluminescent reaction, coelenteramide, (conformation state IV in Fig. 3.1), was solved at 1.96 Å resolution. The W92F mutant of OL-obelin emits two bands of bioluminescence, one blue and the other violet, but the fluorescence of the Ca^{2+} -discharged protein is green the same as for wild type OL-obelin, and other biochemical properties are also the same. The three dimensional structures of W92F-obelin and OL-obelin are also practically identical, so it is valid to assume that comparing the two structures of W92F before and after the bioluminescence, generally represents the conformation change occurring in a photoprotein reaction.

As shown in Fig. 3.15, the Ca^{2+} -discharged W92F-obelin retains the overall scaffold of the undischarged photoproteins. A loop connects two domains of the molecule, N- and C- terminal domains. Each domain is formed by a pair of helix-loop-helix (HLH) motifs. The RMSD from the 185 $\text{C}\alpha$ -atomic positions of Ca^{2+} -discharged W92F-obelin vs. W92F-obelin is 0.306 Å, which shows the well conserved structural features between these two protein states, the one

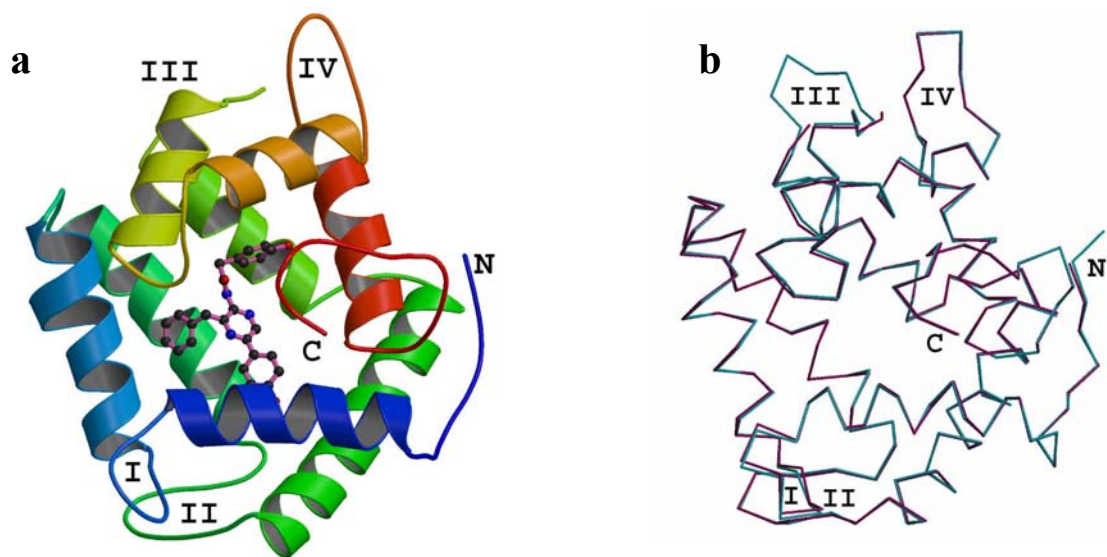


Fig. 3.15 (a) The crystal structure of Ca^{2+} -discharged W92F-obelin with coelenteramide (displayed by the ball-and-stick model in the center of the protein). (b) Superimposition of W92F-obelin (blue) and Ca^{2+} -discharged W92F-obelin (magenta). They greatly resemble each other with an overall RMSD of 0.306 Å.

primed with the hydroperoxy-coelenterazine and the other with the bioluminescence reaction product, coelenteramide. The first six residues at the N-terminus and the four residues at the loop region of the third HLH motif were not observed in the electron density maps and were assumed to be disordered. The binding pocket of coelenteramide is located at the center of the molecule.

3.7 Coelenteramide binding site

The coelenteramide fits precisely into the electron density (Fig. 3.16) and is buried in a highly hydrophobic cavity situated at the center of the protein structure in the same place as its precursor hydroperoxy-coelenterazine, surrounded by residues from each of the eight helices of the protein. This solvent inaccessible cavity apparently provides the necessary environment for efficient generation of the excited state product and its efficient fluorescence. Going from the hydroperoxy-coelenterazine to coelenteramide, the biggest structural change (Fig. 3.16) is in the reaction center around the C2 position (C2, O33 and C10, Fig. 3.17), resulting in an obvious orientational deviation of the phenol group at the C10 position. The other parts of the coelenterazine molecule also adjust their positions a little but not dramatically.

Fig. 3.16 and 3.17 show that key residues facing into the binding cavity and interacting with the hydroperoxy-coelenterazine, do not shift position after the reaction, except His175 and Tyr138. The His175 has flipped its imidazole ring perpendicular (63°) to its original conformation towards the hydroxyl group of Tyr190. The hydrogen bond distance changed from 2.64 Å to 2.54 Å. Tyr190 stayed at the same position connecting to O33 through hydrogen bonding with a water molecule. Tyr138 was in H-bonding interaction with the N1 of hydroperoxy-coelenterazine but has now moved away from the ligand molecule by translating 2.12 Å and rotating about 13° compared to its original position (Fig. 3.17, 4.33 Å). At the same time, one water molecule which was found to bridge Tyr138, His64 and Trp114 in the

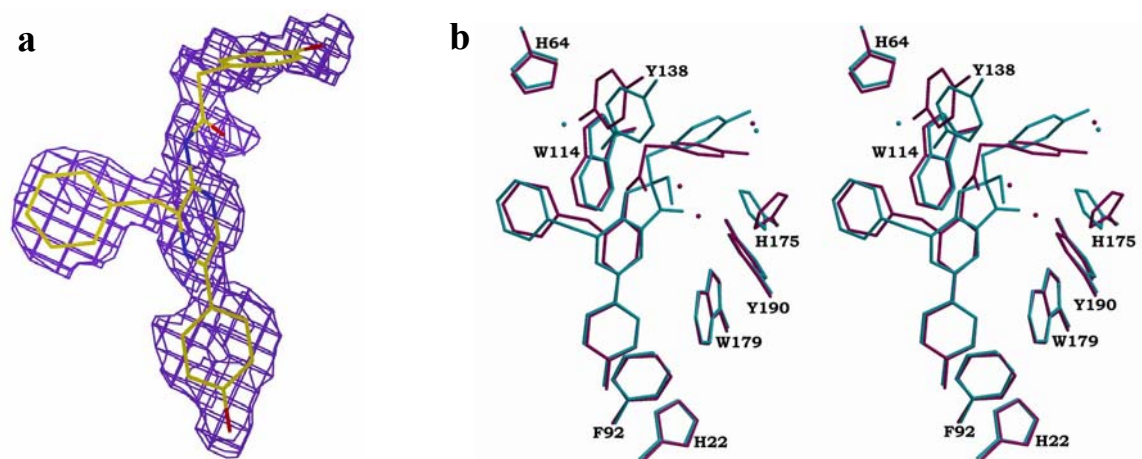


Fig. 3.16 (a) $2F_o - F_c$ electron density map contoured at 1.0σ of Ca^{2+} -discharged W92F-obelin bound coelenteramide. The model fits into the electron density precisely. (b) Stereoview of the superimposition of coelenterazine and coelenteramide with the key residues at their protein active sites. The W92F-obelin is in blue. The Ca^{2+} -discharged W92F-obelin is in magenta. Obvious conformational changes of coelenteramide within the reaction center can be observed. Water molecules are represented by balls.

Fig. 3.17 Two-dimensional drawing of the hydrogen bond (dash lines) network at the binding site of coelenteramide in Ca^{2+} -discharged W92F-obelin. The relative positions of F88 and F92 are also shown in dash-dot lines. Sequence alignment indicates that F88 in obelin corresponds to the Y82 in aequorin, which provides an extra H-bond in that case perhaps accounting for some of these spectral differences between these photoproteins. Distances are in Å: black, Ca^{2+} -discharged W92F-obelin; red, W92F-obelin.

coelenterazine binding site was replaced by Tyr138 in the coelenteramide binding site, resulting in a direct hydrogen bond interaction of Tyr138 with His64 and Trp114.

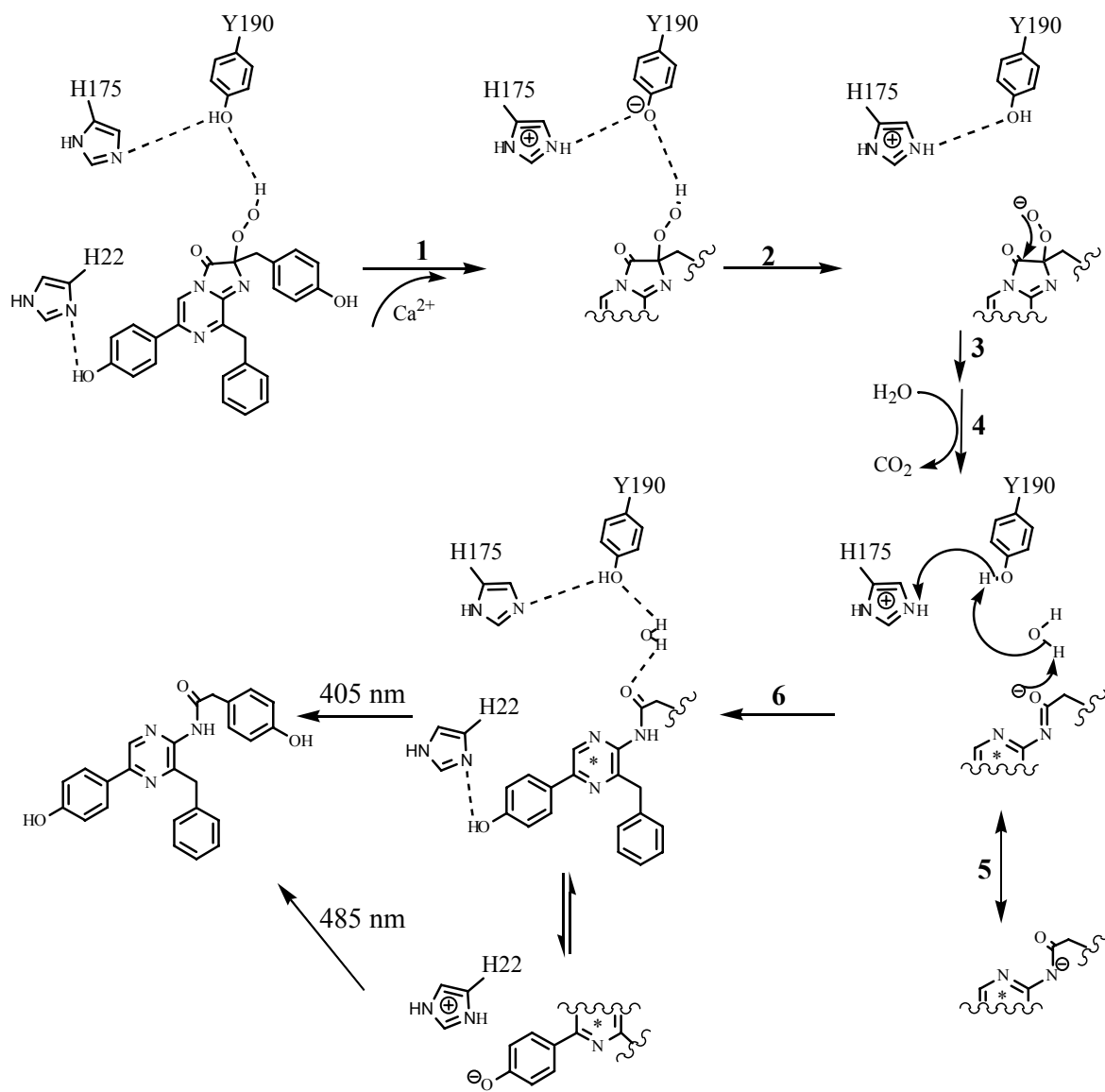
Water molecules are found in both hydroperoxy-coelenterazine and coelenteramide binding sites as shown in Fig. 3.16 and 3.17. There are three water molecules directly hydrogen bonded with the coelenteramide N4, O17 and O33, forming hydrogen bonding bridges between the protein residues, Tyr190, Trp179, Thr172 and Ile111, and coelenteramide. Among those water molecules, the water molecule connecting Ile111, Thr172 and O17 of coelenteramide was present at the same position in coelenterazine binding site. The other two water molecules were introduced by the bioluminescence reaction.

His22 is the only residue with a direct contact to coelenteramide by hydrogen bonding with the 5-p-hydroxyphenyl of coelenteramide with a distance of 2.82 Å (Fig. 3.16, Fig. 3.17). The same interaction was also found in the coelenterazine binding pocket with a distance of 2.84 Å. This remaining interaction implies the importance of His22 in the tuning of bioluminescence and fluorescence spectra of photoproteins. The mutation point Phe92 is located around 3.33 Å away from the 5-p-hydroxyphenyl of coelenteramide, similar to its position in the hydroperoxy-coelenterazine binding site.

3.8 Ca^{2+} -triggered bioluminescence reaction mechanism of photoproteins

As mentioned before, photoproteins presumably generate bioluminescence by a common mechanism proposed by McCapra and Chang (25) (section 1.7). A proton relay mechanism of Ca^{2+} -triggered bioluminescence reaction is proposed based on the structural studies of photoproteins and the detection of the hydrogen bond network in the substrate binding cavity (Scheme 3.1).

Scheme 3.1



Photoproteins all have typical EF-hand spatial structure characteristics but the Ca^{2+} -binding loop structures are not prepositioned for Ca^{2+} binding. Some movements of the residues must have to occur on Ca^{2+} binding to accommodate the coordinating atoms to the required 2.4 Å separation with bipyramidal pentagonal configuration. In photoproteins the first step in the generation of high intensity bioluminescence must be the binding of Ca^{2+} to the loops within the EF-hands. It is observed that certain residues within or near the exiting helices of loop IV, namely His175 and Tyr190 (Fig. 3.8), have critical proximity to the reactive center of the substrate, the peroxy oxygen atom at the C2 position of the chromophore hydroperoxy-coelenterazine. The second peroxy oxygen O34 of coelenterazine-oxygen complex is hydrogen bonded to the hydroxyl group of Tyr190 which itself is hydrogen bonded to the Nε2 of His175. Therefore any conformational adjustment in the binding loops accompanying Ca^{2+} binding can be expected to propagate into shifts of the hydrogen bond donor-acceptor separations around the coelenterazine, the networks apparently essential for the hydroperoxide stability, O34, Tyr190, and His175.

According to the McCapra and Chang mechanism, the hydroperoxy of the coelenterazine has to be deprotonated before the bioluminescence reaction breakdown. The best candidate for the hydroperoxide proton acceptor would be Tyr190. This could happen only when the hydroxyl proton of Tyr190 has been transferred to another proton acceptor, which leads to the assumption that His175 probably plays the essential role for triggering the bioluminescence reaction as in its location, the spatial displacement of His175 would disrupt the hydrogen bond network by which the hydroperoxy-coelenterazine had been stabilized. The notion that residue His175 in this position is very important for photoprotein activity is supported by observations on aequorin. Site-directed mutagenesis of the five histidine residues in aequorin has shown that substitution of

His169 (corresponding to His175 of obelin) to Ala, Phe, or Trp, leads to complete loss of activity, whereas modification of the remaining four histidine residues yielded mutant aequorins with varying bioluminescence activities (29).

The visualization of the coelenteramide binding pocket now allows a precise description of the proposal for the mechanism (Scheme 3.1) by which Ca^{2+} binding to EF-hand loops, sets off the bioluminescence reaction. The hypothesis for the triggering is that the accommodation of calcium binding to the loops requires loop residue displacement. This movement is transmitted to the helices of EF-hand motif IV where the critical residue His175 is located. The His175 flips its imidazole ring perpendicular (63°) to its original conformation towards the hydroxyl group of Tyr190. With the caution that a change of this H-bond distance from 2.64 Å to 2.54 Å is hardly significant at 1.96 Å resolution, reducing this separation would make the H-bond stronger (78) equivalent to a partial protonation of the His175, leaving Tyr190 then partially in the phenolate form. It is simplified in Scheme 3.1 (step 1) by a complete proton transfer from Tyr190 to His175. As a consequence, it would contribute to a stabilization of the activation barrier, leading to catalysis of complete proton transfer from the hydroxyl group of Tyr190 to His175. Since the phenol and hydroperoxy groups have similar pKs, proton transfer from the hydroperoxy is feasible (step 2) leaving the peroxy anion which can immediately undergo nucleophilic attack on the C3-position of the coelenterazine to form the dioxetanone intermediate (step 3). The uphill reaction sequence would be favored by the exergonicity from the irreversible formation of the dioxetanone.

In the product structure (Fig. 3.17) it is also seen that a new water molecule has moved into position bridging Tyr190 and O33 through hydrogen bonds. The primary excited state in the reaction in Scheme 1.2 is the coelenteramide amide anion but within the protein cavity this

would be rapidly protonated to the excited state of the neutral coelenteramide. In the photoprotein Tyr138 makes an H-bond to the N1-atom of hydroperoxy-coelenterazine (2.68 Å) and is an obvious first choice as a proton donor. However as Tyr138 moves away to 4.33 Å after reaction (Fig. 3.17), a more likely candidate would be for the one that protonated His175 (scheme 3.1) to return to Tyr190 with the bridging water molecule now conveniently located to relay a proton to O33 of the coelenteramide (step 6). Thus the proton relay cycle is completed and the photoprotein is essentially “turned over” like an enzyme. Turnover has in fact been demonstrated in the case of aequorin (79).

The present structure does not reveal the loop residue displacements required in the trigger mechanism because, although the crystal was grown from the protein solution in the presence of calcium, the precipitant used for successfully obtaining a good crystal was sodium citrate which is a strong cation chelator. Consequently the crystal contains no bound Ca^{2+} and therefore corresponds to conformation state IV in Fig. 3.1. However the crystals do give a green fluorescence similar to that of a solution of Ca^{2+} -discharged W92F obelin with or without the presence of calcium ions (Fig. 3.18). Calcium binding shifts the fluorescence maximum to longer wavelength only by 10 nm so it is reasonable to assume that the coelenteramide binding environment in the current crystal structure (conformation state IV of Fig. 3.1) is practically the same in the Ca^{2+} loaded state (conformation state III of Fig. 3.1). It is probable that the difference in the NMR spectra observed for these two conformation states arises from the displacement of residues involved in the Ca^{2+} binding.

