A Story About Light and Energy:
Mechanistic and Biosynthetic Investigations of Bioluminescence and Methanogenesis

by

Phong Duy Ngo

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 10, 2016

Keywords: conformational dynamics, dinoflagellate luciferase, methyl-coenzyme M reductase, post-translational modifications, reaction mechanisms, tetrapyrrole biosynthesis

Approved by

Steven Mansoorabadi, Chair, Assistant Professor of Chemistry
Holly Ellis, Professor of Chemistry
Eduardus Duin, Associate Professor of Chemistry
Bradley Merner, Assistant Professor of Chemistry

Abstract

This dissertation seeks to investigate underexplored aspects of dinoflagellate bioluminescence and biological methane production.

Ubiquitous in the four oceans, dinoflagellate bioluminescence is a consequence of an oxidation of a luciferin (LH_2) substrate by the luciferase (LCF) enzyme. Utilizing an array of analytical techniques including high-performance liquid chromatography (HPLC), mass spectrometry (MS), and UV-visible/fluorescence spectroscopy, this work initially probes the biosynthesis of dinoflagellate LH_2. Pyropheophorbide *a* is incubated with cell free extracts of *Pyrocystis fusiformis*, the dinoflagellate model organism of this study. Analysis of the reaction mixtures revealed two novel chlorophyll catabolites, P710 and P680, which we propose are involved in the dinoflagellate LH_2 biosynthetic pathway. Additionally, the structure of P630, the previously discovered immediate precursor of dinoflagellate LH_2, was assigned. Computational methods, including constant pH accelerated molecular dynamics (CpHaMD) and time-dependent long-range corrected density functional theory (TDLCDFT), were used to probe the pH regulation of dinoflagellate LCF and the mechanism of the bioluminescence reaction, respectively. CpHaMD simulations carried out on domain III of dinoflagellate LCF demonstrated that at pH 8, the pH at which LCF is maintained in a physiologically inactive state, the conformation of domain III is relative stable, exhibiting little to no large-scale conformational fluctuations. In contrast, an identical simulation carried out on domain III of LCF at pH 6, which is close to the pH optimum of LCF, displayed large collective movements of the proposed

regulatory α-helical bundle of the N-terminal region. We also carried out TDLCDFT calculations on proposed excited state (peroxy anion, hydroperoxide, *gem*-diol, and *gem*-diolate) intermediates of the LH_2-LCF reaction mechanism in an effort to identify the bioluminophore (the light emitting species). Contrary to previously proposed reaction mechanisms, analysis of the first ten low-lying excited states revealed that the most energetically feasible and consistent excited state transition energy resided with the *gem*-diolate intermediate undergoing a biologically novel twisted intermolecular charge transfer mechanism.

Biological methane production is carried out by microorganisms classified as methanogenic archaea. Arguably the most significant reaction in methanogenesis, the terminal step catalyzed by the enzyme methyl-coenzyme M reductase (MCR), uses methyl-coenzyme M and coenzyme B as substrates, coenzyme F430 as a cofactor, and produces methane gas. In an effort to construct active *holo* MCR in a heterologous host, the possibility of *in vivo* biosynthesis of F430 in *Escherichia coli* was investigated. Coexpression of all previously established coenzyme F430 genes and HPLC/MS analysis of the cell extracts indicated that F430 was not synthesized with the coenzyme F430 genes alone, with the pathway ending prematurely at Ni-sirohydorchlorin *a,c*-diamide. Additionally, to elucidate the origin and purpose of several post-translational modifications (PTMs) discovered within the active site region of MCR, a unique cloning/expression strategy was developed and complemented with MS. Genes from the *mcr* gene cluster, *mcrGBDC,* were ligated into a single plasmid with each gene possessing its own promoter and ribosome-binding site. The *mcrA* gene*,* equipped with a C-terminal hexahistidine tag for protein purification purposes, was coexpressed on a separate plasmid with the *mcrGBDC* genes and plasmids containing genes proposed to be involved in performing the PTMs. MS analysis of purified McrA revealed that a protein methylation gene A (*prmA*) homolog was

responsible for the 1-*N*-methylhistidine and *S*-methylcysteine modifications. Additionally, gene knockout experiments demonstrated that methanogenesis marker 10 (*mm10*) is responsible for the 5-(*S*)-methylarginine PTM.

Dedication

Throughout my academic career, many have influenced my life and the path that I have taken but no one has impacted it more than my mom. From primary school and beyond, she has always encouraged me to follow my dreams while expecting the best from me regardless of path. Throughout all the years of my indecisiveness with regard to my academic and career pursuits, she has unwaveringly supported me while others discounted me. She taught me to work hard, to never quit, and to never settle. For this I will be forever appreciative because it was this mentality that has driven me to continue in spite of egregiously difficult times and hard set backs throughout my two decades of formal education. It is because of her that I am who I am today. Therefore, I dedicate this dissertation to my mom, Tina Ngu. Words cannot express how grateful I am to be your son and for all that you have done for me. Thank you mom, and I love you.

Acknowledgments

It would be impossible for me to acknowledge all that were involved in the success of this dissertation, but first and foremost I would like to acknowledge Steven Mansoorabadi for extending to me the opportunity to pursue a Ph.D. here at Auburn University. Though I expected an exponential expansion of my knowledge base being Steve's student but who would have thought that a longhaired post-doc that I met at the University of Texas would have been able to teach me so much about life as well. Occasionally, he may think I have regressed in some respects; nonetheless I want to personally thank him for teaching me the value and virtue of patience, humility, and deliberateness. I would also like to personally thank my friend and cohort Victoria Owens. We joined the program and Steve's lab at the same time and have been together ever since. I cannot recall the numerous times that my two life quotes—'I can only handle one catastrophe at time' and 'if it weren't for the last moment, nothing would get done'— we have experienced together working in this lab for the passed four years but I could not have gotten through them without you. I would be remiss if I did not acknowledge and thank all of the undergraduates that have worked with me over the passed four years. Thus, I would like to express resolute gratitude to Austin Richert, Nicholas Harbin, Kristty Bondurant, Jennifer Royal, Drew Meehan, Stephen Amann, Rachel Park, Kelly Carlquist, Benjamin Mueller, Amanda Holland and Catherine Scull. The assistance that this extraordinary group of undergraduates selflessly provided to the lab and me was instrumental and invaluable to work that is presented in this dissertation. I would also like to acknowledge Ken Zheng for all of his hard work and

entertaining rants, which made our collaborations all the more interesting. I would also like to acknowledge the rest of my labmates for their continued support. I would also like to acknowledge the Biochemistry division for all of the material and intellectual support.

And last, but certainly not the least, I would like to thank the love of my life, Marla Shifflet, for never giving up on me. I know I can be enormously difficult to deal with at times, especially when a grant, paper, or dissertation is due, but I could not have done this without you. You were there when I needed intellectual support, physical support and most significantly when I needed emotional support. For all the late nights you stayed up with me in the lab and for all times you took care of me while was sick, I will forever be grateful. Thank you for being there when I needed you most and thank you for teaching me that there is more to life than work. The meaning that you give my life knows no bounds and I love you for that.

Financial support was provided by the National Science Foundation Career Award (Grant CHE-1555138 and the United States Department of Energy Advanced Research Projects Agency-Energy (Grants DE-AR000428 and DE-AR000433).

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ALA	5-amino levulinic acid
AMBER	Assisted Model Building with Energy Refinement
aMD	Accelerated molecular dynamics
ANME	Anaerobic methanotrophic archea
AOM	Anaerobic oxidation of methane
aq	Aqueous
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
C	Control
Cys	Cysteine
CFB	Coenzyme F430 biosynthesis
CIEEL	Chemically initiated electron-exchange luminescence
cMD	Conventional molecular dynamics
CpHMD	Constant pH molecular dynamics
DAD	Diode array detector
DFT	Density functional theory
DIET	Direct inter-species electron transfer
DOPR	Light independent protochlorophyllide reductase
DTT	Dithiothreitol

ESI	Electrospray ionization
FMN	Flavin mononucleotide
GPU	Graphics processing unit
H	Histidine
HAB	Harmful algal blooms
His	Histidine
HOMO	Highest occupied molecular orbital
HPLC	High performance liquid chromatography
LBP	Luciferin binding protein
LCF	Dinoflagellate luciferase
LH_2	Dinoflagellate luciferin
$LHOO^-$	Peroxy anion intermediate
LHOOH	Hydroperoxide intermediate
LO	Oxyluciferin
$LOHO^-$	*Gem*-diolate intermediate
LOHOH	*Gem*-diol intermediate
LUMO	Lowest unoccupied molecular orbital
m/*z*	Mass to charge ratio
MALDI-TOF MS	Matrix assisted laser desorption ionization – time of flight mass spectrometry
MC	Magnesium chloride
MCR	Methyl-coenzyme M reductase
MD	Molecular dynamics

MM1	Methanogenesis marker 1
MM10	Methanogenesis marker 10
MM13	Methanogenesis marker 13
MPI	Message passing interface
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
NAP(D)H	Nicotinamide adenine (phosphate) dinucleotide
NAPH	Nicotinamide adenine dinucleotide
NPT	Isothermal-isobaric
NTO	Natural bond transition
PaO	Pheophorbide *a* oxygenase
PCM	Polarizable continuum model
PDB	Protein Data Bank
PDC	Pheophorbide demethoxycarbonylase
PPa	Pyropheophorbide *a*
PTM	Post-translational modification
RMSD	Root mean squared deviation
SA	Sodium ascorbate
SAM	*S*-adenosyl-L-methionine
SD	Sodium dithionite
TB	Terrific broth
TDLCDFT	Time-dependent long-range corrected density functional theory

TICT	Twisted intramolecular charge transfer
TLC	Thin layer chromatography
TUV	Tunable ultra violet
UPLC-MS/MS	Ultra-high performance liquid chromatography tandem mass spectrometry
UV	Ultra-violet

PART I

Mechanistic and Biosynthetic Studies of Dinoflagellate Bioluminescence

Introduction

Displays of bioluminescence throughout nature have intrigued scientists for centuries. While this phenomena has been thoroughly investigated in many organisms at all levels of taxonomy, such as bacteria, jellyfish, and fireflies, fundamental aspects of enzyme catalyzed bioluminescence in members of the alveolate class Dinophyceae (i.e., dinoflagellates) still remain a mystery [1].