The suggested mechanism (Scheme 3.1) also provides a reasonable explanation for the origin of calcium-independent luminescence. Given the energetics of hydrogen bonding thermally driven structural fluctuation suffice to trigger Ca^{2+} -independent light emission, it is conceivable

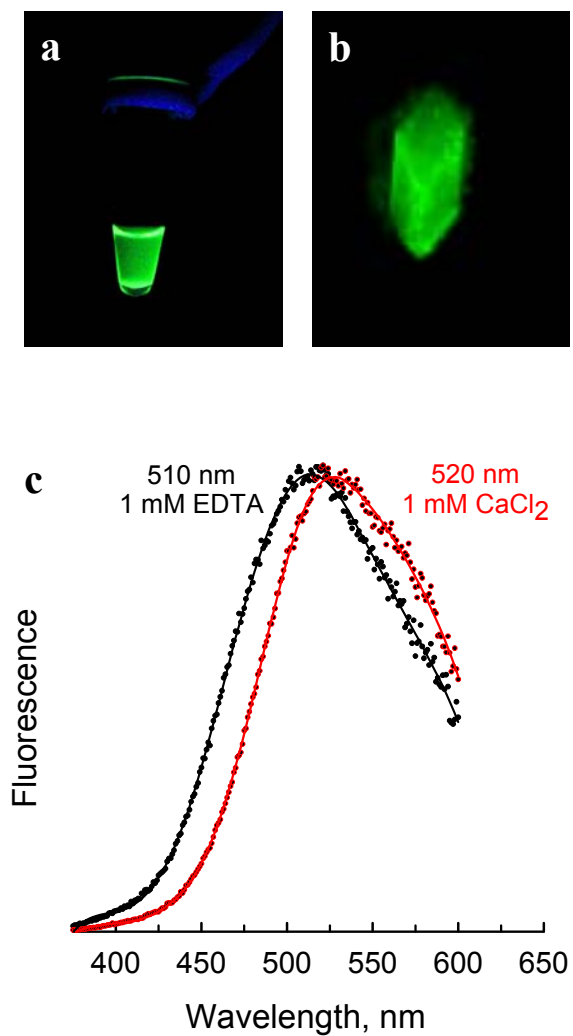


Fig. 3.18 A solution of Ca^{2+} -discharged W92F-obelin (a) and the crystal (b) show the same green fluorescence on excitation by near UV. (c) The fluorescence spectrum of Ca^{2+} -discharged W92F-obelin in solution in the presence of Ca^{2+} (red) shifts 10 nm to shorter wavelength on removal of the Ca^{2+} (black).

that proton transfer from Tyr190 to His175 could also occur in the absence of Ca^{2+} through spontaneous motion of these residues resulting in a decrease in H-bond distance and the extremely small probability of this occurrence would probably account for this phenomenon. The mobility of coelenterazine at the active sites of photoproteins was detected and described in section 3.4.

3.9 Bioluminescence emitter

The bioluminescence spectral distributions of Ca^{2+} -regulated photoproteins are broad with maxima depending on the types in the range 465 - 495 nm. Identification of the bioluminescence emitter is a key requirement in solving the reaction mechanism. In the past the approach taken has been to study the chemiluminescence reaction of the substrate or close substrate analogues, under various solution conditions that might mimic the protein reaction site environment as mentioned in section 1.6. To qualify as the product of the light reaction, the candidate should have a fluorescence spectrum the same as the bioluminescence as well as a fluorescence quantum yield greater than the bioluminescence quantum yield. In the earliest study, McCapra and Chang (25) proved the mechanism of the chemiluminescence of the coelenterazine analogue in Scheme 1.2. In DMSO and a strong base such as potassium *tert*-butoxide, chemiluminescence of this compound occurs with a spectral emission maximum at 455 nm. The amide is produced in high yield, and under the same basic solution conditions, the fluorescence spectrum of the synthesized amide matches the chemiluminescence spectrum. The fluorescing species was identified as the excited amide anion

The McCapra and Chang mechanism (Scheme 1.2) (25) demands that the amide anion is the initially formed excited state and is the proven emitter of the chemiluminescence in aprotic medium. However, in bioluminescence systems the excited coelenteramide anion resides in an

active site of a protein, and the attendant hydrogen bond interactions with surrounding residues allow the possibility of excited state proton transfer changing the nature of the emitting state. Such processes are known to be fast, on the nanosecond timescale and therefore highly competitive with the radiative rate of the amide anion.

The first proton transfer to be expected is to form the excited neutral coelenteramide. An amide group is only dissociable in an aprotic solvent under very basic conditions, and although it is known that its excited state pK could be lowered over the ground state; the proton sources in the binding site (Fig. 3.17) would favor rapid formation of the neutral excited state. As discussed in section 3.7, the most convenient proton source for generating the neutral species is the water molecule bridging O33 and Tyr190. Tyr138 was originally thought to donate its proton to N1 of coelenterazine in hydroperoxy-coelenterazine binding pocket, is not likely to be important for the bioluminescence emissions of photoproteins as it moves away from coelenteramide, and probably serves to position the coelenterazine.

Another position for excited state proton transfer is at the 5-p-hydroxy-phenyl substituent of coelenteramide. As shown in Fig. 3.17, His22 of Ca^{2+} -discharged W92F-obelin remains in the same position and is ready to accept proton from the hydroxyl group of the 5-p-hydroxy-phenyl substituent. The excited state pK of phenol is known to be lowered by several units from the ground state pK and so the excited state pK of the 5-hydroxy-phenyl substituent could be near or even less than the pK of the proximate His22. Transfer of the H atom in the H-bond to protonate His22 will then yield the excited ion-pair state.

Thus, the possibility of proton transfer at these two positions around coelenteramide could result in a mixture of neutral and ion-pair excited states and in fact, the bioluminescence emissions from the obelins (14, 22) as well as the *Renilla* luciferase reaction (80), all show a

small contribution from a 400 nm band on the major band around 480 nm. The cited model fluorescence studies assign the 480 nm band as from a coelenteramide phenolate ion-pair excited state (23) and 400 nm from the neutral species. A deuterium effect and the influence of pH on the relative population of the excited-state species of W92F-obelin (32) give additional support to these excited state proton transfer ideas. In conclusion, it is proposed that the bioluminescence of wild type obelin originates from the coelenteramide phenolate ion-pair excited state with a small admixture of the neutral excited state, both rapidly formed from the primary excited amide anion.

As a matter of record, Shimomura et al. (79, 81) were the first to observe a bimodal bioluminescence spectrum from what they called *e*-aequorin, an artificial photoprotein formed by charging the apoaquorin with a coelenterazine analogue containing an extra ethano bridge. Ohmiya et al. (30) produced the aequorin mutant W86F, which showed an enhanced 405 nm contribution just using coelenterazine but this 405 nm band became dominant if the fused ring analogue was employed. Without structural information it is impossible to know what is going on with the analogues but a steric effect is likely, where the fused ring structure alters the separation of the *p*-hydroxyl group and the His16 in the case of aequorin. The contribution of the 405 nm emission originating from the excited state of the neutral coelenteramide in the corresponding mutant of obelin, W92F, is even greater than that in W86F aequorin (31). The idea of dual proton transfer after primary excited amide anion formation and of the 480 nm bioluminescence originating from the excited phenolate in an ion-pair interaction with the His22 also allows other observations to fall into place. At a pH below 7, the His22 is protonated and can no longer accept a proton to form the excited phenolate in competition with the radiative rate from the neutral excited coelenteramide. Hydrogen bonding from the Trp92 in OL-obelin could

stabilize the ion-pair state favoring longer wavelength emission, and that is why its removal produces the violet bioluminescence and the bimodal emission spectrum. We need to caution here that it is not possible to exclude that this substitution could change dielectric properties of the environment of the p-hydroxyl group. As a result the relevant pK's could be changed with resulting influence on the proton-transfer rate from p-hydroxyl group to His22.

The fluorescence spectra of OL-obelin and W92F-obelin are both monomodal with the same 510-nm maximum (31). As the bioluminescence and fluorescence are believed to originate from the same excited state species, this 25-nm red-shift of the fluorescence from the bioluminescence (485 nm) would result if the final structure of the binding cavity was more polar than that at the instant of bioluminescence emission. Model studies do show that the fluorescence spectra of coelenteramide anions shift to longer wavelength in more polar solvents. It is noted that in comparison with hydroperoxy-coelenterazine in the photoprotein cavity, two new water molecules are now in H-bonding interactions with the coelenteramide (Fig. 3.17). If the water molecule interacting with N4 were to arrive there only in some relaxation period following the bioluminescence emission, then plausibly this would render the environment more polar for the fluorescence measurement than during the bioluminescence.

Why in aequorin the product fluorescence spectral distribution is the same as the bioluminescence (465 nm) (21), but the obelin product fluorescence (510 nm) is shifted to longer wavelength than its bioluminescence (485 nm for OL-obelin, 495 nm for OG-obelin) (14, 22), is not clear yet. Without a structure of Ca^{2+} -discharged aequorin but taking into account that aequorin and obelin are homologous, it is assumed that the cavity structure of Ca^{2+} -discharged aequorin is similar to that of Ca^{2+} -discharged W92F-obelin. The answer could be in small variation of hydrogen bond distances between coelenterazine and key residues in these

photoproteins that can influence proton transfer rates of the excited coelenteramide, and in the structure of the cavity enclosing the excited coelenteramide immediately following the bioluminescence triggering that can provide both an appearance of new hydrogen bonds with excited coelenteramide and a change of polarity of the environment. It is well known that the fluorescence energy level of a molecule can be influenced by the surroundings, solvent polarity, viscosity, etc. An extra hydrogen bond formed by Tyr82 (corresponding to Phe88 in W92F-obelin, Fig. 3.17) and the coelenteramide could be a critical factor accounting for the spectral overlap of aequorin bioluminescence and its product fluorescence by tuning the effective dielectric constant of the active site.

3.10 Structures of Ca^{2+} -loaded apo-aequorin and Ca^{2+} -loaded apo-obelin

Ca^{2+} -binding proteins are probably one of the most extensively studied protein families. The main reason for such attention above all is because these proteins regulate numerous vital intracellular events in living cells and organisms. They face a variety of completely different environments. Hence these Ca^{2+} -binding proteins have affinities and kinetics of binding and releasing calcium ions that finely tuned to exert their precise biological functions, such as directing Ca^{2+} -dependent enzyme activities, modulating the calcium signal and controlling calcium homeostasis in the cell (82). The members of this family are related not necessarily by any similarity in function but by the fact that most of them selectively bind calcium through a homologous structural unit known as the "EF-hand" (15, 26).

EF-hand proteins are distinguished from other Ca^{2+} -binding proteins in that its members have a common calcium binding helix-loop-helix motif consisting of two helices that flank a "canonical" sequence loop region of 12 contiguous residues from which the oxygen ligands for the calcium ion are derived. Calcium ions is co-ordinated in a pentagonal bipyramidal array with

about 2.4 Å separation to side-chain oxygens from residues numbered 1,3,5,7,9, and 12 in the loop (56). Oxygen ligands are generally found in carboxylic side-chain groups of aspartate or glutamate or carbonyl oxygen atoms of the main-chain or side chain of asparagine and glutamine. This allows the formation of bonds with oxygen groups from different amino acids of a protein thus generating a particular three-dimensional complex structure. The Ca^{2+} -induced conformational change of a protein can lead to the activation or inhibition of enzymatic activities or the binding to specific ligands either in the presence or absence of bound calcium (83).

Another distinguishing characteristic of the EF-hand protein family is the fact that the Ca^{2+} -binding occur in intimately linked pairs (56). This double EF-hand motif appears to be a basic structural feature of all EF-hand Ca-binding proteins. The short β -type interactions between the two binding loops is important for the pairing (84). This organization seems to be important for correct structural folding during protein synthesis and also strongly affects the functional properties of individual calcium-binding sites. It is thought to increase the affinity of each EF-hand for calcium and to provide a means for cooperativity in binding calcium to two loops. The tendency of EF-hands to pair together is so strong that synthetic peptides with the sequence of an EF-hand will form homodimers in solution (85), and if peptides representing the two EF-hands normally paired together in a protein are mixed, they will preferentially pair to form the 'natural' heterodimer (86).

The primary sequence of aequorin determined after almost 25 years later than its discovery revealed the canonical Ca^{2+} -binding EF-hands (11, 12), and this is now known to be common in the sequences of photoproteins from other genera, such as obelin (13, 14), indicating that they are members of the EF-hand Ca^{2+} -binding protein family. The structures of Ca^{2+} -loaded apo-

aequorin and Ca^{2+} -loaded apo-obelin (conformation state V of Fig. 3.1) solved at 1.7 Å and 2.2 Å respectively further confirmed the EF-hand characteristics of photoproteins.

There are two ways to obtain Ca^{2+} -loaded apo-photoproteins *in vitro*. One is to add Ca^{2+} directly to an apo-photoprotein solution (35). The other is to remove the coelenteramide from a Ca^{2+} -loaded Ca^{2+} -discharged photoproteins. The resulting structures of Ca^{2+} -loaded apo-photoproteins may not be the same. The Ca^{2+} -loaded apo-photoproteins described here were obtained unintentionally by removing the coelenteramide which was extracted by organic solvent precipitants during the crystallization process (see details in section 2.2.2).

As shown in Fig. 3.19, the crystal structures of Ca^{2+} -loaded apo-aequorin and Ca^{2+} -loaded apo-obelin still maintain the overall globular folding as seen in those of aequorin and obelins. Two clearly separated domains, N-terminal and C-terminal domain, are connected by a short loop region. The N-termini of both proteins are disordered. The C-terminus of Ca^{2+} -loaded apo-obelin is not observed in the electron density map as well and assumed to be disordered. The protein consists of four HLH motifs arranged in pairs to form the globular molecule. Three calcium ions are found at the expected Ca^{2+} -binding sites of the three EF-hand motifs. Although the second HLH motif can not bind calcium ion, it has the three-dimensional architecture of EF-hand motifs and pairs with the first Ca^{2+} -binding EF-hand motif, which meets the typical structural characteristic of being EF-hand motifs.

The second helix of each EF-hand motif is oriented almost perpendicular to the first one. The canonical loop region tightly holds calcium ion. All Ca^{2+} -binding sites (Fig. 3.20) of the two proteins contribute six oxygen ligands derived from the carboxylic side-chain groups of aspartate and glutamate residues, carbonyl groups of the peptide backbone or side-chain of asparagine to the metal ion with an average distance of separation of about 2.39, 2.52 Å in Ca^{2+} -loaded apo-

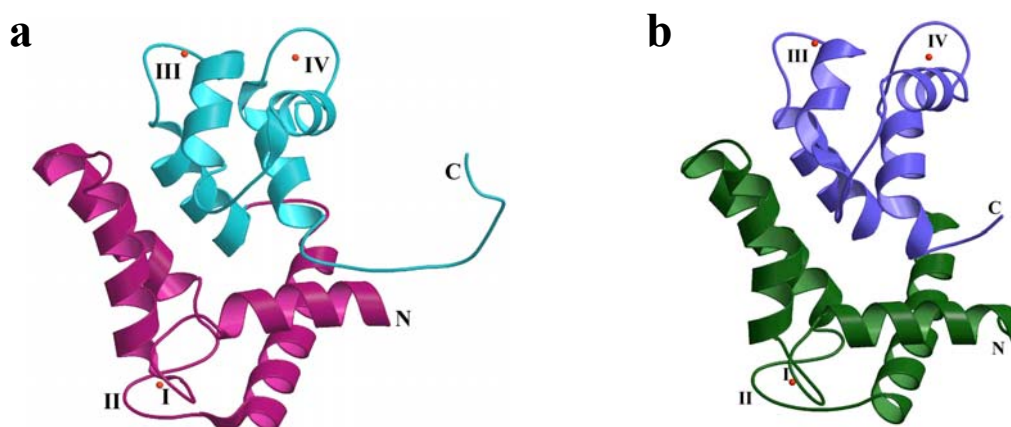


Fig. 3.19 The crystal structures of (a) Ca^{2+} -loaded apo-aequorin, and (b) Ca^{2+} -loaded apo-obelin. Calcium ions are represented by red balls. Part of C-terminus of Ca^{2+} -loaded apo-obelin is disordered. The well defined N-terminal (magenta, green) and C-terminal domains (cyan, blue) are easily observed.

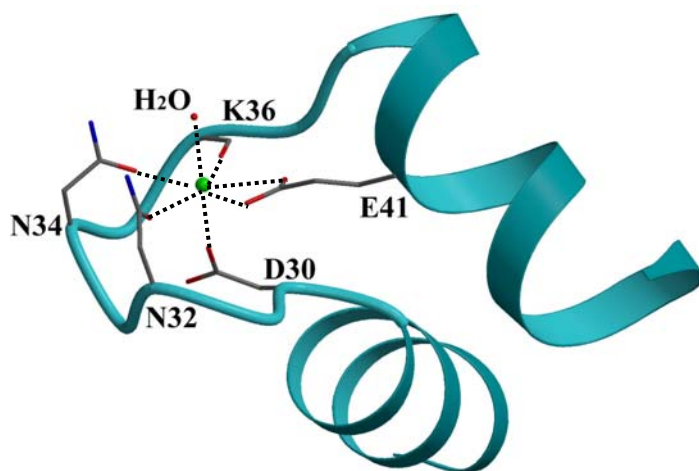


Fig. 3.20 The Ca²⁺-binding site at the loop region of EF-hand motif I of Ca²⁺-loaded apo-obelin shows typical pentagonal bipyramid Ca²⁺-binding geometry. All other Ca²⁺-binding sites in Ca²⁺-loaded apo-obelin and Ca²⁺-loaded apo-aequorin all look similar as the one shown here.

aequorin and Ca^{2+} -loaded apo-obelin respectively. The seventh ligand comes from the oxygen atom of a water molecule. The geometrical arrangement of the ligands can be described as a pentagonal bipyramid with the Ca^{2+} ion occupying the center of the pyramid.

Although Ca^{2+} -loaded apo-aequorin and Ca^{2+} -loaded apo-obelin share a similar overall fold, easily observed local structure shifts happen on almost every part of the structures, the helices regions and the loop regions (Fig. 3.21a). Unlike apo-protein bound with coelenterazine (conformation state II of Fig. 3.1), obvious deviations occur at the solvent exposed loop regions of EF-hand motifs III and IV. The RMSD between the $\text{C}\alpha$ atoms of the two Ca^{2+} -loaded apo-protein structures (from residue 15 to 181 of Ca^{2+} -loaded apo-obelin) is 2.27 Å, whereas the RMSD of the same residue range between aequorin and OL-obelin, for example, is only 1.48 Å. This difference shows that the Ca^{2+} -binding does introduce appreciable different conformational changes depending on photoprotein species, hence suggests different influences on other Ca^{2+} -binding related characteristics of photoproteins, such as different Ca^{2+} -binding kinetics of photoproteins.