Dinoflagellates are eukaryotic microorganisms that are ubiquitous throughout the global oceans. During periods of normal population density, dramatic light demonstrations by this phylogenetic group draw much attention, but instances of harmful algal blooms (HABs) are an unwelcome occurrence. This terrifying sight, more commonly known as red tide, is a sudden exponential proliferation of dinoflagellates in coastal waters that is both economically costly and deleterious to human health [2]. Mass mortalities of marine livestock in areas of red tide are the result of specific toxins produced by dinoflagellates. Oxygen depletion of coastal waters also occurs rapidly with some HABs, due to the biochemical decomposition of any deceased red tide organisms [2]. This can cause even more loss of fish and marine animals. Additionally, many groups of dinoflagellates are capable of producing toxins that are harbored in shellfish during feeding. Humans that consume shellfish originating from red tide afflicted areas are susceptible to shellfish poisoning, which can be neurotoxic with accompanying symptoms that can include diarrhea, amnesia, or paralysis [3]. In extreme cases, exposure to saxitoxin (Figure 1), the most

potent of dinoflagellate toxins, induces an array of neurological symptoms that culminates in respiratory distress, cardiovascular shock, and death [3]. Conservative estimates of the cost of red tides and its effect on the environment, seafood and tourism industries, and human health are upwards of $82,000,000 annually [2]. In spite of the economical and public health issues that arise from algal blooms, minimal research has been carried out to investigate ecological and biochemical means of remediating red tides.

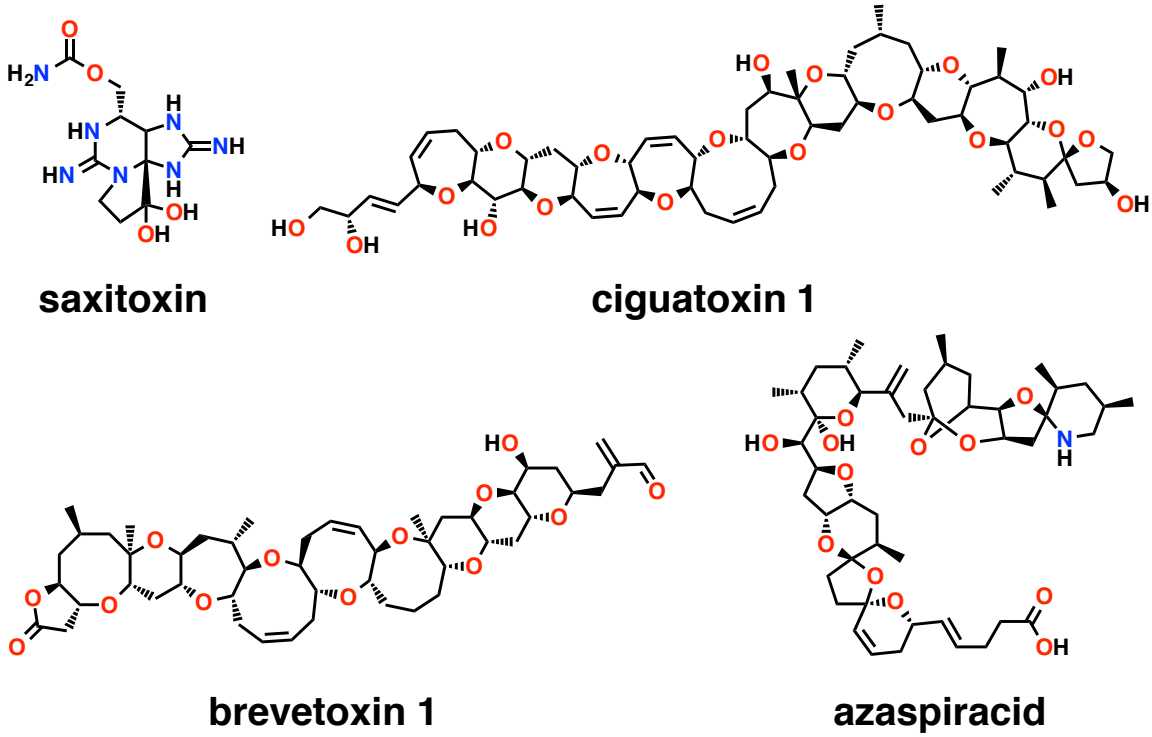


Figure 1. Dinoflagellate toxins.

In contrast to other bioluminescence systems, which can be initiated by quorum sensing, fluorescent proteins, or constitutive expression, bioluminescence exhibited by dinoflagellates is unique in that light emission is activated by physical agitation. At the organismal level, living organisms display light emission for a variety of behavioral reasons. The most well-known of these organismal behaviors is the use of flashing yellow luminosity at 560 nm by fireflies in an

attempt to attract a mate [1, 4, 5]. Other bioluminescent organisms employ light emission in hunting strategies and in camouflaging techniques. For dinoflagellates, bioluminescence is a part of a well evolved, albeit counterintuitive defense mechanism, that biologist have termed the 'burglar alarm hypothesis' [6]. Dinoflagellates are preyed upon by various species of crustaceans and copepods. As these small predators approach their dinoflagellate victims they disturb water around the dinoflagellates triggering bright blue bioluminescence [6]. Bright blue light of wavelength 475 nm travels farthest in water, such that it alerts secondary predators, such as fish or mollusks, to prey on the dinoflagellate predators, thus sparing the dinoflagellates themselves. Dinoflagellates, in effect, save themselves by revealing their location to predators.