Since the Ca^{2+} -loaded apo-aequorin has more complete structural features than the Ca^{2+} -loaded apo-obelin at the resolutions obtained, the Ca^{2+} -loaded apo-aequorin and aequorin are used for the purpose of structural comparison. Fig. 3.21b shows that the N-terminal domain of aequorin is more conserved after Ca^{2+} binding than C-terminal domain. Despite the obvious deviations observed between the whole structures of Ca^{2+} -loaded apo-aequorin and aequorin (RMSD is 2.04 Å of $\text{C}\alpha$ atoms from residue 11 to 177 of Ca^{2+} -loaded apo-aequorin) (Fig. 3.21), there are two major differences between them. The biggest difference is located at their C-termini. In the aequorin structure, the C-terminus caps the ligand within the protein molecule, producing a solvent inaccessible chromophore binding site, apparently providing the necessary

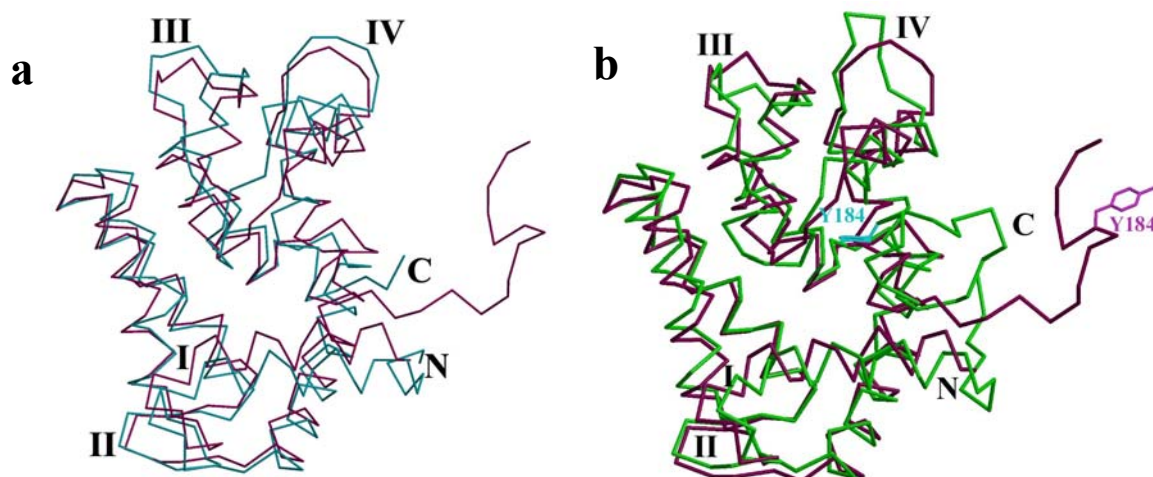


Fig. 3.21 (a) Superimposition of Ca²⁺-loaded apo-aequorin (magenta) and Ca²⁺-loaded apo-obelin (cyan). Obvious local structure shifts happen on almost every part of the structures, the helices regions and the loop regions. (b) Superimposition of Ca²⁺-loaded apo-aequorin (magenta) and aequorin (light green). Y184 which has direct hydrogen bonding contact with the hydroperoxy group of coelenterazine substrate of aequorin is marked in light magenta in Ca²⁺-loaded apo-aequorin, in cyan in aequorin.

environment to generate the excited state product for its efficient bioluminescence and fluorescence, whereas, in the Ca^{2+} -loaded apo-aequorin structure, the C-terminus opens up, as a consequence, the ligand could easily dissociate as is known. Another noticeable difference is that aequorin stretches out its whole molecule in order to make enough space to accommodate the ligand, however without the chromophore coelenteramide bound inside, Ca^{2+} -loaded apo-aequorin squeezed the whole molecule in by bringing the eight helices closer, especially helices from EF-hand motifs III and IV. The same is also found between Ca^{2+} -loaded apo-obelin and Ca^{2+} -discharged W92F-obelin. These observations show that Ca^{2+} -binding does introduce noticeable residue displacements to photoproteins, but the detailed process of how the Ca^{2+} -binding triggers the bioluminescence reaction of photoproteins is still not clear without the structure of Ca^{2+} -loaded Ca^{2+} -discharged photoproteins (conformation state III of Fig. 3.1).

REFERENCES

1. Herring, P. J. (1978) *Bioluminescence in Action* (Academic Press, London).
2. White, E. H., Rapaport, E. & Schliger, H. H. (1971) *Biorg. Chem.* **1**, 92-122.
3. Shimomura, O. & Johnson, F. H. (1966) in *Bioluminescence in Progress*, eds. Johnson, F. H. & Haneda, Y. (Princeton University Press, pp. 495-521.
4. Shimomura, O., Johnson, F. H. & Saiga, Y. (1962) *J. Cell Comp. Physiol.* **59**, 223-239.
5. Morin, J. G. & Hastings, J. W. (1971) *J Cell Physiol.* **77**, 305-312.
6. Morin, J. G. (1974) in *Coelenterate Biology: Reviews and New Perspectives*, eds. Muscatine, L. & Lenhoff, H. M. (Academic Press, New York), pp. 397-438.
7. Shimomura, O. & Johnson, F. H. (1975) *Nature* **256**, 236-238.
8. Ohmiya, Y. & Hirano, T. (1996) *Chem. Biol.* **3**, 337-347.
9. Hastings, J. W., Mitchell, G., Mattingly, P. H., Blinks, J. R. & Van Leeuwen, M. (1969) *Nature* **222**, 1047-1050.
10. Allen, D. G., Blinks, J. R. & Prendergast, F. G. (1977) *Science* **195**, 996-998.
11. Inouye, S., Noguchi, M., Sakaki, Y., Takagi, Y., Miyata, T., Iwanaga, S. & Tsuji, F. I. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 3154-3158.
12. Charbonneau, H., Walsh, K. A., McCann, R. O., Prendergast, F. G., Cormier, M. J. & Vanaman, T. C. (1985) *Biochemistry* **24**, 6762-6771.
13. Illarionov, B. A., Bondar, V. S., Illarionova, V. A. & Vysotski, E. S. (1995) *Gene* **153**, 273-274.

14. Markova, S. V., Vysotski, E. S., Blinks, J. R., Burakova, L. P., Wang, B.-C. & Lee, J. (2002) *Biochemistry* **41**, 2227-2236.
15. Moncrief, N. D., Kretsinger, R. H. & Goodman, M. (1990) *J. Mol. Evol.* **30**, 522-562.
16. Inouye, S. & Tsuji, F. I. (1993) *FEBS Lett* **315**, 343-346.
17. Fagan, T. F., Ohmiya, Y., Blinks, J. R., Inouye, S. & Tsuji, F. I. (1993) *FEBS Lett* **333**, 301-305.
18. Watterson, D. M., Sharief, F. & Vanaman, T. C. (1980) *J. Biol. Chem.* **255**, 962-975.
19. Tsuji, F. I., Ohmiya, Y., Fagan, T. F., Toh, H. & Inouye, S. (1995) *Photochem Photobiol* **62**, 657-661.
20. Shimomura, O. & Johnson, F. H. (1978) *Proc Natl Acad Sci U S A* **75**, 2611-2615.
21. Shimomura, O. & Johnson, F. H. (1970) *Nature* **227**, 1356-1357.
22. Markova, S. V., Vysotski, E.S. & Lee, J. (2001) in *Bioluminescence & Chemiluminescence 2000* (World Scientific Publishing Company, Singapore, pp. 115-118.
23. Shimomura, O. & K., T. (2000) *Luminescence* **15**, 51-58.
24. Imai, Y., Shibata, T., Maki, S., Niwa, H., Ohashi, M. & Hirano, T. (2001) *Photochem. Photobiol. A: Chemistry* **146**, 95-107.
25. McCapra, F. & Chang, Y. C. (1967) *Chem. Commun.* **19**, 1011-1012.
26. Kawasaki, H., Nakayama, S. & Kretsinger, R. H. (1998) *Biometals* **11**, 277-295.
27. Nomura, M., Inouye, S., Ohmiya, Y. & Tsuji, F. I. (1991) *FEBS Lett* **295**, 63-66.
28. Kurose, K., Inouye, S., Sakaki, Y. & Tsuji, F. I. (1989) *Proc Natl Acad Sci U S A* **86**, 80-84.
29. Ohmiya, Y. & Tsuji, F. I. (1993) *FEBS Lett* **320**, 267-270.

30. Ohmiya, Y., Ohashi, M. & Tsuji, F. I. (1992) *FEBS Lett* **301**, 197-201.
31. Deng, L., Vysotski, E. S., Liu, Z.-J., Markova, S. V., Malikova, N. P., Lee, J., Rose, J. & Wang, B.-C. (2001) *FEBS Lett.* **506**, 281-285.
32. Vysotski, E. S., Liu, Z.-J., Markova, S. V., Blinks, J. R., Deng, L., Frank, L. A., Herko, M., Malikova, N. P., Rose, J. P., Wang, B.-C. & Lee, J. (2003) *Biochemistry* **42**, 6013-6024.
33. Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000) *Protein Sci.* **11**, 2085-2093.
34. Head, J. F., Inouye, S., Teranishi, K. & Shimomura, O. (2000) *Nature* **405**, 372-376.
35. Lee, J., Glushka, J. N., Markova, S. V. & Vysotski, E. S. (2001) in *Bioluminescence & Chemiluminescence 2000*, eds. Case, J. F., Herring, P. J., Robison, B. H., Haddock, S. H. D., Kricka, L. J. & Stanley, P. E. (World Scientific Publishing Company, Singapore), pp. 99-102.
36. Illarionov, B. A., Frank, L.A., Illarionova, V.A., Bondar, V.S., Vysotski, E.S. & Blinks, J.R. (2000) *Methods Enzymol* **305**, 223-249.
37. Jancarik, J. & Kim, S.-H. (1991) *J. Appl. Cryst.* **24**, 409-411.
38. Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990) *J. Appl. Cryst.* **23**, 297-302.
39. D'Arcy, A., Elmore, C., Stihle, M. & Johnston, J. E. (1996) *J. Crystal Growth* **168**, 175-180.
40. Teng, T.-Y. (1990) *J. Appl. Cryst.* **23**, 387-391.
41. Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol* **276**, 307-326.
42. Wang, B. C. (1985) *Methods Enzymol* **115**, 90-112.

43. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54**, 905-921.
44. Jones, T. A., Zou, J. Y., Cowan, S. W. & M. Kjeldgaard (1991) *Acta Crystallogr A* **47**, 110-119.
45. Terwilliger, T. C. & Berendzen, J. (1999) *Acta Crystallogr D Biol Crystallogr* **55**, 849-861.
46. Terwilliger, T. C. (2000) *Acta Cryst.* **D56**, 965-972.
47. McRee, D. E. (1999) *J. Struct. Biol.* **125**, 156-165.
48. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997) *Acta Cryst. D* **53**, 240-255.
49. Luzzati, V. (1952) *Acta Crystallogr.* **5**, 802-810.
50. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283-291.
51. Sheldrick, G. M. & Schneider, T. R. (1997) *Methods Enzymol* **277**, 319-343.
52. Lovell, S. C., Davis, I. W., Arendall III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003) *Proteins* **50**, 437-450.
53. Kretsinger, R. H. & Nakayama, S. (1993) *J Mol Evol* **36**, 477-488.
54. Van Eerd, J. P. & Takahashi, K. (1976) *Biochemistry* **15**, 1171-1180.
55. Wnuk, W., Schoechlin, M. & Stein, E. A. (1984) *J Biol Chem* **259**, 9017-9023.
56. Strynadka, N. C. & James, M. N. (1989) *Annu Rev Biochem* **58**, 951-998.
57. Bondar, V. S., Frank, L. A., Malikova, N. P., Inzhevatkin, E. V., Illarionova, V. A. & Vysotski, E. S. (1999) *Bioluminescent activity of the recombinant obelin after chemical modification of histidine and cysteine residues* (John Wiley & Sons, Chichester, UK).

58. Shimomura, O. & Johnson, F. H. (1972) *Biochemistry* **11**, 1602-1608.
59. Shimomura, O., Musicki, B. & Kishi, Y. (1989) *Biochem J* **261**, 913-920.
60. Shimomura, O., Saiga, Y. & Johnson, F. H. (1963) *J. Cell Compar. Physiol.* **62**, 9-15.
61. Levine, L. D. & Ward, W. W. (1982) *Comp. Biochem. Physiol* **B72**, 77-85.
62. Campbell, A. K. (1974) *Biochem J* **143**, 411-418.
63. Stephenson, D. G. & Sutherland, P. J. (1981) *Biochim Biophys Acta* **678**, 65-75.
64. Bondar, V. S., Trofimov, K.P., Vysotski, E.S. (1992) *Biochemystry-Moscow* **57**, 1020-1027.
65. Ward, W. W. & Seliger, H. H. (1974) *Biochemistry* **13**, 1491-1499.
66. Ward, W. W. & Seliger, H. H. (1974) *Biochemistry* **13**, 1500-1510.
67. Prasher, D., McCann, R. O. & Cormier, M. J. (1985) *Biochem Biophys Res Commun* **126**, 1259-1268.
68. Prasher, D. C., McCann, R. O., Longiaru, M. & Cormier, M. J. (1987) *Biochemistry* **26**, 1326-1332.
69. Blinks, J. R. & Caplow, D. D. (1991) *Physiologist* **34**, 110.
70. Shimomura, O. & Shimomura, A. (1985) *Biochem J* **228**, 745-749.
71. Hart, R. C., Stempel, K. E., Boyer, P. D. & Cormier, M. J. (1978) *Biochem Biophys Res Commun.* **81**, 980-986.
72. Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000) *Proc Natl Acad Sci U S A.* **97**, 623-628.
73. Swaminathan, G. J., Holloway, D. E., Veluraja, K. & Acharya, K. R. (2002) *Biochemistry* **41**, 3341-3352.

74. Longhi, S., Czjzek, M., Lamzin, V., Nicolas, A. & Cambillau, C. (1997) *J. Mol. Biol.* **268**, 779-799.
75. Bonagura, C. A., Bhaskar, B., Sundaramoorthy, M. & Poulos, T. L. (1999) *J. Biol. Chem.* **274**, 37827-37833.
76. Bode, W., Schwager P. (1975) *J Mol Biol.* **98**, 693-717.
77. Holmes, M., Matthews BW. (1982) *J Mol Biol.* **160**, 623-639.
78. Jeffrey, G. A. (1997) *An introduction to hydrogen bonding* (Oxford University Press, New York, NY).
79. Shimomura, O., Musicki, B. & Kishi, Y. (1988) *Biochem J* **251**, 405-410.
80. Hori, K., Charbonneau, H., Hart, R. C. & Cormier, M. J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4285-4287.
81. Shimomura, O. (1995) *Biol Bull* **189**, 1-5.
82. Nelson, M. R. & Chazin, W. J. (1998) *Biometals* **11**, 297-318.
83. Schwaller, B. (2001) in *Nature Encyclopedia of Life Sciences (April)* (Nature Publishing Group, London).
84. Nelson, M. R. & Chazin, W. J. (2000) in *Nature Encyclopedia of Life Sciences (July)* (Nature Publishing Group, London).
85. Shaw, G. S., Hodges, R. S. & Sykes, B. D. (1990) *Science* **249**, 280-283.
86. Finn, B. E., Kördel, J., Thulin, E., Sellers, P. & Forsén, S. (1992) *FEBS Lett.* **298**, 211-214.

APPENDIX A

Crystallization and Preliminary X-ray Analysis of GlcNAc α 1,4Gal-releasing Endo- β -galactosidase from *Clostridium perfringens*

Crystallization and preliminary X-ray analysis of GlcNAc α 1,4Gal-releasing endo- β -galactosidase from *Clostridium perfringens*

Lu Deng,^a Zhi-Jie Liu,^b Hisashi Ashida,^c Su-Chen Li,^c Yu-Teh Li,^c Peter Horanyi,^b Wolfram Tempel,^b John Rose^b and Bi-Cheng Wang^{b*}

^aDepartment of Chemistry, University of Georgia, Athens, GA 30602, USA, ^bDepartment of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA, and ^cDepartment of Biochemistry, Tulane University Health Sciences Center School of Medicine, New Orleans, Louisiana 70112, USA

Correspondence e-mail:
 wang@bcl1.bmb.uga.edu

The unique clostridial endo- β -galactosidase (Endo- β -Gal_{GnGa}) capable of releasing the disaccharide GlcNAc α 1,4Gal from O-glycans expressed in the gastric gland mucous cell-type mucin has been crystallized. The crystal belongs to space group *P*6₃, with unit-cell parameters *a* = 160.4, *c* = 86.1 Å. Under cryocooled conditions and using a synchrotron X-ray source, the crystals diffract to 1.82 Å resolution. The asymmetric unit contains two or three molecules.

Received 13 October 2003
 Accepted 15 December 2003

1. Introduction

Clostridium perfringens is known to cause a wide variety of diseases in humans and animals (Smith, 1979; Niilo, 1980; Sterne, 1981). A unique endoglycosidase named GlcNAc α 1,4Gal-releasing endo- β -galactosidase (Endo- β -Gal_{GnGa}) isolated from *C. perfringens* was found to specifically release the disaccharide GlcNAc α 1,4Gal from O-glycans expressed in the gastric gland mucous cell-type mucin. This enzyme hydrolyzes the endo- β -galactosyl linkage not only in the GlcNAc α 1,4Gal β 1,4GlcNAc sequence, but also in GlcNAc α 1,4Gal β 1,3GalNAcSer/Thr. The strict specificity for Endo- β -Gal_{GnGa} to release the disaccharide GlcNAc α 1,4Gal distinguishes this enzyme from the other three types of reported endo- β -galactosidases (Ashida *et al.*, 2001). The primary sequence study shows that this enzyme consists of 420 amino-acid residues, including a 17-residue signal peptide at the N-terminus. It was also found to contain the EXDX(X)E sequence (Glu168–Glu173) that has been identified as the catalytic motif of the retaining glycoside hydrolase families 16 and 7. Based on site-directed mutagenesis, Glu168 and Glu173 were found to be essential for Endo- β -Gal_{GnGa} activity (Ashida *et al.*, 2002). The roles of catalytic nucleophile and general acid/base, respectively, are highly likely for these two residues in the catalytic pathway, based on the sequence-homology study of the catalytic motif. The existence of such an unusual endo- β -galactosidase in *C. perfringens* is very intriguing. This unique endo- β -galactosidase should become useful for studying the structure and biological function of glycoconjugates containing the GlcNAc α 1,4Gal epitope.

In this paper, we describe the crystallization and preliminary X-ray diffraction data of Endo- β -Gal_{GnGa}. Further analysis of the crystal structure of this novel enzyme may reveal its functional domain and provide insights into its reaction mechanism.

2. Materials and methods

The pure recombinant enzyme was prepared according to the previously described procedure (Ashida *et al.*, 2002). For crystallization trials, Endo- β -Gal_{GnGa} was concentrated to 16.7 mg ml⁻¹ in 0.1 M ammonium acetate pH 6.0. Crystallization conditions were screened using the sparse-matrix sampling method (Jancarik & Kim, 1991) with reagents from both Hampton Research and Emerald Biostructures. The modified microbatch method (Chayen *et al.*, 1990; D'Arcy *et al.*, 1996) was used by mixing equal volumes (0.5 μ l) of the protein and the screening solutions in the wells of a Nunc HLA plate. The HLA plate was then sealed with 4 ml of a combination of silicon and paraffin oil (5:5). Two hits were found with small crystals, but only one of them gave decent X-ray diffraction. After optimization, the best conditions for crystallization were found to be 42% ammonium sulfate, 2% PEG 400 in 0.1 M Na HEPES pH 7.4. Beautiful crystals (Fig. 1) with maximum dimensions of about 0.3 \times 0.3 \times 0.2 mm were obtained after incubating at 277 K for 2 d.

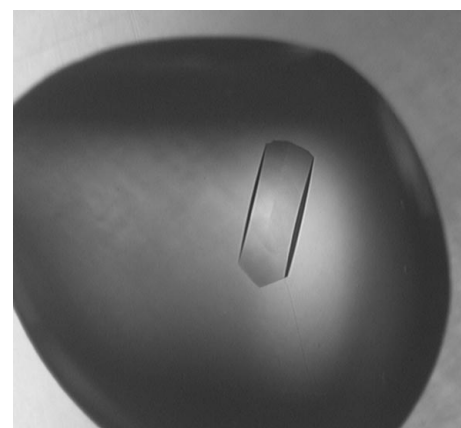


Figure 1
 Native crystal of GlcNAc α 1,4Gal-releasing endo- β -galactosidase from *C. perfringens* with dimensions of 0.15 \times 0.15 \times 0.4 mm.

Table 1
Data-processing statistics.

Values for the outer shell are in parentheses.

Resolution range (Å)	50–1.82 (1.90–1.82)
Unique reflections	107796 (13186)
Completeness (%)	95.8 (94.4)
$I/\sigma(I)$	25.1 (8.1)
$R_{\text{merge}}^{\dagger}$ (%)	5.8 (21.8)

$\dagger R_{\text{merge}} = \sum_{hkl} [\sum_i (|I_{hkl,i} - \langle I_{hkl} \rangle|) / \sum_i I_{hkl,i}]$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with the Miller indices h , k and l and $\langle I_{hkl} \rangle$ is the mean intensity of that reflection.

3. Results and discussion

Similar-shaped Endo- β -Gal_{GnGa} crystals were obtained by growing crystals at both 291 and 277 K under the same crystallization conditions. However, crystals produced at 291 K gave very weak X-ray diffraction, indicating that the enzyme crystal is sensitive to temperature and prefers to grow at low temperature. Fully grown crystals were soaked with artificial mother liquor (42% ammonium sulfate, 2% PEG 400, 0.1 M Na HEPES pH 7.4 and 15% glycerol) for 1 min prior to flash-freezing (Hope, 1988) and data collection. Preliminary diffraction data were

collected at beamline 22-ID in the facilities of the South East Regional Collaborative Access team (SER-CAT) at the Advanced Photon Source, using a MAR Research 165 mm CCD detector and 0.98 Å wavelength X-rays. The crystal diffracted to 1.82 Å resolution. Data processing was carried out using *HKL2000* (Otwinowski & Minor, 1997). The data-processing statistics are given in Table 1. Systematic absences suggest that the crystal belongs to space group $P6_3$, with unit-cell parameters $a = 160.4$, $c = 86.1$ Å. Assuming one molecule per asymmetric unit, the Matthews coefficient (Matthews, 1968) was calculated to be $6.48 \text{ Å}^3 \text{ Da}^{-1}$, indicating there to be two or three molecules per asymmetric unit, corresponding to 60 or 40% solvent content, respectively. Structure determination is currently in progress.

We thank the support from the National Institutes of Health, the University of Georgia Research Foundation and Georgia Research Alliance. Data were collected at Southeast Regional Collaborative Access

Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. W-31-109-Eng-38.

References

- Ashida, H., Anderson, K., Nakayama, J., Maskos, K., Chou, C.-W., Cole, R. B., Li, S.-C. & Li, Y.-T. (2001). *J. Biol. Chem.* **276**, 28226–28232.
- Ashida, H., Maskos, K., Li, S.-C. & Li, Y.-T. (2002). *Biochemistry*, **41**, 2388–2395.
- Chayen, N. E., Shaw Stewart, P. D., Meader, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
- D'Arcy, A., Elmore, C., Stihle, M. & Johnston, J. E. (1996). *J. Cryst. Growth*, **168**, 175–180.
- Hope, H. (1988). *Acta Cryst.* **B44**, 22–26.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Niilo, L. (1980). *Can. Vet. J.* **21**, 141–148.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Smith, L. D. (1979). *Rev. Infect. Dis.* **1**, 254–262.
- Sterne, M. (1981). *Br. Vet. J.* **137**, 443–454.

APPENDIX B

Structure Determination of Fibrillarin from the Hyperthermophilic Archaeon

Pyrococcus furiosus

Structure determination of fibrillarin from the hyperthermophilic archaeon *Pyrococcus furiosus*

Lu Deng,^{a,b,1} Natalia G. Starostina,^{c,1} Zhi-Jie Liu,^{b,c} John P. Rose,^{b,c} Rebecca M. Terns,^c Michael P. Terns,^{c,*} and Bi-Cheng Wang^{b,c}

^a Department of Chemistry, University of Georgia, Athens, GA 30602, USA

^b Southeast Collaboratory for Structural Genomics, Athens, GA 30602, USA

^c Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA

Received 12 January 2004

Abstract

The methyltransferase fibrillarin is the catalytic component of ribonucleoprotein complexes that direct site-specific methylation of precursor ribosomal RNA and are critical for ribosome biogenesis in eukaryotes and archaea. Here we report the crystal structure of a fibrillarin ortholog from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.97 Å resolution. Comparisons of the X-ray structures of fibrillarin orthologs from *Methanococcus jannashii* and *Archaeoglobus fulgidus* reveal nearly identical backbone configurations for the catalytic C-terminal domain with the exception of a unique loop conformation at the *S*-adenosyl-L-methionine (AdoMet) binding pocket in *P. furiosus*. In contrast, the N-terminal domains are divergent which may explain why some forms of fibrillarin apparently homodimerize (*M. jannashii*) while others are monomeric (*P. furiosus* and *A. fulgidus*). Three positively charged amino acids surround the AdoMet-binding site and sequence analysis indicates that this is a conserved feature of both eukaryotic and archaeal fibrillarins. We discuss the possibility that these basic residues of fibrillarin are important for RNA-guided rRNA methylation.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Fibrillarin; snoRNA; Methyltransferase; *Pyrococcus furiosus*; Ribosome biogenesis; X-ray crystallography; Archaea; Methylation

Small nucleolar ribonucleoprotein particles (snoRNPs) predominantly catalyze nucleotide modification of precursor ribosomal RNA (rRNA) [1–3]. Individual snoRNP complexes are composed of a unique modification guide RNA (snoRNA) and several associated proteins [1,2,4]. Numerous snoRNP species interact with different regions of pre-rRNA (and other cellular RNAs) to guide either site-specific ribose methylation (at the 2' hydroxyl group) or uridine isomerization (i.e., conversion to pseudouridine) [1–3,5–7]. The site of target RNA modification is selected by complementary base-pairing between the specific snoRNA and the target RNA molecule [1,2,8,9]. In turn, the multiple post-transcriptional nucleotide modifications catalyzed by the snoRNPs effect the folding and

stability of rRNA and influence the formation and activity of ribosomes [10–13].

Two major classes of snoRNAs were identified based on conserved sequence elements and secondary structure: box C/D (methylation guide) and box H/ACA (pseudouridylation guide) RNAs. Each class of snoRNA is associated with a specific subset of common (or core) proteins: fibrillarin, Nop56, Nop58, and 15.5 kDa (box C/D) [14–18] or Cbf5p, GAR1, Nop10, and Nhp2 (box H/ACA) [19–22]. Fibrillarin and Cbf5p are the enzymatic components of the snoRNPs and catalyze methylation and pseudouridylation, respectively [19,23–25]. The precise roles of the remaining snoRNP proteins are less clear but most individual proteins are required for snoRNA accumulation and all proteins are essential for snoRNP activity [1,2].

Fibrillarin is a key component of the box C/D class of snoRNPs and orthologs of fibrillarin have been found in all eukaryotic organisms examined [15,26,27].

* Corresponding author. Fax: 1-706-542-1752.

E-mail address: mterns@bmb.uga.edu (M.P. Terns).

¹ Authors contributed equally.

All eukaryotic fibrillarlin species have a similar two-domain structure. The amino terminal domain of fibrillarlin is composed of glycine/arginine repeats and this glycine/arginine-rich (GAR) domain has been implicated in nucleolar localization of fibrillarlin [28] and interaction with the survival of motor neuron (SMN) protein [29,30]. The remaining carboxy terminal residues adopt a structure common to *S*-adenosylmethionine (AdoMet)-dependent methyltransferases [31–33]. Until recently, the structural similarities between fibrillarlin and other methyltransferases [33] combined with evidence that certain temperature-sensitive alleles of the yeast fibrillarlin ortholog were defective in rRNA methylation [23] provided highly suggestive evidence that fibrillarlin was the methyltransferase of snoRNPs. More recent analysis with cell-free systems and purified components provided direct evidence that fibrillarlin is the enzyme that catalyzes site-specific ribose methylation of rRNA [24,25,34,35].

Here we present the X-ray structure of fibrillarlin from *Pyrococcus furiosus* (referred to herein as Pf-fib) as the first step in our study of structure and function of box C/D ribonucleoprotein complexes in this organism. *P. furiosus* is a hyperthermophilic euryarchaeote (optimal growth temperature 100 °C) [36]. Comparison with the reported crystal structures of fibrillarlin from two other euryarchaeotes, *Methanococcus jannashii* [33] (Mj-fib) and *Archaeoglobus fulgidus* [37] (Af-fib), shows that while the overall folding patterns are highly similar, significant deviations are observed in the N-terminal domain and the AdoMet binding site. Additionally, Pf-fib (like Af-fib [37]) is a monomer whereas Mj-fib appears to homodimerize [33]. Electrostatic potential distribution analysis revealed the presence of three positively charged amino acids that flank the AdoMet (methyl donor) binding pocket of Pf-fib. Sequence analysis and available structural data indicate that the clustering of the three basic residues at the AdoMet binding site is a common feature of both eukaryotic and archaeal fibrillarlin orthologs suggesting the potential importance for these residues in the interaction of fibrillarlin with snoRNA and/or target pre-rRNA substrate.

Materials and methods

Protein expression and purification. The *P. furiosus* fibrillarlin gene was amplified by PCR from *P. furiosus* genomic DNA, cloned into a modified version of pET21b(+) vector (Novagen) via *Bam*HI and *Nor*I restriction sites (P. Brereton and M. Adams, Department of Biochemistry and Molecular Biology, University of Georgia), and transformed into *Escherichia coli* BL21 (DE3, Invitrogen). The cells were grown in 300 ml LB medium containing 100 µg/ml ampicillin at 37 °C until OD₆₀₀ of 0.7, and expression of the N-terminal, 6× histidine-tagged fibrillarlin protein was induced with 1 mM of isopropylthio-β-D-galactoside (IPTG) for 6 h at room temperature. The cells were pelleted and disrupted by sonication in 30 ml of the binding buffer (20 mM

sodium phosphate, pH 6.5; 500 mM sodium chloride; and 0.1 mM phenylmethylsulfonyl fluoride (PMSF)). The sonicated sample was centrifuged at 20,000 rpm (Beckman, JA-21 rotor) for 30 min at 4 °C. The supernatant was heated at 80 °C for 20 min and centrifuged at 20,000 rpm (Beckman, JA-21 rotor) at 4 °C. Following filtration (0.8 µm pore size, Millex Filter Units, Millipore) and addition of 5 mM imidazole, the supernatant was applied to a 2 ml Ni-NTA Agarose (Qiagen) column equilibrated with binding buffer (pH 6.5) supplemented with 5 mM imidazole. Fibrillarlin was eluted in 10 ml elution buffer (20 mM sodium phosphate, pH 6.5; 500 mM sodium chloride; and 250 mM imidazole).

The procedure was carried out at room temperature. The protein sample was dialyzed against 50 mM sodium phosphate (pH 6.5) containing 200 mM sodium chloride, 0.5 mM DTT, and 0.5 mM EDTA, and concentrated to 10 mg/ml using a Centricon Plus-20 filter device, PL-10 (Millipore). The purity of expressed fibrillarlin was confirmed by SDS-PAGE and by gel filtration (the protein fractionated on a Superdex 200 HiLoad 26/60 column (Pharmacia/Biotech) as a single peak, corresponding to monomeric fibrillarlin).

Crystallization and data collection. Crystallization was achieved by the modified microbatch method under several different conditions. The best conditions were 10% PEG 8000, 200 mM calcium acetate in 100 mM imidazole, pH 8.0, and 12% PEG 6000, 100 mM di-ammonium hydrogen phosphate in 100 mM Tris-hydrochloride, pH 8.5. Crystals appeared after overnight incubation at 18 °C with the largest growing to a size of approximately 0.3 × 0.3 × 0.2 mm. Two data sets were collected at cryogenic temperatures using beamline 17-ID (IMCA-CAT), Advanced Photon Source (APS), Argonne National Laboratory, and a MarResearch 165 mm CCD detector; a data set to

Table 1
X-ray crystallography data statistics

<i>(a) Data collection and processing statistics</i>	
Data set used for phasing	
Wavelength (Å)	1.74
Resolution (Å)	2.20
Completeness (%)	84.8
Bijvoet redundancy	2.8
R_{merge}^a (%)	6.6 (22.4)
Data set used for refinement	
Wavelength (Å)	1.0
Resolution (Å)	1.97
Completeness (%)	82.3
R_{merge}^a (%)	5.2 (37.8)
<i>(b) Refinement statistics</i>	
<i>R</i> value all data	0.228
Free <i>R</i> value	0.272
Free <i>R</i> value test set	1749 reflections
RMS deviations from ideality	
Bond lengths (Å)	0.006
Bond angles	1.3°
Dihedral angles	23.9°
Improper angles	0.79°
Wilson <i>B</i> value (Å ²)	24.2
Mean <i>B</i> value (Å ²)	31.9
Coordinate error ^b (Å)	0.28
Resolution range (Å)	40.06–1.97
Data cutoff	2.0 σ_F
Number of reflections	17,950 reflections
Number of protein atoms	1813
Number of solvent atoms	160

$$^a R_{\text{merge}} = \sum_{hkl} [\sum_i (|I_{hkl,i}| - \langle I_{hkl} \rangle)] / \sum_{hkl,i} \langle I_{hkl,i} \rangle.$$

^b Estimated coordinate error from the Luzzati plot [52].

2.5 Å resolution collected using 1.74 Å X-rays on a potassium iodide soaked crystal for phasing purposes and a data set to 1.97 Å collected using 1.0 Å X-rays on a native crystal for structure refinement. The Pf-fib crystals belong to space group $P2_12_12_1$ with unit-cell dimensions $a = 43.97$ Å, $b = 62.08$ Å, and $c = 103.36$ Å. Based on one molecule per asymmetric unit the solvent content is calculated [38] to be 55%.

Structure determination. Analysis of the phasing data (SHELXD [39]) gave 15 sites which were assigned as iodide and used to estimate the protein phases (ISAS; [45]) to 3.0 Å. The resulting electron density map was used for the preliminary chain tracing and showed most of the secondary structure elements and significant side-chain density. Using this map, the polypeptide chain (residues 4–229, residues 1–3 were not observed and are presumed to be disordered) was easily traced using XTALVIEW [41]. The initial model was refined against the phasing data using CNS 1.0 [42]. Each cycle of refinement was followed by a manual model adjustment coupled with MolProbity [43] analysis. The resulting model was subsequently refined against the high-resolution native data by gradual phase extension to 1.97 Å. The refinement statistics are shown in Table 1. The final model was validated using MolProbity [43] and PROCHECK [44] prior to submission to the Protein Data Bank (accession code 1PRY).

Results and discussion

The structure of *P. furiosus* fibrillarlin (Pf-fib) was determined at 1.97 Å resolution by the method of iterative single-wavelength anomalous scattering (ISAS) [45]. Table 1 gives the data and refinement statistics. The final model of the protein contains 226 residues and 160 water molecules. The crystal structure shown in Fig. 1A consists of 7 α -helices and 12 β -strands that fold into two well-defined domains connected by a short α -helix ($\alpha 0$). The smaller, N-terminal domain is composed of five short β -strands ($\beta 1'$ – $\beta 5'$). The larger, C-terminal catalytic domain is made up of seven β -strands ($\beta 1$ – $\beta 7$) connected by six alternating α -helices ($\alpha 1$ – $\alpha 6$).