At the cellular level, bioluminescence in dinoflagellates involves organelles that are formed by membrane outpocketings of acidic vacuoles known as scintillons [7-9]. All bioluminescent dinoflagellates appear to possess scintillons. Contained within scintillons are the bioluminescence system of dinoflagellates, which in most genera (e.g., *Lingulodinium*, *Alexandrium*, and *Protoceratium*) consists of a luciferin substrate, the luciferin binding protein (LBP), and the luciferase enzyme [10]. Interestingly, not all species of bioluminescent dinoflagellates possess LBP [11, 12]. In particular, the luciferin binding protein does not appear to be conserved in the *Pyrocystis* genus [13]. Additionally, the *Pyrocystis* genus possesses scintillons that are individually much less dense than those of any other genera, which could be related to the lack of a luciferin binding protein [11]. Physical agitation induces an action potential that travels along the vacuole membrane, opening voltage-gated proton channels and allowing protons to flood the interior of the scintillon (Figure 2) [14]. As the pH drops, luciferin is released from the LBP and luciferase is activated, oxidizing luciferin and emitting light in the process [14, 15].

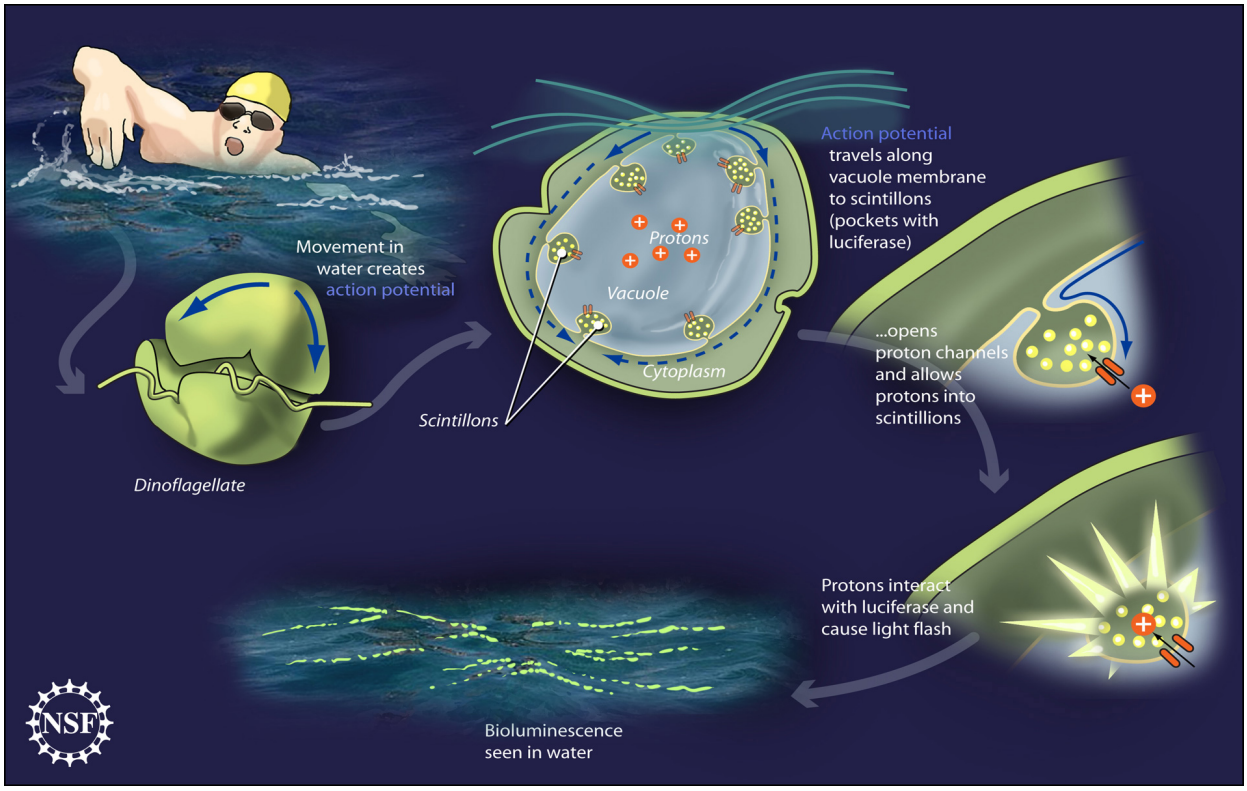


Figure 2. Luminescence exhibited by dinoflagellates (Courtesy: National Science Foundation).