The crystal structures of fibrillarlin orthologs from *M. jannaschii* [33] and *A. fulgidus* [37] were previously determined. Af-fib was solved as a subunit in complex with another core snoRNP protein Nop56/58 (a single protein in archaea is related to two eukaryotic snoRNP proteins, Nop56 and Nop58) and the cofactor AdoMet (i.e., the methyl donor in the ribose methylation reaction). All three fibrillarins adopt very similar global folding patterns (Fig. 1B) despite having ~50% amino acid identity (there is a 56% amino acid identity between Pf-fib and Mj-fib, 46% identity between Pf-fib and Af-fib, and 49% identity between Mj-fib and Af-fib). The overall root-mean-squared deviations (r.m.s.d.) of the C α -atomic positions are 1.79 Å between Pf-fib and Mj-fib, 2.82 Å between Pf-fib and Af-fib, and 3.05 Å between Mj-fib and Af-fib. As shown in the structure superimposition of Pf-fib, Mj-fib, and Af-fib (Fig. 1B), the greatest structural differences are located in the N-terminal domain and in the loop linking $\beta 1$ and $\alpha 1$ in the C-terminal domain.

Amino acid residues at the amino termini of fibrillarlin molecules from different organisms exhibit significant

variability. Indeed, the glycine/arginine-rich (GAR) domain characteristic of eukaryotic fibrillarins is absent in archaeal fibrillarins. The sequences of the N-terminal domains of the three structurally characterized archaeal fibrillarins are less similar than their C-terminal (catalytic) domains [37]. In addition, the N-terminal domain of Af-fib is considerably shorter than those of the other two archaeal fibrillarins [37]. Consistent with this variability, the three archaeal fibrillarins exhibit visible folding differences in their N-terminal domains (Fig. 1B). We cannot exclude the possibility that the histidine tag at the N-terminus of Pf-fib (which was not observed in the electron density maps and was assumed to be disordered in crystal) might exert some influence on the structure and account for some of the N-terminal deviations of Pf-fib from the other two proteins, which lacked a histidine tag. However, it appears that the N-terminal domain of fibrillarlin varies significantly both at the primary sequence level and structurally.

The N-terminal domain of Mj-fib was reported to facilitate dimerization of fibrillarlin molecules through specific β -strand interactions [33]. In contrast, the available evidence indicates that both Pf-fib and Af-fib exist as monomers both in solution and in a crystalline state. No fibrillarlin/fibrillarlin interactions were observed in either the Pf-fib or Af-fib [37] X-ray structures. Furthermore, gel filtration chromatography profiles of Pf-fib (data not shown) and Af-fib [37] are consistent with the crystallographic data regarding their monomeric configuration. These results indicate that fibrillarlin homodimerization is not a general property of fibrillarlin molecules in archaea.

The C-terminal domains of the archaeal fibrillarlin proteins exhibit the hallmark topology (a seven-stranded β -sheet flanked by three α -helices on each side), conserved motifs, and invariant residues present in the catalytic domains of all known AdoMet-dependent methyltransferases (MTases) [31,32,46,47]. This topology provides both the AdoMet methyl donor binding pocket and the active site. In all AdoMet-dependent MTases, the AdoMet cofactor interacts predominantly with residues from loops following strands $\beta 1$, $\beta 2$, and $\beta 3$ [32,46,47]. From analysis of the X-ray structure of the Af-fib/Nop56/58 complex bound with AdoMet, the specific residues of fibrillarlin involved in AdoMet interaction have been elucidated [37]. The residues are located in the loops following $\beta 1$ (Af-fib residues Ala67 and Ser68), $\beta 2$ (Glu88, which interacts with the two hydroxyl groups of the ribose moiety of AdoMet), and $\beta 3$ (Ala114), and also within the $\alpha 1$ helix (Thr70) and the loop following $\beta 4$ (Asp133, Ile134, and Gln136) [37]. Furthermore, for fibrillarlin, it has been suggested that one of these residues, Af-fib Asp133 (corresponding to Pf-fib Asp150) located in the loop following $\beta 4$, is the active site base that extracts the proton from the 2' hydroxyl of the target rRNA during ribose methylation [37]. A nearby invariant lysine, Af-fib Lys42

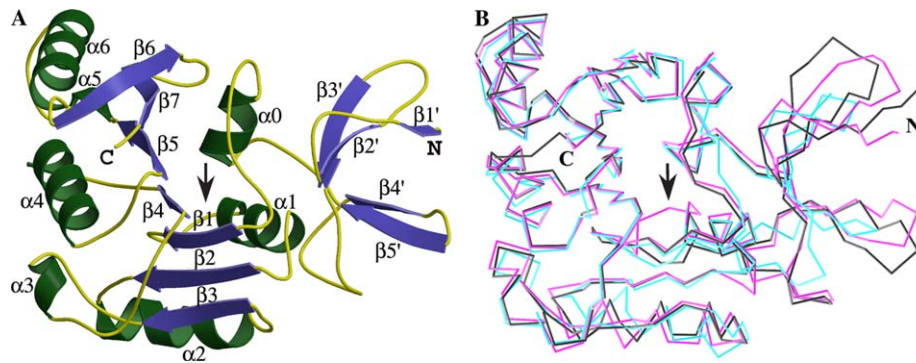


Fig. 1. (A) Ribbon diagram of the fibrillarlin ortholog from *P. furiosus*. The crystal structure consists of 7 α -helices (green) and 12 β -strands (blue) that fold into two well-defined, N- and C-terminal domains and a short connecting helix (α 0). Loop regions are in yellow. B: The superposition of fibrillarlin orthologs from *P. furiosus* (magenta), *M. jannaschii* (black), and *A. fulgidus* (cyan) is shown. The arrow points to a loop (connecting β 1 and α 1) in the *P. furiosus* structure that deviates from the other fibrillarlin structures and is a part of the AdoMet binding site.

		α 0	β 1	α 1		
H. sapiensFib	140	FRSKLAAAILGGV..DQIH	IKPGAKVLYLGAA	SGTTVSHVSDIVGPDGLV		
M. musculusFib	146	FRSKLAAAILGGV..DQIH	IKPGAKVLYLGAA	SGTTVSHVSDIVGPDGLV		
A. thalianaFib	124	FRSKLAAAILGGVD	LDNIWIKPGAKVLYLGAA	SGTTVSHVSDIVGPEGCV		
S. cerevisiaeFib	147	FRSKLAAAGIMGG..LD	ELFTAPGKGVLYLGAA	SGTSVSHVSDIVGPEGVV		
P. abyssiFib	54	HRSKLGAAIMNG..LKN	FPKPGKSVLYLGT	SGTTASHVSDIVGEGKI		
P. horikoshiiFib	54	NRSKLGAAIMNG..LKN	FPKPGKSVLYLGT	SGTTASHVSDIVGEGKI		
P. furiosusFib	54	HRSKLGAAIVNG..LKN	FPKPGKSVLYLGT	SGTTASHVSDIVGEGKI		
M. vannieliiFib	55	NKSKLGAAIING..LKK	MPIKKGTVLYLGAS	AGTTPSHVSDISE..DTIV		
M. voltaeFib	53	NKSKLGASIIING..LKE	MPIKKGKSVLYLGAS	AGTTPSHVADVAE..DSPV		
M. jannaschiiFib	55	NKSKLAAAIKNG..LKV	MPIKGRDSKLYLGAS	AGTTPSHVADTAD..KGIV		
M. thermoautotFib	50	RRSKLAAAILNG..LRG	FSLNDSRVLYLGAS	AGTTASHISDIV..TDGRV		
A. fulgidusFib	39	WRSKLAAMILKCHRLK...	LRGDERVLYLGAA	SGTTVSHLADIVD..EGIT		
		β 2	α 2	β 3	α 3	β 4
H. sapiensFib	188	YAVEFSHRSGRDLINLAKKR	TNIIPVIEDARHPHKYRML	IAMVDVIFADV		
M. musculusFib	194	YAVEFSHRSGRDLINLAKKR	TNIIPVIEDARHPHKYRML	IAMVDVIFADV		
A. thalianaFib	174	YAVEFSHRSGRDLVNMAKKR	TNIIPVIEDARHPAKYRML	VGMDVIFSDV		
S. cerevisiaeFib	195	YAVEFSHRSGRELISMAKKR	TNIIPVIEDARHPQKYRML	LGMDVCFADV		
P. abyssiFib	102	FGIEFSPRVLRRELVP	IVEDRKNIPILGDATKPEEYRALVP	KVDVIFEDV		
P. horikoshiiFib	102	FGIEFSPRVLRRELVP	IVEDRKNIPILGDATKPEEYRALVP	KVDVIFEDV		
P. furiosusFib	102	YGIEFSPRVLRRELVP	IVEDRKNIPILGDATKPEEYRALVP	KVDVIFEDV		
M. vannieliiFib	102	YAVEFAPRIMREFIDSCNER	INLIPILGDANKRPF	FEYSNIVGKVDVIFEDV		
M. voltaeFib	100	YAVEFAPRIMREFIESCEGR	KNLFPILGDANKRPF	EYANIVGKVDVIFEDV		
M. jannaschiiFib	102	YAVEFAPRIMRELIDACAERN	IIPILGDANKRPF	EYANIVGKVDVIFEDV		
M. thermoautotFib	97	YCIIEFSPRMMRELLGVCE	SRKNMPLLEDASRPL	SYLRMVEAADLVYCDV		
A. fulgidusFib	85	YAVEYSAKPFKLLLELVRE	RNNIIPILFDASKFP	WKYSGLVEKVDLIVQDI		
		α 4	β 5	α 5	α 6	
H. sapiensFib	238	AQPDQTRIVALNAHTFLRN	GGHFVISIKANCIDSTASAE	AVFASEVVKMQ		
M. musculusFib	244	AQPDQTRIVALNAHTFLRN	GGHFVISIKANCIDSTASAE	AVFASEVVKMQ		
A. thalianaFib	224	AQPDQARILALNASFFLKT	GGHFVISIKANCIDSTVAE	AVFQSEVVKLQ		
S. cerevisiaeFib	245	AQPDQARILALNSHMF	LKDQGGVVVISIKANCIDSTVDAETVFAREVQKLE			
P. abyssiFib	152	AQPTQAKILIDNAEVLKRG	GYGMIAVKSRSIDVTKEPEQVFRVERELS			
P. horikoshiiFib	152	AQPTQAKILIDNAEVLKRG	GYGMIAVKSRSIDVTKEPEQVFRVERELS			
P. furiosusFib	152	AQPTQAKILIDNAEVLKRG	GYGMIAVKSRSIDVTKEPEQVFRVERELS			
M. vannieliiFib	152	AQPNQAEILVKNKWLKKG	GYGMIAVKSRSIDVTKEPNKEIFEKEILE			
M. voltaeFib	150	AQPNQAEILIKNAKWLKKG	GYGMIAVKSRSIDVTENPRVIFEAEQKEIME			
M. jannaschiiFib	152	AQPNQAEILIKNAKWLKKG	GYGMIAVKSRSIDVTDPKEIFEKEILE			
M. thermoautotFib	147	AQPDQTRIFLENMDCFLKRD	GYGLIMIKARSIDVTRSPKRIFREVGKLE			
A. fulgidusFib	135	AQKNQIEILKANAEFFLKEK	GEVIMVIMIKARSIDSTAEPEEVFKSVLKEM.			
		β 6	β 7			
H. sapiensFib	288	QENMKPQEQLTLEPYERD	HAVVVGVRPPPKVKN~~~			
M. musculusFib	294	QENMKPQEQLTLEPYERD	HAVVVGVRPPPKVKN~~~			
A. thalianaFib	274	QEQFKPAEQVTLPEPFRD	HACVVGGRMPKKQAPAS			
S. cerevisiaeFib	295	EERIKPLEQLTLEPYERD	HCIVVGRYMRSGLKK~~~			
P. abyssiFib	202	EY..FEVIERLNLPEYKDH	ALFVVRKT~~~~~			
P. horikoshiiFib	202	EY..FEVIERLNLPEYKDH	ALFVVRKT~~~~~			
P. furiosusFib	202	EY..FEVIERLNLPEYKDH	ALFVVRKT~~~~~			
M. vannieliiFib	202	EGGFIDVIDEINTEPFEK	DHIMFVGIVWGN~~~~~			
M. voltaeFib	200	QNGFKIVDAINTTEPFEK	DHMLFVGIVWGN~~~~~			
M. jannaschiiFib	202	AGGFKIVDEVIDTEPFEK	DHVMFVGIVWEGK~~~~~			
M. thermoautotFib	197	SSGFRIIDQVGLNPEYK	DHMAVLVKRDV~~~~~			
A. fulgidusFib	184	EGDFKIVKHGSLMPYHRD	H..IFIHAYRF~~~~~			

Fig. 2. Sequence alignment of the C-terminus of fibrillarlin orthologs. Helices are indicated as wavy green lines and β -strands as blue arrows. Strictly conserved residues are shaded in black and semi-conserved in grey. The three strictly conserved positive residues near the AdoMet binding pocket are indicated in turquoise (see also Fig. 3). The residues of the loop following β 1 (where the Pf-fib structure significantly deviates from Mj-fib and Af-fib) are indicated in red. Sequence accession numbers are: *Homo sapiens* (A38712), *Mus musculus* (S38342), *Arabidopsis thaliana* (AAF00542), *Saccharomyces cerevisiae* (S25421), *Pyrococcus abyssi* (NP_125752), *Pyrococcus horikoshii* (NP_142069), *Pyrococcus furiosus* (PDB1PRY; NP_577788), *Methanococcus voltae* (S34646), *Methanococcus vannielii* (S34645), *Methanococcus jannaschii* (PDB1FBN; NP_247681), *Methanococcus thermoautotrophicus* (NP_276343), and *Archaeoglobus fulgidus* (PDB1NT2; NP_070911).

(corresponding to Lys57 in Pf-fib), located in $\alpha 0$ is proposed to stabilize the transition state of the reaction [37]. The residues involved in AdoMet interaction are generally well conserved in the primary sequence of eukaryotic and archaeal fibrillarins (Fig. 2) and generally occupy analogous positions in the three-dimensional structures of the three structurally characterized archaeal fibrillarins (Fig. 1B).

However, we do observe a significant structural deviation in the Pf-fib structure (relative to Af-fib and Mj-fib structures) in one region of the AdoMet binding pocket, the loop following $\beta 1$ (Fig. 1B). Specifically, Pf residues Ser84 and Gly85 of this loop adopt a different conformation than that observed for the corresponding amino acids in the other two fibrillarin species. It is perhaps noteworthy that Pf-Ser84 corresponds to Af-Ser68 which is known to interact with AdoMet [37]. The deviation in the loop structure of Pf-fib may stem from differences in amino acid residues between the three proteins in this immediate region. These differences include the presence of an isoleucine (Ile82; rather than a less hydrophobic Ala) within the loop region and an alanine (Ala88; rather than Val or Pro) in the adjacent $\alpha 1$ helix of the Pf-fib structure (Fig. 2). The aforementioned Ile82 and Ala88 are conserved in all *Thermococci* examined. Analysis of the structures of numerous MTase/AdoMet co-structures reveals that there is considerable flexibility in the structural requirements for many of the contacts that support AdoMet interaction [46–48]. Our results indicate that diversity exists in the AdoMet binding pockets of fibrillarin molecules.

Fibrillarin functions in the context of a ribonucleoprotein complex and the activity of fibrillarin in vivo is dependent upon its association with other components of the complex including snoRNP proteins as well as a box C/D snoRNA [24,25,34,35]. The Af-fib-Nop56/58 co-structure clearly shows that fibrillarin can directly interact with Nop56/58 with 1:1 stoichiometry [37] and it is likely that the heterodimer is incorporated into native snoRNP complexes [34]. We have observed that Pf-fib also heterodimerizes with Pf-Nop56/58 when the two proteins are co-expressed and purified by gel filtration (Starostina, Terns and Terns, unpublished data). The binding of fibrillarin and Nop56/58 is largely dependent on surface (convex/concave) complementary rather than specific amino acid interactions [37].

Although there are three fibrillarin structures available now, including a fibrillarin–Nop56/58 complex structure, important questions remain. For example, it is unclear how fibrillarin interacts with the RNA components of the methyltransferase reaction (i.e., the snoRNA (mediator), rRNA (substrate), and snoRNA/rRNA duplex) and becomes positioned for accurate catalysis. Biochemical data indicate that eukaryotic fibrillarin makes direct contacts with box C/D snoRNAs [49,50]. However, considerable evidence indicates that fibrillarin

is recruited to snoRNP complexes at a late stage of a sequential assembly pathway and does not interact specifically with box C/D snoRNAs in the absence of the other core snoRNP proteins [16,24,34,35]. A yet undescribed molecular mechanism must exist to accurately position the enzyme such that it will selectively methylate the target rRNA nucleotide specified by formation of an rRNA/snoRNA duplex (the modified nucleotide is invariably located in the duplex precisely five nucleotides upstream of the box D (or box D') element of the interacting snoRNA) [40,51].

Analysis of the electrostatic potential distribution on the surface of the Pf-fib molecule (Fig. 3) revealed three positively charged residues neighboring the AdoMet-binding region: Lys57, Arg109, and Lys179. These amino acid residues are strictly conserved in all fibrillarins examined including both archaeal and eukaryotic fibrillarins (Fig. 2 and data not shown). Furthermore, in the Af-fib/Nop56/58 complex structure, the corresponding residues are close to the positively charged region on the surface of the Nop56/58 C-terminal domain that mediates interaction with the box C/D RNA [37]. The positively charged amino acids on the surfaces of fibrillarin and Nop56/58 may cooperate to interact with the negatively charged phosphate backbones of the box C/D RNA and/or the box C/D RNA/rRNA duplex. The highly conserved, basic residues on the surface of fibrillarin may promote fibrillarin/rRNA contacts to facilitate accurate transfer of the methyl group to the target 2'-hydroxyl group of rRNA. Indeed, one of the basic residues that we have noted, Pf-fib Lys57 corresponding to Af-fib Lys42, has been proposed to stabilize the transition state of the methyl transfer reaction [37].

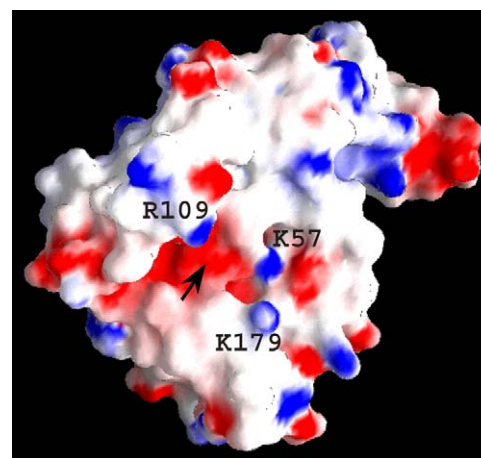


Fig. 3. A view of the solvent-accessible surface of the *Pyrococcus furiosus* fibrillarin ortholog colored by calculated electrostatic potential (negative is in red, positive is in blue, and neutral is in white). Three positively charged residues (Lys57, Arg109, and Lys179) around the AdoMet binding site (indicated with an arrow) are strictly conserved in the fibrillarin protein family.

Functional box C/D RNPs have recently been reconstituted from recombinant proteins, in vitro transcribed box C/D snoRNAs, and model rRNA substrates [24,34,35]. Obtaining high resolution structures of these functional complexes promises valuable insight into the architecture of methylation guide RNPs and the molecular mechanism by which they catalyze site-specific rRNA methylation.

Acknowledgments

This work is supported by grants from the National Science Foundation to M. Terns and R. Terns, and from the National Institute of Health and the University of Georgia Research Foundation and Georgia Research Alliance to B.-C. Wang. 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. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. W-31-109-Eng-38.

References

- [1] M.P. Terns, R.M. Terns, Small nucleolar RNAs: versatile transacting molecules of ancient evolutionary origin, *Gene Expr.* 10 (2002) 17–39.
- [2] W. Filipowicz, V. Pogacic, Biogenesis of small nucleolar ribonucleoproteins, *Curr. Opin. Cell Biol.* 14 (2002) 319–327.
- [3] W.A. Decatur, M.J. Fournier, RNA-guided nucleotide modification of ribosomal and other RNAs, *J. Biol. Chem.* 278 (2003) 695–698.
- [4] L.B. Weinstein, J.A. Steitz, Guided tours: from precursor snoRNA to functional snoRNP, *Curr. Opin. Cell Biol.* 11 (1999) 378–384.
- [5] J.P. Bachellerie, J. Cavaille, A. Huttenhofer, The expanding snoRNA world, *Biochimie* 84 (2002) 775–790.
- [6] P.P. Dennis, A. Omer, T. Lowe, A guided tour: small RNA function in Archaea, *Mol. Microbiol.* 40 (2001) 509–519.
- [7] T. Kiss, Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs, *EMBO J.* 20 (2001) 3617–3622.
- [8] J. Cavaille, J.P. Bachellerie, SnoRNA-guided ribose methylation of rRNA: structural features of the guide RNA duplex influencing the extent of the reaction, *Nucleic Acids Res.* 26 (1998) 1576–1587.
- [9] T. Kiss, Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions, *Cell* 109 (2002) 145–148.
- [10] G. Badis, M. Fromont-Racine, A. Jacquier, A snoRNA that guides the two most conserved pseudouridine modifications within rRNA confers a growth advantage in yeast, *RNA* 9 (2003) 771–779.
- [11] J. Ofengand, M.J. Fournier, The pseudouridine residues of rRNA: Number, location, biosynthesis, and function, in: H. Grosjean, B. Benne (Eds.), *Modification and Editing of RNA*, ASM Press, Washington, DC, 1998, pp. 229–253.
- [12] J.A. Steitz, K.T. Tycowski, Small RNA chaperones for ribosome biogenesis, *Science* 270 (1995) 1626–1627.
- [13] W.A. Decatur, M.J. Fournier, rRNA modifications and ribosome function, *Trends Biochem. Sci.* 27 (2002) 344–351.
- [14] P. Wu, J. Brockenbrough, A. Metcalfe, S. Chen, J. Aris, Nop5p is a small nucleolar ribonucleoprotein component required for pre-18 S rRNA processing in yeast, *J. Biol. Chem.* 273 (1998) 16453–16463.
- [15] T. Schimmang, D. Tollervey, H. Kern, R. Frank, E.C. Hurt, A yeast nucleolar protein related to mammalian fibrillarin is associated with small nucleolar RNA and is essential for viability, *EMBO J.* 8 (1989) 4015–4024.
- [16] N.J. Watkins, V. Segault, B. Charpentier, S. Nottrott, P. Fabrizio, A. Bachi, M. Wilm, M. Rosbash, C. Branlant, R. Luhrmann, A common core RNP structure shared between the small nucleolar box C/D RNPs and the spliceosomal U4 snRNP, *Cell* 103 (2000) 457–466.
- [17] D.L. Lafontaine, D. Tollervey, Synthesis and assembly of the box C+D small nucleolar RNPs, *Mol. Cell. Biol.* 20 (2000) 2650–2659.
- [18] T. Gautier, T. Berges, D. Tollervey, E. Hurt, Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis, *Mol. Cell Biol.* 17 (1997) 7088–7098.
- [19] D.L. Lafontaine, C. Bousquet-Antonelli, Y. Henry, M. Caizergues-Ferrer, D. Tollervey, The box H+ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase, *Genes Dev.* 12 (1998) 527–537.
- [20] N.J. Watkins, A. Gottschalk, G. Neubauer, B. Kastner, P. Fabrizio, M. Mann, R. Luhrmann, Cbf5p, a potential pseudouridine synthase, and Nhp2p, a putative RNA-binding protein, are present together with Gar1p in all H BOX/ACA-motif snoRNPs and constitute a common bipartite structure, *RNA* 4 (1998) 1549–1568.
- [21] A. Henras, Y. Henry, C. Bousquet-Antonelli, J. Noaillac-Depierre, J.P. Gélugne, M. Caizergues-Ferrer, Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs, *EMBO J.* 17 (1998) 7078–7090.
- [22] J.P. Girard, H. Lehtonen, M. Caizergues-Ferrer, F. Amalric, D. Tollervey, B. Lapeyre, GAR1 is an essential small nucleolar RNP protein required for pre-rRNA processing in yeast, *EMBO J.* 11 (1992) 673–682.
- [23] D. Tollervey, H. Lehtonen, R. Jansen, H. Kern, E.C. Hurt, Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly, *Cell* 72 (1993) 443–457.
- [24] A.D. Omer, S. Ziesche, H. Ebhardt, P.P. Dennis, In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex, *Proc. Natl. Acad. Sci. USA* 99 (2002) 5289–5294.
- [25] S. Galardi, A. Fatica, A. Bachi, A. Scaloni, C. Presutti, I. Bozzoni, Purified box C/D snoRNPs are able to reproduce site-specific 2'-O-methylation of target RNA in vitro, *Mol. Cell. Biol.* 22 (2002) 6663–6668.
- [26] B. Lapeyre, P. Mariottini, C. Mathieu, P. Ferrer, F. Amaldi, F. Amalric, M. Caizergues-Ferrer, Molecular cloning of *Xenopus* fibrillarin, a conserved U3 small nuclear ribonucleoprotein recognized by antisera from humans with autoimmune disease, *Mol. Cell. Biol.* 10 (1990) 430–434.
- [27] J.P. Aris, G. Blobel, cDNA cloning and sequencing of human fibrillarin, a conserved nucleolar protein recognized by autoimmune antisera, *Proc. Natl. Acad. Sci. USA* 88 (1991) 931–935.
- [28] S. Snaar, K. Wiesmeijer, A.G. Jochemsen, H.J. Tanke, R.W. Dirks, Mutational analysis of fibrillarin and its mobility in living human cells, *J. Cell Biol.* 151 (2000) 653–662.
- [29] L. Pellizzoni, J. Baccon, B. Charroux, G. Dreyfuss, The survival of motor neurons (SMN) protein interacts with the snoRNP proteins fibrillarin and GAR1, *Curr. Biol.* 11 (2001) 1079–1088.
- [30] K.W. Jones, K. Gorzynski, C.M. Hales, U. Fischer, F. Badbanchi, R.M. Terns, M.P. Terns, Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin, *J. Biol. Chem.* 276 (2001) 38645–38651.

- [31] A. Niewmierzycka, S. Clarke, S-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase, *J. Biol. Chem.* 274 (1999) 814–824.
- [32] X. Cheng, R.J. Roberts, AdoMet-dependent methylation, DNA methyltransferases and base flipping, *Nucleic Acids Res.* 29 (2001) 3784–3795.
- [33] H. Wang, D. Boisvert, K.K. Kim, R. Kim, S.H. Kim, Crystal structure of a fibrillarin homologue from *Methanococcus jannaschii*, a hyperthermophile, at 1.6 Å resolution, *EMBO J.* 19 (2000) 317–323.
- [34] R. Rashid, M. Aittaleb, Q. Chen, K. Spiegel, B. Demeler, H. Li, Functional requirement for symmetric assembly of archaeal box C/D small ribonucleoprotein particles, *J. Mol. Biol.* 333 (2003) 295–306.
- [35] E.J. Tran, X. Zhang, E.S. Maxwell, Efficient RNA 2'-O-methylation requires juxtaposed and symmetrically assembled archaeal box C/D and C'/D' RNPs, *EMBO J.* 22 (2003) 3930–3940.
- [36] G.M. Garrity, J.G. Holt, in: R.W. Castenholz, D.R. Boone, G.M. Garrity (Eds.), *Bergey's Manual of Systematic Bacteriology, The Archaea and Deeply Branching and Phototrophic Bacteria*, Springer-Verlag, New York, 2001, pp. 211–355.
- [37] M. Aittaleb, R. Rashid, Q. Chen, J.R. Palmer, C.J. Daniels, H. Li, Structure and function of archaeal box C/D sRNP core proteins, *Nat. Struct. Biol.* 10 (2003) 256–263.
- [38] B.W. Mathews, Solvent content of protein crystals, *J. Mol. Biol.* 33 (1968) 491–497.
- [39] T.R. Schneider, G.M. Sheldrick, Substructure solution with SHELXD, *Acta Crystallogr. D Biol. Crystallogr.* 58 (2002) 1772–1779.
- [40] J.P. Bachellerie, J. Cavallé, Guiding ribose methylation of rRNA, *Trends Biochem. Sci.* 22 (1997) 257–261.
- [41] D.E. McRee, XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density, *J. Struct. Biol.* 125 (1999) 156–165.
- [42] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, Crystallography & NMR system: a new software suite for macromolecular structure determination, *Acta Crystallogr. D Biol. Crystallogr.* 54 (Pt 5) (1998) 905–921.
- [43] S.C. Lovell, I.W. Davis, W.B. Arendall 3rd, P.I. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by C α geometry: phi, psi and C β deviation, *Proteins* 50 (2003) 437–450.
- [44] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.* 26 (1993) 283–291.
- [45] B.C. Wang, Resolution of phase ambiguity in macromolecular crystallography, *Methods Enzymol.* 115 (1985) 90–112.
- [46] J.L. Martin, F.M. McMillan, SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold, *Curr. Opin. Struct. Biol.* 12 (2002) 783–793.
- [47] H.L. Schubert, R.M. Blumenthal, X. Cheng, Many paths to methyltransfer: a chronicle of convergence, *Trends Biochem. Sci.* 28 (2003) 329–335.
- [48] M.A. Kurowski, J.M. Sasin, M. Feder, J. Debski, J.M. Bujnicki, Characterization of the cofactor-binding site in the SPOUT-fold methyltransferases by computational docking of S-adenosylmethionine to three crystal structures, *BMC Bioinform.* 4 (2003) 9.
- [49] A. Fatica, S. Galardi, F. Altieri, I. Bozzoni, Fibrillarin binds directly and specifically to U16 box C/D snoRNA, *RNA* 6 (2000) 88–95.
- [50] N.M. Cahill, K. Friend, W. Speckmann, Z.H. Li, R.M. Terns, M.P. Terns, J.A. Steitz, Site-specific cross-linking analyses reveal an asymmetric protein distribution for a box C/D snoRNP, *EMBO J.* 21 (2002) 3816–3828.
- [51] Z. Kiss-László, Y. Henry, J.P. Bachellerie, M. Caizergues-Ferrer, T. Kiss, Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs, *Cell* 85 (1996) 1077–1088.
- [52] V. Luzzati, Traitement statistique des erreurs dans la détermination des structures cristallines, *Acta Crystallogr.* 5 (1952) 802–810.

APPENDIX C

Structural Basis for the Emission of Violet Bioluminescence from a W92F Obelin Mutant

Structural basis for the emission of violet bioluminescence from a W92F obelin mutant

Lu Deng^a, Eugene S. Vysotski^{b,c}, Zhi-Jie Liu^c, Svetlana V. Markova^{b,c}, Natalia P. Malikova^b, John Lee^{c,*}, John Rose^c, Bi-Cheng Wang^c

^aDepartment of Chemistry, University of Georgia, Athens, GA 30602, USA

^bPhotobiology Laboratory, Institute of Biophysics, Russian Academy of Sciences, Siberian Branch, Krasnoyarsk 660036, Russia

^cDepartment of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA

Received 20 June 2001; revised 30 August 2001; accepted 10 September 2001

First published online 26 September 2001

Edited by Richard Cogdell

Abstract Mutation of the Trp92 that is known to lie within the active site of the photoprotein obelin from *Obelia longissima*, results in a shift of the bioluminescence color from blue ($\lambda_{\text{max}} = 485$ nm) to violet. The corrected spectrum shows a new band with $\lambda_{\text{max}} = 410$ nm now contributing equally to the one at longer wavelength. The crystal structure of this W92F obelin determined at 1.72 Å resolution shows that there is no significant change in the dimensions of the active site between WT obelin (recombinant Ca^{2+} -regulated photoprotein from *Obelia longissima*) and the mutant. It is proposed that the bioluminescence spectral shift results from removal of a hydrogen bond from the indole of W92 nearby a hydroxyl belonging to the 6-phenyl substituent of the substrate coelenterazine. Propagation of this change through a conjugated bond system in the excited state of the product coelenteramide affects the coupling of the N1-position and the hydrogen-bonded Y138. © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Calcium-regulated photoprotein; X-ray crystallography; Fluorescence; Coelenterazine; Aequorin

1. Introduction

Calcium-regulated photoproteins are found in the marine bioluminescent coelenterates [1], the most well-known being Aequorin, from the bioluminescent jellyfish *Aequorea* [2]. They are named because it only requires the addition of Ca^{2+} to trigger the light reaction, with the emission of bioluminescence of blue to blue-green color. Their bioluminescence spectral distributions are broad with maxima depending on the species in the range 465–495 nm. It had been proposed that the lack of any substrate requirement for the light reaction is because the protein contains an imidazopyrazinone derivative named ‘coelenterazine’ tightly bound within the

protein's active site and as a stabilized peroxide derivative (I). Many chemical and biochemical studies have established the reaction to be an oxidative decarboxylation of this peroxycoelenterazine with the exergonicity deposited into the first electronic state of the product, coelenteramide (Scheme 1).

Fluorescence studies of coelenteramide and analogues have shown that several ionic species can exist, a neutral species (II) with fluorescence spectral maximum around 400 nm, the amide monoanion (III) around 450 nm, a phenolate anion (IV) 480–490 nm, and other anionic species with maxima at longer wavelengths [3]. Small variations are also introduced by solvent dielectric constant and solvent viscosity changes. The general consensus is that the excited state III is the source of the bioluminescence spectrum but in obelin ($\lambda_{\text{max}} = 485$ nm) there is a small band at 400 nm [4] evidently from the excited II [3]. Part of the identifying evidence for III as the primary excited state therefore was in the case of aequorin, a close match between its bioluminescence spectrum ($\lambda_{\text{max}} = 465$ nm) and the fluorescence of the protein-bound product. However for other photoproteins such as obelin, the product fluorescence is not the same and is at longer wavelength than the bioluminescence [4]. It is suggested that the Ca^{2+} -discharged obelin fluorescence is from IV, as it is matched by the fluorescence of coelenteramide in certain basic apolar solvents [3].

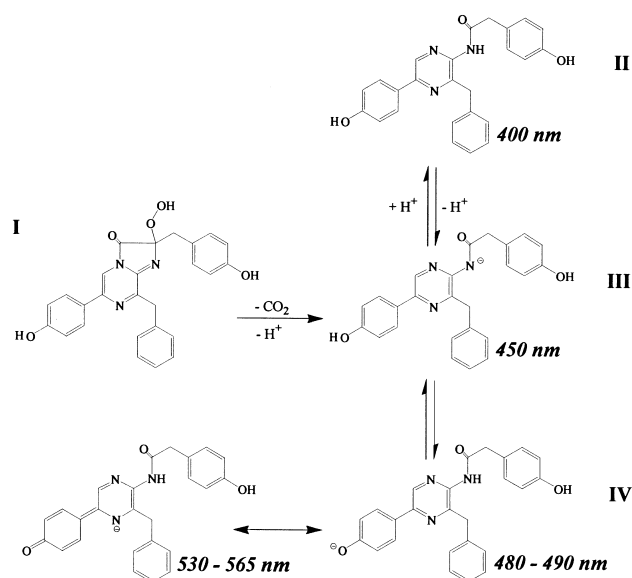
Very recently the three-dimensional structures of two recombinant photoproteins have been determined, aequorin at resolution 2.3 Å [5], and obelin from the hydroid *Obelia longissima*, at resolutions 1.7 [6] and 1.1 Å [7]. Both photoproteins exhibit spatial structure characteristics of the super-family of Ca^{2+} -regulated proteins, four helix–turn–helix motifs, but only three motifs are Ca^{2+} -binding EF-hands. Of considerable importance for mechanistic study is that these structures show the substrate stabilized within the active site. Both protein structures indicate that the coelenterazine is substituted at the C2-position. In the aequorin crystal, weak electron density at this C2-position is consistent with the presence of a peroxy group but there is clearly only one oxygen atom in both obelin structures. It has been suggested that the latter represents the substrate prior to oxygen addition, perhaps bound as a hydrate [7].

Ohmiya et al. [8] produced a number of Trp to Phe mutants of aequorin and found that one of them, W86F, showed a bioluminescence emission with the 400-nm band enhanced, i.e. a contribution to the emission from II. We show here that the equivalent mutation in obelin, W92F, produces an even greater enhancement and that there is no significant change in the

*Corresponding author. Fax: (1)-706-542 1738.

E-mail addresses: ldeng@sunchem.chem.uga.edu (L. Deng), vys@bmb.uga.edu (E.S. Vysotski), zjliu@arches.uga.edu (Z.-J. Liu), sveta@bmb.uga.edu (S.V. Markova), nata.malik@usa.net (N.P. Malikova), jlee@arches.uga.edu (J. Lee), rose@bcl4.bmb.uga.edu (J. Rose), wang@bcl1.bmb.uga.edu (B.-C. Wang).

Abbreviations: WT obelin, recombinant Ca^{2+} -regulated photoprotein from *Obelia longissima*



Scheme 1.

dimensions of the active site between WT obelin and the mutant. The spatial structure now provides a rational basis for explaining the enhancement of the 400-nm bioluminescence. It is proposed that the bioluminescence spectral shift results from removal of the influence of hydrogen bonding from the indole of W92 which lies nearby the hydroxyl belonging to the 6-phenyl substituent of the substrate coelenterazine.