Additionally, many dinoflagellate species are capable of both bioluminescence and photosynthesis. Given that bioluminescence cannot compete with the sun, and photosynthetic light-harvesting proteins are counter productive during bioluminescence, the necessity for regulation of the bioluminescence reaction on a circadian rhythm becomes apparent. Therefore, bioluminescent dinoflagellates have evolved in such a way that light production occurs at night while photosynthesis occurs during the day [16, 17]. As dusk approaches, plastids in the cytoplasm containing photosynthetic pigments and proteins are retracted towards the nucleus, making way for bioluminescence and resolving issues with diminution due to photosynthetic light-absorbing antennae proteins. Sunlight ends bioluminescence and initiates the optimization of photoabsorption from the sun in which plastids are then extended from the nucleus of the cell

through the cytoplasm to maximize the surface area that photosynthetic light-harvesting proteins cover (Figure 3.) [16, 18-20]. For some genera of dinoflagellates, LBP may play a dual regulatory role by binding luciferin during the day and preventing it from being consumed unnecessarily at non-beneficial times [10, 21]. This is supported by evidence that the mRNA level for LBP remains constant throughout the circadian cycle of most bioluminescent dinoflagellates [10, 12, 15, 21, 22].

Not all bioluminescent dinoflagellates are created equal in terms of their luciferin-luciferase system [23]. For instance, members of the red tide causing dinoflagellate genus *Lingulodinium* engage in circadian regulation of bioluminescence that involves diurnal biosynthesis and destruction of luciferase [16, 17, 23]. In contrast, *Pyrocystis,* the model organisms of this dissertation, show no variation in the expression levels of luciferase or the intracellular luciferin concentration throughout each phase of the circadian cycle. In fact, some species have been determined to contain higher intracellular luciferin concentrations during the day [10, 18, 22-24]. Therefore, it is speculated that species such as *Pyrocystis fusiformis*, *Pyrocystis lunula*, and *Pyrocystis noctiluca* are capable of luminescing throughout the circadian cycle, although it may be difficult to observe during the day phase because the bioluminescent machinery is relocated elsewhere in the cell [16]. Interestingly, members of the genus *Pyrocystis* also appear to lack the luciferin binding protein.

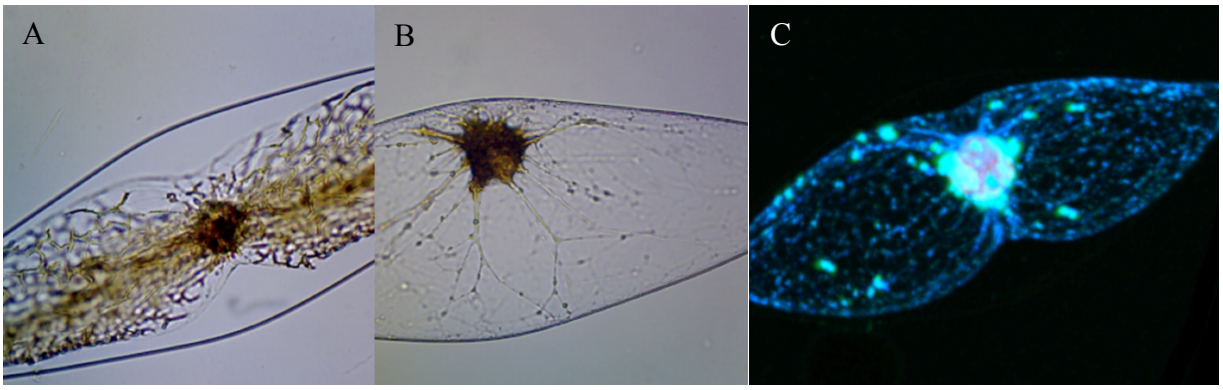


Figure 3. (A) *Pyrocystis fusiformis* observed under a light field microscope during the day phase showing extended photosynthetic machinery throughout the cell. (B) Night phase *P. fusiformis* with photosynthetic machinery withdrawn toward the nucleus of the cell. (C) Confocal laser scanning microscopy image of night phase *P. fusiformis* (exciting at 406 nm, 488 nm, and 640 nm) revealing the location of scintillons in cytoplasm and highly fluorescent photosynthetic pigments localized in close proximity to the nucleus.

At the biochemical level, light emission from living organisms is dependent on a class of enzymes known as luciferases and their cognate substrates known as luciferins. Interestingly, luciferases and luciferins are not structurally conserved throughout bioluminescent phyla [1, 25]. In fact, while divergent evolution has dictated the variability in the function of bioluminescence organisms with the capability, convergent evolution has adapted vastly diverse enzymes and substrates for the biochemical purpose of producing light [1].

Dinoflagellate luciferin is an open-chain tetrapyrrole that bears an resemblance to the photosynthetic pigment chlorophyll *a* [26]. Though the exact mechanism of bioluminescence has yet to be elucidated, the reaction endogenous to dinoflagellates proceeds in a manner that is superficially analogous to those occurring in other luminescent species. Dinoflagellate luciferase catalyzes the oxidation of luciferin by molecular oxygen, producing an excited state intermediate that relaxes with release of electromagnetic radiation in form of blue light with a λ_{max} of 475 nm and yields an oxyluciferin product (Figure 4) [27, 28].

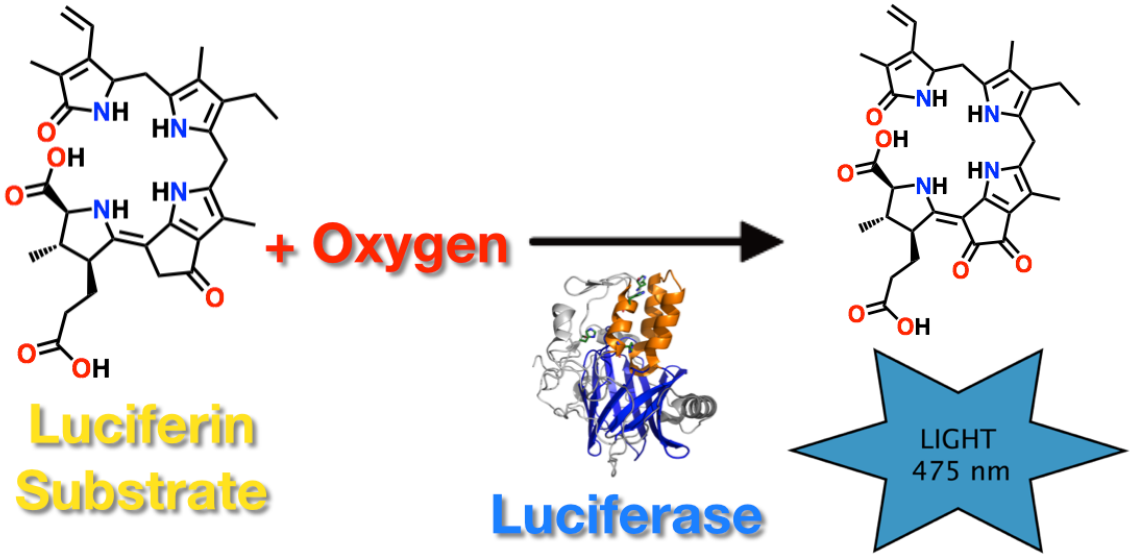


Figure 4. General dinoflagellate luciferase-catalyzed bioluminescence reaction involving the substrates dinoflagellate luciferin and molecular oxygen and the products oxyluciferin and 475 nm light.

While details about the biosynthesis of luciferin continue to elude, the molecule itself is well understood. For instance, dinoflagellate luciferin can be converted to a number of products, depending on the mode of oxidation (Figure 5) [29]. Only oxidation catalyzed by luciferase produces what is termed the enzymatic air-oxidation product (oxyluciferin) that is accompanied by the emission of light. In the absence of enzyme, luciferin spontaneously oxidizes to the non-enzymatic air oxidation product. Curiously, the blue oxidation product can be obtained during the process of purifying luciferin from natural sources or oxidizing luciferin with 0.1% I_2 in ethanol [29]. All other auto-oxidation processes decompose the tetrapyrrole [30].