2. Materials and methods

2.1. Molecular biology

Site-directed mutagenesis was done on the template pET19-OL8 *Escherichia coli* expression plasmid carrying the *O. longissima* wild-type apo-obelin (herein called WT obelin) [4]. Mutation resulting in the amino acid change W92F was introduced using the QuickChange site-directed mutagenesis kit from Stratagene (La Jolla, CA, USA) according to the manufacturer's instructions. Two complementary oligonucleotides containing the desired mutations were the sense primer 5'-CAATTCCTCGATGGATTCAACAATTAGCG-3' and antisense primer 5'-CGCCAATTGTTTGAATCCATCGAGGAATG-3'. The changed nucleotides causing the mutations are underlined. Plasmid harboring the mutation was verified by DNA sequence analysis.

For protein production, the *E. coli* BL21-Gold cells with the W92F-obelin plasmid were cultivated with vigorous shaking at 37°C in LB medium containing ampicillin and induced with 1 mM isopropyl β -D-thiogalactopyranoside when the culture reached an OD₆₀₀ of 0.5–0.6. After induction, the cultivation was continued for 3 h.

2.2. Protein purification, crystallization and data collection

W92F obelin was purified as previously reported [4,9]. The purified protein was concentrated to approximately 8–10 mg/ml, desalted on a BioGel P2 column equilibrated with 1 mM EDTA, 10 mM Na/K phosphate buffer, pH 7.3, and concentrated again to approximately the same concentration. Protein was homogeneous according to LC-Electrospray Ionization Mass Spectrometry and the mass was in excellent agreement with that calculated from sequence excluding Met1, as also observed for WT obelin [10]. Protein concentration was determined with the Bio-Rad DC Protein Assay Kit with bovine serum albumin as a protein standard.

W92F obelin was crystallized under conditions similar to WT obelin [7] by the hanging drop vapor diffusion technique at 4°C. The best precipitant was a solution containing 18% polyethylene glycol 8000, 50 mM KH₂PO₄, pH 6.0, with the addition of hexaminecobalt chloride into the drop. Obelin crystals had a size of about 0.5 × 0.25 × 0.25

mm after 10–15 days. Crystals exposed to calcium ion before and after X-ray irradiation, emit light confirming that the crystals consist of an active photoprotein. A complete diffraction data set to 1.72 Å resolution has been collected at 100 K, on a RAXIS IV image plate detector using MSC blue osmic confocal optics focused 5.0 kW Cu K α X-rays generated on a Rigaku RU200 rotating anode. The data were indexed, integrated, and scaled using DENZO/SCALEPACK.

2.3. Refinement

The structure was determined by the Molecular Replacement Method using WT obelin (protein data base entry: 1EL4) as a search model. The space group is P4₁2₁2 with unit-cell dimensions $a = b = 53.45$ Å and $c = 144.49$ Å. The refinement was carried out using CNS1.0 [11]. The final R -value is 23.4%. The free R -factor value is 27.6% using 8% reflection test set. The root-mean-square deviation (r.m.s.d.) in bond length and angles are 0.005 Å and 1.1° respectively. Analysis of the Ramachandran plot (PROCHECK) [12] showed that 94% residues are in most favored regions and 6% residues are in additional allowed regions. No residues are in disallowed or generously allowed regions. Atomic coordinates have been deposited with the Protein Data Bank, accession code 1JF2.

2.4. Bioluminescence assay

The bioluminescence emission was measured with a home-made photometer or a Turner TD-20e luminometer by rapid injection of 10 μ l of the photoprotein solution into the luminometer cell containing 1 ml of 100 mM CaCl₂, 100 mM Tris-HCl, pH 7.0, at room temperature.

2.5. Spectral measurements

Absorption spectra were obtained with a Hewlett-Packard 8453 UV-visible spectrophotometer. Bioluminescence and fluorescence spectra were measured with an SLM-8000 spectrofluorometer. Emission spectra were corrected by reference to the absolute spectral distribution of quinine sulfate, which was also used as the reference for the fluorescence quantum yield. The bioluminescence spectra were measured at 20°C from the obelin in 1 mM EDTA, 10 mM Tris-HCl, pH 7.0, and initiated by injection of CaCl₂. The concentration of free calcium was 0.5 μ M in order to provide a constant light level during the spectral scan. The concentration of calcium was calculated with the MAXICHELATOR program.

3. Results

Fig. 1 shows the bioluminescence emission from W92F obelin mutant together with the fluorescence of the product. The left panel is the uncorrected bioluminescence spectrum to demonstrate that the band at shorter wavelength is much more enhanced than for the W86F aequorin published by Ohmiya et al. [8]. The right panel shows that after correction,

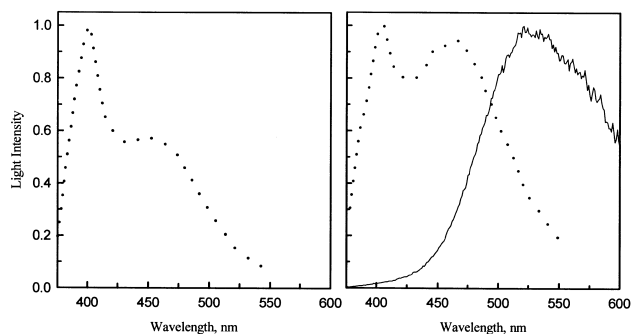


Fig. 1. Bioluminescence and fluorescence spectra. Left panel: Uncorrected bioluminescence spectrum (dotted line) from W92F obelin shows a much enhanced 400-nm contribution over that of aequorin W86F published by Ohmiya et al. [8]. Right panel: Corrected bioluminescence (dotted line) from W92F obelin and fluorescence (solid line) of the Ca²⁺-discharged protein product ($\lambda_{\text{max}} = 510$ nm; $\text{Ex} = 350$ nm).

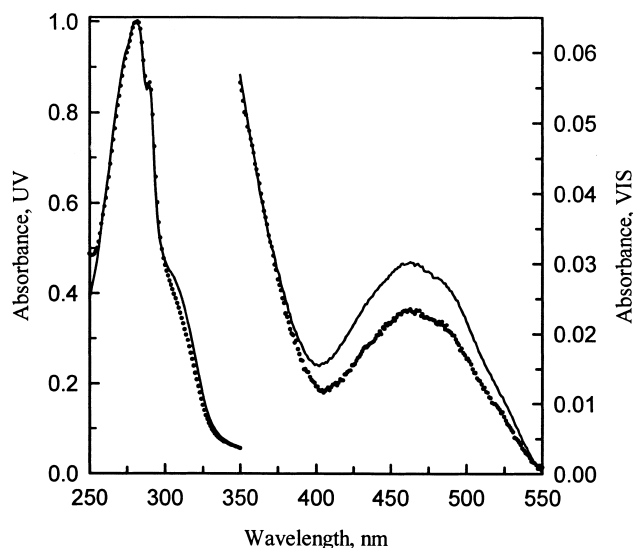


Fig. 2. Absorbance spectra of WT-obelin (dotted line) and W92F (solid line) both normalized at 280 nm.

the two bands are about equally intense, the bioluminescence maxima are at 410 and 470 nm. Probably the longer wavelength band is the same as for WT obelin (485 nm) only distorted by the tail of the shorter wavelength band.

The product fluorescence does not correspond to the bioluminescence emission (right panel) as also found for WT obelin ($\lambda_{\text{max}} = 510$ nm). For aequorin the fluorescence of the Ca^{2+} -discharged corresponds to the bioluminescence but Ohmiya et al. [8] did not report on the case with Ca^{2+} -discharged W86F-aequorin.

Fig. 2 shows that the absorption spectra of W92F mutant and WT obelin have the same maxima, 275 and 460 nm, with the visible absorbance relative to that at 275 nm being lowered in the mutant ($F_{1\text{ cm}}^{0.1\%}$ at 280 nm is 1.99 and 2.5 for the W92F and WT obelin, respectively) since one Trp was replaced for Phe. The Ca^{2+} -discharged proteins also have the same maxima, at 344 and 278 nm in the presence of Ca^{2+} . The uncorrected excitation maximum of the Ca^{2+} -discharged proteins is around 340 nm. The fluorescence emission anisotropy

is 0.33 and constant over the 340-nm excitation band indicating that the fluorophore is rigidly bound to the protein.

The overall structure of the W92F mutant of obelin shows the same sets of helix–turn–helix motifs as WT obelin, with an r.m.s.d. between the $\text{C}\alpha$ -positions of the two proteins of only 0.525 Å. Details of the crystallization and structure solution will be given elsewhere. Of more relevance to the spectral properties is a comparison of the structures of the coelenterazine–oxygen binding site. The electron density map in Fig. 3 shows a relatively weak density around the coelenterazine C2-position, consistent with the substitution by a peroxy group the same as observed for aequorin by Head et al. [5], but in contrast to the single oxygen at this position for the WT obelin.

Fig. 4 is the substrate in the active site and compares the distances to the residues forming the binding cavity between the WT obelin and W92F. Apart from the insertion of a second oxygen between the C2-position and the Y190, the mutation produces no significant change among the other separations including the two nearby water molecules.

It also needs to be noted that the W92F mutant does not have significant differences in the specific activity, in the level of calcium-independent luminescence, in the temperature stability, and in the kinetics of coelenterazine binding (data not shown), in comparison with WT obelin.

4. Discussion

Proteins in this Ca^{2+} -regulated super-family, undergo a conformation change on binding Ca^{2+} and so it can be reasonably supposed that an analogous change is responsible for triggering the chemical reaction leading to the coelenteramide excited state from which the bioluminescence emission occurs. Indeed nuclear magnetic resonance evidence indicates that the obelin apo-protein can assume several conformations including that of obelin itself (with coelenterazine), of the Ca^{2+} -discharged protein (with coelenteramide) in the presence of Ca^{2+} , and of the same after removing Ca^{2+} [13]. The bioluminescence intensity reflects the overall reaction rate and this intensity reaches a maximum on the millisecond time scale, a time sufficient for an adjustment of the protein structure, particu-

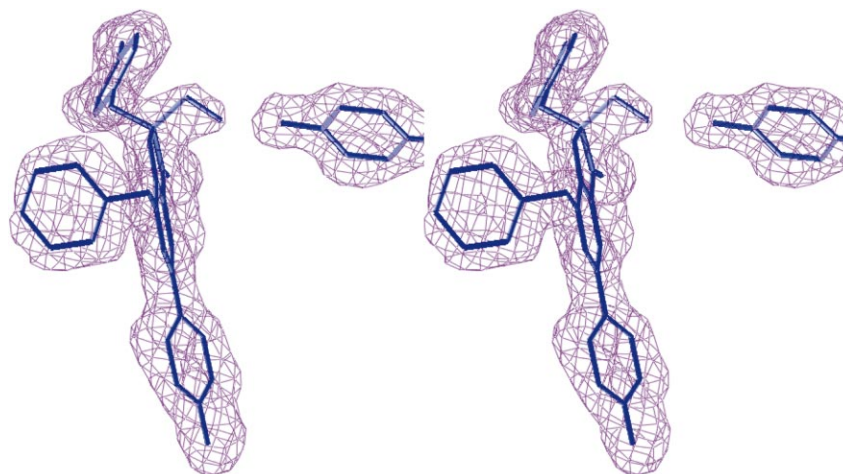


Fig. 3. Stereoview of the electron density map and substrate structure including residue Y190, within the binding cavity of W92F obelin. There is sufficient electron density around the C2-position of coelenterazine to account for a peroxy substitution. The electron density is weaker here than over the rest of the molecule as also observed in aequorin by Head et al. [5].

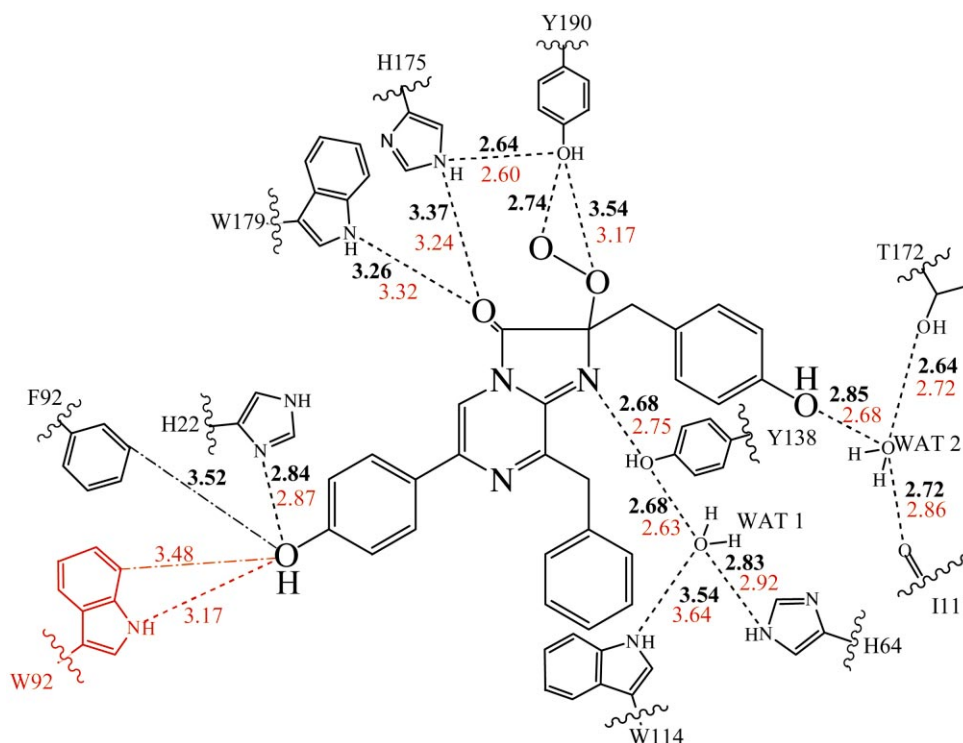


Fig. 4. Two-dimensional picture showing that the W92F mutation produces no significant change in the dimensionality of the 2-peroxycoelenterazine within the photoprotein binding site. Distances are in Å: red, WT-obelin; bold, W92F.

larly if only a minor alteration around the active site is involved.

The problem here is that if we seek an explanation for the difference in the bioluminescence spectra in structural terms, we only have the structures of aequorin, obelin and the W92F obelin mutant, in hand. In the overall structure of W92F obelin which will be presented in more detail elsewhere, there are some slight (0.5) displacements of some residues compared with WT obelin. In the binding cavity however, Fig. 4 shows that there is no significant (± 0.3 Å) dimensional difference between WT obelin and the W92F mutant. The equivalent dimensions in aequorin are also nearly the same as theseobelins. Therefore, how is the strong emission from **II** so favored in the W92F mutant? Besides the appearance of the C2-peroxide substitution in W92F mutant, the controlling factor must be in the presence or absence of the Trp92 hydrogen bond to the *p*-hydroxy group of the 6-phenyl substituent (Fig. 4).

Studies of the chemiluminescence in an aprotic medium as a model, indicate that **III** is the primary excited state produced by the bioluminescence reaction [14,15] (Scheme 1). In the ground state an amide hydrogen is not acidic but in the protein binding site the protonation rate of **III** might not be competitive with the rate of fluorescence transition. But in fact the bioluminescence spectrum from obelin but not aequorin, already shows a small contribution at 400 nm, identified as emission from **II**. The excited amide presumably receives its proton from the proximate hydroxy group of Y138. The W92F change removes the proton coupling via the hydrogen bond from the indole in the vicinity of the hydroxy of the 6-phenyl. Because the conjugation in the coelenteramide extends from this hydroxy to the N1-position, it would not be surprising to find an influence on the proton coupling between

the N1 and Y138 hydroxyl. This proposition does not require any dimensional change although we have no knowledge of how the protein structure might have changed by the time the initially excited coelenteramide is formed. In the aequorin study of a number of Trp mutants, only the equivalent W86F mutant showed a clear bioluminescence spectral contribution from **II** although it is much smaller than in the case of W92F-obelin.

For aequorin the Ca^{2+} -discharged protein has a fluorescence spectrum corresponding to the bioluminescence. This would occur if there was a *pK* shift of the excited coelenteramide within the binding site to favor dissociation of the amide proton. However, for both Ca^{2+} -discharged WT obelin and W92F mutant the final state has a green fluorescence, maximum 510 nm. For obelin therefore, there must be a further structural change going to the final product state. Consequently, a direct spectroscopic examination of the amide anion in Ca^{2+} -discharged obelins is not experimentally accessible but could be feasible for the Ca^{2+} -discharged W86F aequorin.

Further structural and spectroscopic investigation of photoprotein mutants are planned to follow up these ideas.

Acknowledgements: 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.

References

- [1] Morin, J.G. (1974) in: *Coelenterate Biology: Reviews and New Perspectives* (Muscatine, L. and Lenhoff, H.M., Eds.), pp. 397–438, Academic Press, New York.

- [2] Shimomura, O., Johnson, F.H. and Saiga, Y. (1962) *J. Cell Comp. Physiol.* 59, 223–239.
- [3] Shimomura, O. and Teranishi, K. (2000) *Luminescence* 15, 51–58.
- [4] Markova, S.V., Vysotski, E.S. and Lee, J. (2001) in: *Bioluminescence and Chemiluminescence 2000* (Case, J.F., Herring, P.J., Robison, B.H., Haddock, S.H.D., Kricka, L.J. and Stanley, P.E., Eds.), pp. 115–118, World Scientific Publishing Company, Singapore.
- [5] Head, J.F., Inouye, S., Teranishi, K. and Shimomura, O. (2000) *Nature* 405, 372–376.
- [6] Liu, Z.-J., Vysotski, E.S., Chen, C.-J., Rose, J.P., Lee, J. and Wang, B.C. (2000) *Protein Sci.* 9, 2085–2093.
- [7] Vysotski, E.S., Liu, Z.-J., Deng, L., Rose, J., Wang, B.C. and Lee, J. (2001) in: *Bioluminescence and Chemiluminescence 2000* (Case, J.F., Herring, P.J., Robison, B.H., Haddock, S.H.D., Kricka, L.J., and Stanley, P.E., Eds.), pp. 135–138, World Scientific Publishing Company, Singapore.
- [8] Ohmiya, Y., Ohashi, M. and Tsuji, F.I. (1992) *FEBS Lett.* 301, 197–201.
- [9] Illarionov, B.A., Frank, L.A., Illarionova, V.A., Bondar, V.S., Vysotski, E.S. and Blinks, J.R. (2000) *Methods Enzymol.* 305, 223–249.
- [10] Vysotski, E.S., Liu, Z.-J., Rose, J., Wang, B.C. and Lee, J. (1999) *Acta Crystallogr. D. Biol. Crystallogr.* 55, 1965–1966.
- [11] Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gross, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) *Acta Crystallogr. D. Biol. Crystallogr.* 54, 905–921.
- [12] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Crystallogr.* 26, 283–291.
- [13] Lee, J., Glushka, J.N., Markova, S.V. and Vysotski, E.S. (2001) in: *Bioluminescence and Chemiluminescence 2000* (Case, J.F., Herring, P.J., Robison, B.H., Haddock, S.H.D., Kricka, L.J. and Stanley, P.E., Eds.), pp. 99–102, World Scientific Publishing Company, Singapore.
- [14] Hori, K., Wampler, J.E. and Cormier, M.J. (1973) *Chem. Comm.* 492–493.
- [15] Hori, K., Wampler, J.E., Matthews, J.C. and Cormier, M.J. (1973) *Biochemistry* 12, 4463–4468.