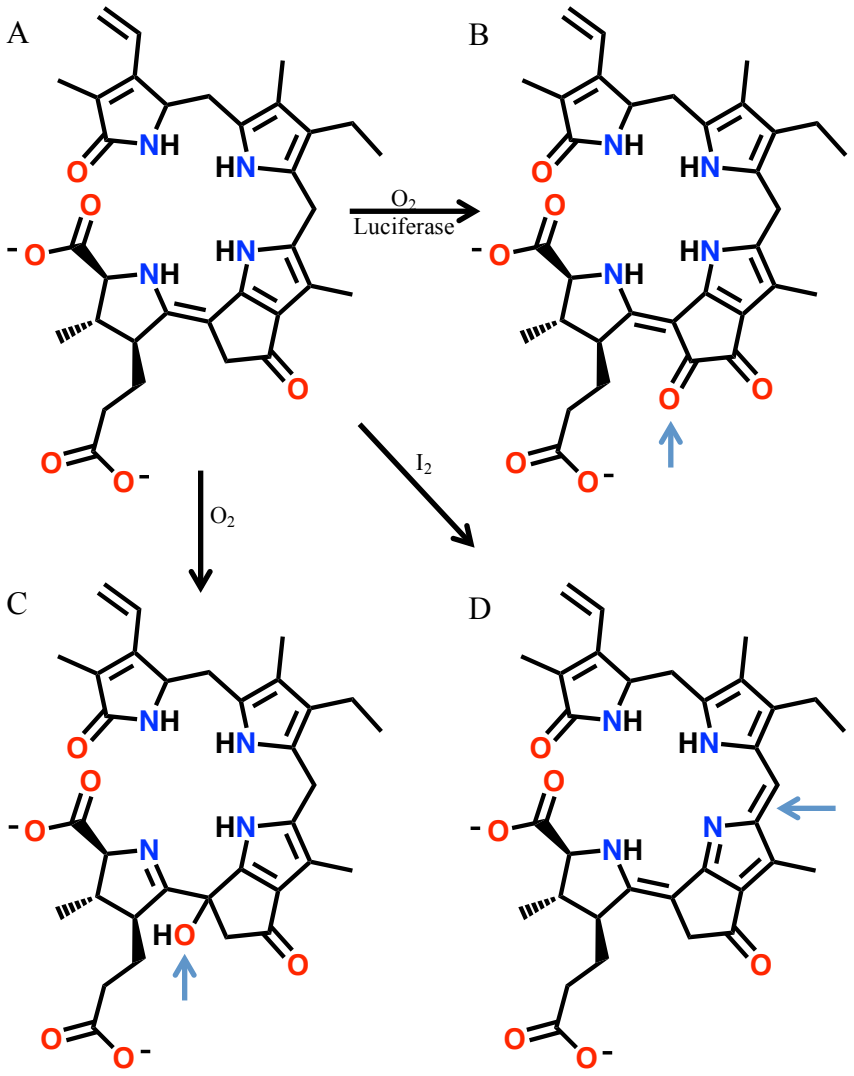


Figure 5. (A) Dinoflagellate luciferin. (B) Enzymatic air oxidation product (oxyluciferin). (C) Non-enzymatic air-oxidation product. (D) Blue oxidation product.

The spectral properties of dinoflagellate luciferin and its oxidation products have also been thoroughly investigated. Luciferin possesses fluorescence properties much like other tetrapyrroles and bioluminescent molecules. The UV-visible absorbance and fluorescence excitation spectra of luciferin are nearly identical, with absorption maxima near 390 nm. Interestingly, the fluorescence emission spectrum of luciferin is very similar to the bioluminescence spectrum, with λ_{max} values of 470 nm and 475 nm, respectively [31]. Despite these spectral similarities, the structure of the light-emitting excited state intermediate of the

luciferase-catalyzed reaction is unknown [1, 31]. Named for its characteristic intense blue color in aqueous solution, the blue oxidation product exhibits an absorption maximum at 630 nm. Interestingly, the blue oxidation product and the enzymatic air-oxidation product vary significantly in their UV-visible absorption spectra. Oxyluciferin produced during the course of bioluminescence features two prominent absorption peaks at 250 nm and 390 nm (Figure 6) [29].

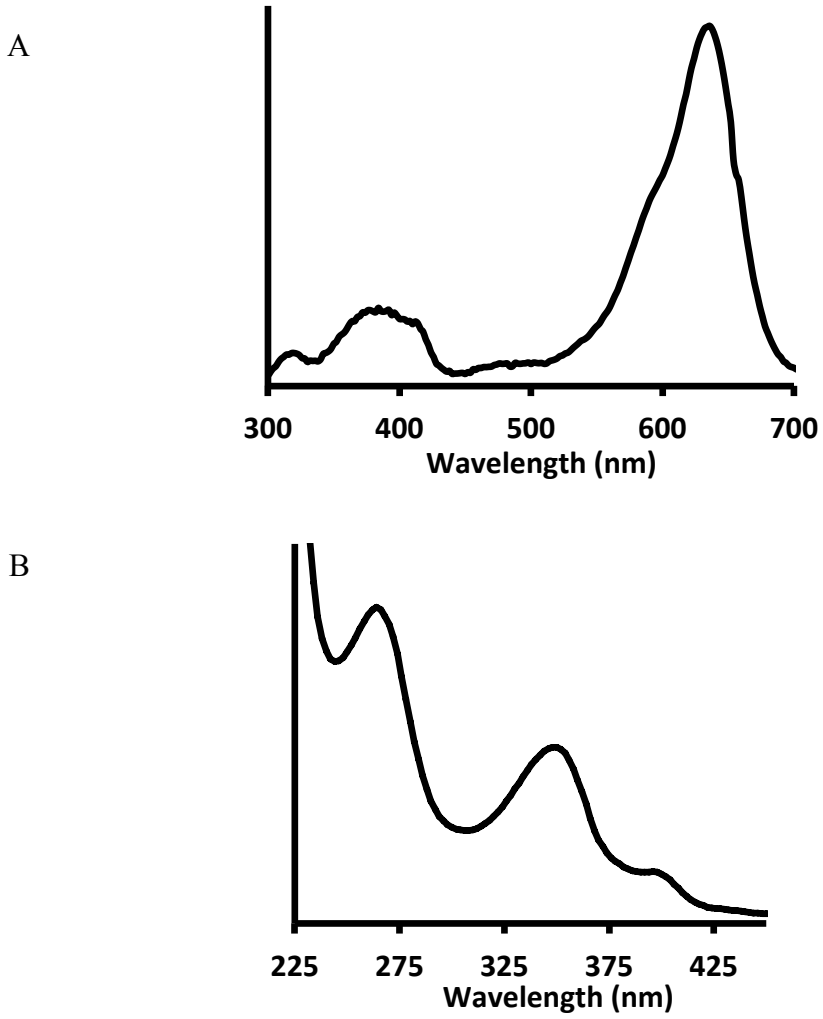


Figure 6. (A) UV-visible absorption spectra of the blue oxidation product (B) and the enzymatic air-oxidation product.

The biosynthesis of luciferin itself is not fully understood but it has been postulated that this substrate is a degradation product of a ubiquitous molecule involved in photosynthesis, chlorophyll *a*, a tetrapyrrole with a magnesium metal center (Figure 7) [32].