APPENDIX D

Preparation and X-ray Crystallographic Analysis of the Ca^{2+} -discharged Photoprotein Obelin

Lu Deng,^a Svetlana V.
Markova,^{b,c} Eugene S.
Vysotski,^{b,c} Zhi-Jie Liu,^b John
Lee,^b John Rose^b and Bi-Cheng
Wang^{b*}

^aDepartment of Chemistry, University of Georgia, Athens, GA 30602, USA, ^bDepartment of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA, and ^cInstitute of Biophysics, Russian Academy of Sciences, Siberian Branch, Krasnoyarsk 660036, Russia

Correspondence e-mail:
wang@bcl1.bmb.uga.edu

Preparation and X-ray crystallographic analysis of the Ca²⁺-discharged photoprotein obelin

Received 13 October 2003

Accepted 10 December 2003

Ca²⁺-regulated photoproteins belong to the EF-hand Ca²⁺-binding protein family. The addition of calcium ions initiates bright blue bioluminescence of the photoproteins, a result of the oxidative breakdown of coelenterazine peroxide to coelenteramide. Crystals of the Ca²⁺-discharged W92F mutant of obelin from *Obelia longissima* have been grown, representing the first crystallization of a photoprotein after the Ca²⁺-triggered bioluminescence. A green fluorescence observed from the crystals clearly demonstrates that coelenteramide, the bioluminescence product of coelenterazine peroxide, is bound within the protein. The diffraction pattern exhibits tetragonal Laue symmetry. Systematic absences indicate that the space group is either *P*₄₃₂₁₂ or *P*₄₁₂₁₂. The unit-cell parameters are *a* = *b* = 53.4, *c* = 144.0 Å. The crystals diffract to 1.9 Å resolution.

1. Introduction

The Ca²⁺-regulated photoproteins consist of a single polypeptide chain (~22 kDa) to which an imidazolopyrazinone derivative (called coelenterazine) is tightly but non-covalently bound (Shimomura & Johnson, 1972). The light-yielding reaction of Ca²⁺-regulated photoproteins is an oxidative decarboxylation of coelenterazine, with the elimination of a mole of carbon dioxide and the generation of the protein-bound product (called coelenteramide) in an excited state (Shimomura & Johnson, 1972).

Two features of the calcium-regulated photoproteins distinguish them from the luciferin–luciferase bioluminescence systems that operate through this same chemical mechanism but have classical enzyme–substrate kinetics. The first is that the bioluminescence kinetics of photoproteins is not influenced by oxygen concentration (Shimomura *et al.*, 1962; Morin & Hastings, 1971). It was suggested therefore that in aequorin the coelenterazine was bound as the stabilized peroxycoelenterazine intermediate (Shimomura & Johnson, 1978) and this proposal has been confirmed in a recent structural study (Head *et al.*, 2000). The second feature is that although calcium is not essential for the luminescence (photoproteins alone give off a very low level of light emission called ‘calcium-independent luminescence’), the light intensity is increased up to 10⁶-fold or more on the addition of calcium. It is supposed therefore that the binding of Ca²⁺ induces conformational changes in the photoprotein which destabilize the peroxycoelenterazine and allow it to decompose to the excited state of the

protein-bound coelenteramide, followed by the emission of the blue bioluminescence.

The three-dimensional structures of obelin from *Obelia longissima* (Liu *et al.*, 2000) and aequorin from *Aequorea* (Head *et al.*, 2000) revealed that the photoprotein molecule is almost entirely helical, highly compact and globular. The overall folds of obelin and aequorin are very similar and are characterized by four sets of helix–loop–helix (HLH) structural motifs wrapping around to form a hydrophobic core cavity that accommodates the coelenterazine peroxide. Similar to other Ca²⁺-binding proteins, the four HLH motifs are arranged in pairs. The loops of HTH motifs I, III and IV form typical calcium ion-binding sites and these motifs are therefore referred as EF-hand motifs. The loop of HTH motif II is not functional for calcium binding because it does not have the canonical sequence for calcium binding. The structures of obelin and aequorin confirmed the prediction made earlier that Ca²⁺-regulated photoproteins belong to the EF-hand Ca²⁺-binding protein family (Charbonneau *et al.*, 1985; Inouye *et al.*, 1985; Illarionov *et al.*, 1995).

Although aequorin and obelin are very similar to each other in primary and spatial structures, there are some obvious differences. Both photoproteins emit blue light, but the light emission of obelin ($\lambda_{\text{max}} = 485 \text{ nm}$) (Markova *et al.*, 2002) is shifted to longer wavelength in comparison with aequorin bioluminescence ($\lambda_{\text{max}} = 465 \text{ nm}$). Also, a shoulder at 400 nm that is clearly revealed in the obelin bioluminescence spectrum is not observed in the aequorin spectrum. Another difference is in the fluorescence of calcium-discharged photoproteins (after biolumines-

cence has been triggered by the addition of Ca^{2+} ions to the solution) on excitation by near-UV. Although both proteins have approximately the same excitation maximum (~ 350 nm), the Ca^{2+} -discharged aequorin displays a blue fluorescence with $\lambda_{\text{max}} = 465$ nm that matches its bioluminescence spectrum, whereas the Ca^{2+} -discharged obelin has a green fluorescence with $\lambda_{\text{max}} = 510$ nm that is shifted from its bioluminescence maximum by 25 nm. It is assumed that the coelenteramide-binding pockets in these Ca^{2+} -discharged photoproteins differ from each other. Recently, we have produced a W92F obelin mutant that displays a bimodal light-emission spectrum with violet and blue emission bands and determined its three-dimensional structure (Deng *et al.*, 2001; Vysotski *et al.*, 2003). In spite of the difference in bioluminescence spectra between this mutant and wild-type obelin, the fluorescence spectrum of Ca^{2+} -discharged W92F obelin is green ($\lambda_{\text{max}} = 510$ nm), similar to that of Ca^{2+} -discharged wild-type obelin, suggesting that the environment of coelenteramide bound in these two Ca^{2+} -discharged proteins is the same.

An HSQC-NMR spectroscopy study of obelin showed that the protein undergoes

clear conformational changes through five states: the first is the apoprotein, the second is when the Ca^{2+} is bound with the apoprotein, the third is on charging the apoprotein with coelenterazine in the absence of Ca^{2+} , the fourth on the addition of Ca^{2+} to cause the bioluminescence reaction and the fifth on removal of calcium ions from the product (Lee *et al.*, 2001). In order to completely elucidate the mechanism of this bioluminescent reaction and the structural transients accompanying each step of the mechanism, the three-dimensional structures of each conformational state need to be solved.

This paper reports the successful production and characterization of crystals of the Ca^{2+} -discharged form of W92F obelin that diffract to 1.96 Å resolution. These crystals represent the first crystallization of a photoprotein after the bioluminescence reaction triggered by calcium ions. The three-dimensional structure of the protein bound with coelenteramide, the product of bioluminescent reaction of Ca^{2+} -regulated photoproteins, will provide more insight into the mechanism of coelenterazine-dependent bioluminescence and the role of the protein moiety in that bioluminescence.

2. Materials and methods

2.1. Protein production and purification

Site-directed mutagenesis was performed on the template pET19-OL8 *Escherichia coli* expression plasmid carrying the *O. longissima* wild-type apo-obelin (Markova *et al.*, 2001). A mutation resulting in the amino-acid change W92F was introduced using the QuickChange site-directed mutagenesis kit from Stratagene (La Jolla, CA) as described elsewhere (Deng *et al.*, 2001; Vysotski *et al.*, 2003).

For protein production, *E. coli* BL21-Gold cells with the W92F obelin plasmid were cultivated with vigorous shaking at 310 K in LB medium containing ampicillin and induced with 1 mM isopropyl β -D-thiogalactopyranoside when the culture reached an OD_{600} of 0.5–0.6. After induction, cultivation was continued for 3 h.

The W92F obelin was purified as previously reported (Deng *et al.*, 2001; Vysotski *et al.*, 2003). The purified protein was concentrated to approximately 8–10 mg ml⁻¹, desalted on a BioGel P2 column equilibrated with 1 mM EDTA, 10 mM sodium/potassium phosphate buffer pH 7.3 and concentrated again to approximately the same concentration. Protein was homogeneous according to LC-Electrospray

Ionization Mass Spectrometry and the mass was in excellent agreement with that calculated from the sequence. The apo W92F obelin was converted to photoprotein with synthetic coelenterazine (Prolume Ltd, Pittsburgh, USA).

To prepare Ca^{2+} -discharged W92F obelin, a solution of W92F obelin was diluted tenfold with 10 mM bis-Tris buffer pH 7.0. CaCl_2 solution in water was then gradually added into the diluted protein solution with stirring at room temperature. The final concentration of calcium was 1 mM. During this procedure, a bright violet bioluminescence was observed. After the bioluminescence emission ceased, the yellow protein solution had turned colorless, indicating that coelenterazine had been converted into coelenteramide. To test for the presence of bound coelenteramide, the final product was excited with near-UV. The Ca^{2+} -discharged W92F obelin displayed green fluorescence. The Ca^{2+} -discharged protein was then concentrated to 28 mg ml⁻¹ with the use of Millipore (Bedford, MA, USA) centrifugal tubes. The protein concentration was measured by the Bradford method with chicken albumin in 1 mM CaCl_2 , 10 mM bis-Tris pH 7.0 as a standard.

2.2. Crystallization

Crystals (Fig. 1a) were grown by the modified microbatch method (Chayen *et al.*, 1990; D'Arcy *et al.*, 1996) using an ORYX 1–6 protein crystallization robot from Douglas Instruments Ltd (East Garston, UK) for screening initial crystallization conditions. Optimization was performed manually. Equal volumes (0.5 μ l) of the protein and the crystallization solutions were mixed in the wells of a Nunc HLA plate. The mixed solutions were covered with 10 μ l of paraffin oil. After all the crystallization droplets had been set up, the HLA plate was sealed using 4 ml of a combination of silicon and paraffin oil (7:3) and was incubated at 277 K for over two weeks. The best conditions for crystallization of Ca^{2+} -discharged W92F obelin were 1.5 M trisodium citrate in 0.1 M Na HEPES buffer pH 7.5.

3. Results and discussion

It took one month for the initial crystals to appear and they were relatively small (maximum dimensions of $\sim 0.05 \times 0.05 \times 0.1$ mm) in a precipitation droplet with a protein concentration of 14 mg ml⁻¹. Success in obtaining crystals was also variable and by the time the crystals appeared

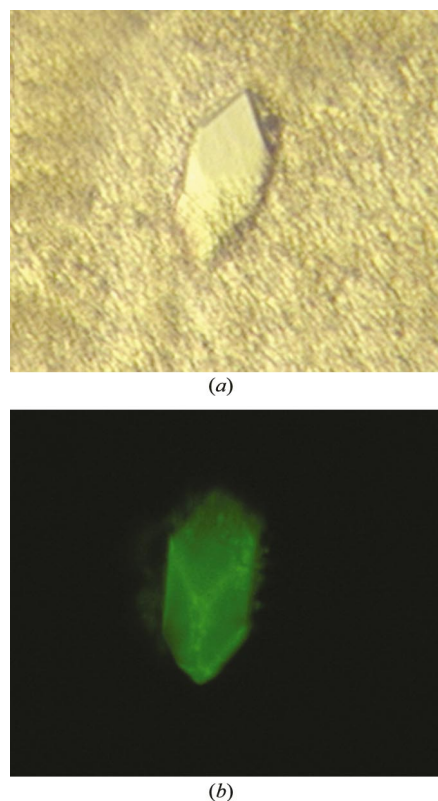


Figure 1
(a) Crystal of Ca^{2+} -discharged W92F obelin ($0.05 \times 0.1 \times 0.25$ mm); (b) fluorescence of the crystal on excitation by near-UV.

Table 1
Data-processing statistics.

Values in parentheses are for the outer shell.

Resolution range (Å)	20–1.96 (2.05–1.96)
Unique reflections	15775 (1982)
Mosaicity	0.4
Completeness (%)	99.5 (99.2)
Redundancy	14.9 (14.1)
$I/\sigma(I)$	9.37 (2.86)
$R_{\text{merge}}^{\dagger}$ (%)	6.0 (18.8)

$\dagger R_{\text{merge}} = \sum_{hkl} (\sum_i (|I_{hkl,i} - \langle I_{hkl} \rangle|)) / \sum_{hkl,i} I_{hkl,i}$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with the Miller indices h , k and l and $\langle I_{hkl} \rangle$ is the mean intensity of that reflection.

the droplet was almost dried out, indicating possible high solubility of the protein. This observation and the length of time it took to form crystals in the incubator partially explains our earlier failure using lower protein concentration (8 mg ml⁻¹). To be sure that coelenteramide is still bound in the protein crystal, the fluorescence of the Ca²⁺-discharged W92F obelin crystals was examined. A single crystal was transferred to a droplet containing only the crystallization mother liquor. The green fluorescence from the crystal observed under a stereomicroscope with excitation at 350 nm (Fig. 1*b*) clearly indicates that coelenteramide is bound within the protein.

For X-ray analysis, the crystal of the Ca²⁺-discharged W92F obelin was directly mounted on a fiber loop (Teng, 1990) containing a minimal amount of mother liquor and flash-frozen (Hope, 1988) in liquid nitrogen before the data-collection process. No cryoprotectant was used prior to the flash-freezing of the crystals. A 360° data

set was collected from the Ca²⁺-discharged W92F obelin crystal on an in-house copper X-ray source with a Bruker Smart 6000 CCD detector. An offset of 8° in 2θ was used for the purpose of extending the data resolution to 1.96 Å. The Bruker program *Proteum* was used for data processing. The data-processing statistics are given in Table 1. The diffraction pattern of the Ca²⁺-discharged W92F obelin exhibits tetragonal Laue symmetry. Systematic absences indicate that its space group is either $P4_32_12$ or $P4_12_12$. The unit-cell parameters are $a = b = 53.4$, $c = 144.0$ Å. Assuming one molecule per asymmetric unit, the Matthews coefficient is 2.34 Å³ Da⁻¹, corresponding to a solvent content of 45%. The determination of the three-dimensional structure of the Ca²⁺-discharged W92F obelin is currently in progress.

We thank Drs Richard B. Meagher and Muthugapatti K. Kandasamy (Department of Genetics, University of Georgia) for assistance with the crystal fluorescence image in Fig. 1. This work was supported by the University of Georgia Research Foundation, the Georgia Research Alliance and grant 02-04-49419 from the Russian Foundation for Basic Research.

References

- Charbonneau, H., Walsh, K. A., McCann, R. O., Prendergast, F. G., Cormier, M. J. & Vanaman, T. C. (1985). *Biochemistry*, **24**, 6762–6771.
- Chayen, N. E., Shaw Stewart, P. D., Meader, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
- D'Arcy, A., Elmore, C., Stihle, M. & Johnston, J. E. (1996). *J. Cryst. Growth*, **168**, 175–180.
- Deng, L., Vysotski, E. S., Liu, Z.-J., Markova, S. V., Malikova, N. P., Lee, J., Rose, J. & Wang, B.-C. (2001). *FEBS Lett.* **506**, 281–285.
- Head, J. F., Inouye, S., Teranishi, K. & Shimomura, O. (2000). *Nature (London)*, **405**, 372–376.
- Hope, H. (1988). *Acta Cryst.* **B44**, 22–26.
- Illarionov, B. A., Bondar, V. S., Illarionova, V. A. & Vysotski, E. S. (1995). *Gene*, **153**, 273–274.
- Inouye, S., Noguchi, M., Sakaki, Y., Takagi, Y., Miyata, T., Iwanaga, S. & Tsuji, F. I. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 3154–3158.
- Lee, J., Glushka, J. N., Markova, S. V. & Vysotski, E. S. (2001). *Bioluminescence and Chemiluminescence 2000*, edited by J. F. Case, P. J. Herring, B. H. Robison, S. H. D. Haddock, L. J. Kricka & P. E. Stanley, pp. 99–102. Singapore: World Scientific Publishing Company.
- Liu, Z.-J., Vysotski, E. S., Chen, C. J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
- Markova, S. V., Vysotski, E. S., Blinks, J. R., Burakova, L. P., Wang, B.-C. & Lee, J. (2002). *Biochemistry*, **41**, 2227–2236.
- Markova, S. V., Vysotski, E. S. & Lee, J. (2001). *Bioluminescence and Chemiluminescence 2000*, edited by J. F. Case, P. J. Herring, B. H. Robison, S. H. D. Haddock, L. J. Kricka & P. E. Stanley, pp. 115–118. Singapore: World Scientific Publishing Company.
- Morin, J. G. & Hastings, J. W. (1971). *J. Cell. Physiol.* **77**, 305–311.
- Shimomura, O. & Johnson, F. H. (1972). *Biochemistry*, **11**, 1602–1608.
- Shimomura, O. & Johnson, F. H. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2611–2615.
- Shimomura, O., Johnson, F. H. & Saiga, Y. (1962). *J. Cell. Comput. Physiol.* **59**, 223–239.
- Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
- Vysotski, E. S., Liu, Z.-J., Markova, S. V., Blinks, J. R., Deng, L., Frank, L. A., Herko, M., Malikova, N. P., Rose, J., Wang, B.-C. & Lee, J. (2003). *Biochemistry*, **42**, 6013–6024.