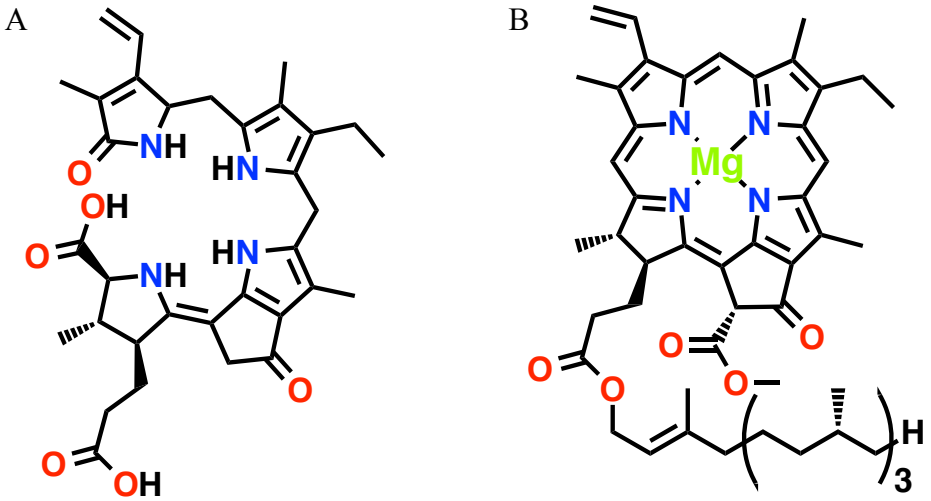


Figure 7. Structures of (A) luciferin and (B) chlorophyll *a*.

The natural chlorophyll degradation pathway in the process of senescence in plants and other species have been heavily studied and show that an enzyme, chlorophyllase, is involved in the initial dephytylation, which is an ester hydrolysis that removes the phytol tail from the chlorin ring [33]. The magnesium is removed from the center of the ring by a magnesium dechelatase enzyme that replaces the magnesium with two protons [34]. At this point, the degraded chlorophyll is known as pheophorbide *a*. Two pathways exist for further degradation of pheophorbide *a*. In one pathway, an additional ester hydrolysis occurs followed by a decarboxylation that removes the methyl ester from the chlorin ring, producing pyropheophorbide *a*. These reactions yield products that are reasonable candidates for the precursors of luciferin. In the other pathway, an oxidation occurs on the ring that facilitates ring opening at the α bridge (as opposed to the δ bridge required for the biosynthesis of luciferin),

generating the red chlorophyll catabolite (Figure 8) [35]. Therefore, an alternative chlorophyll degradation pathway is likely responsible for the biosynthesis of luciferin. In dinoflagellates, chlorophyll degradation may involve an oxygenase with altered regiochemistry for the oxidation of the correct bridging carbon to yield luciferin [32].

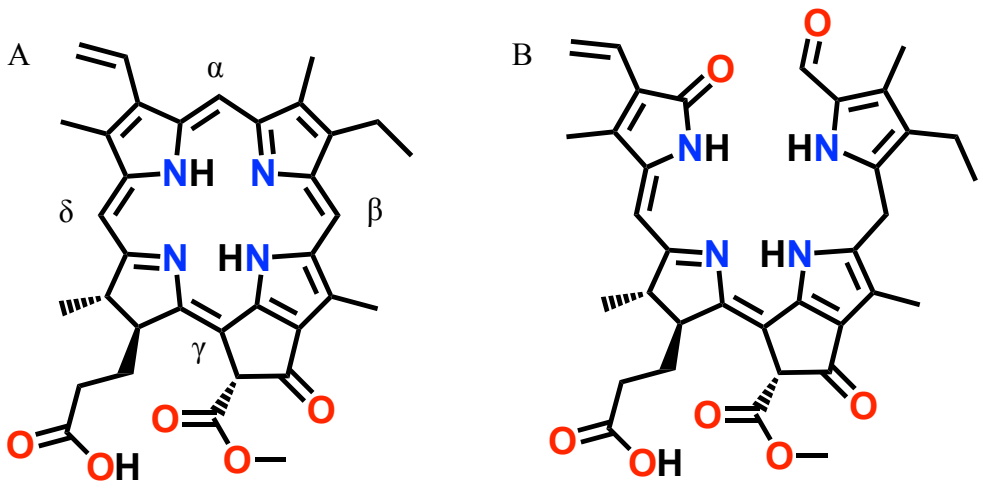


Figure 8. The red chlorophyll catabolite is a degradation product of chlorophyll in which an oxidative cleavage reaction occurred at the α carbon of the chlorin ring. Luciferin is a degradation product of chlorophyll wherein oxidative ring cleavage occurs at the δ carbon.

Luciferases are a class of enzymes that catalyze the oxidation of a luciferin substrate in the presence of oxygen resulting in an oxyluciferin product and bioluminescence. Structurally, dinoflagellate luciferase consists of an N-terminal region and three homologous domains, each containing an active site, on a single polypeptide (Figure 9) [36]. The N-terminal region, the first ~110 amino acids of full-length luciferase, shares sequence homology with the luciferin binding protein (46% for *Lingulodinium polyedrum*), suggesting that these regions share a similar unknown function [36].

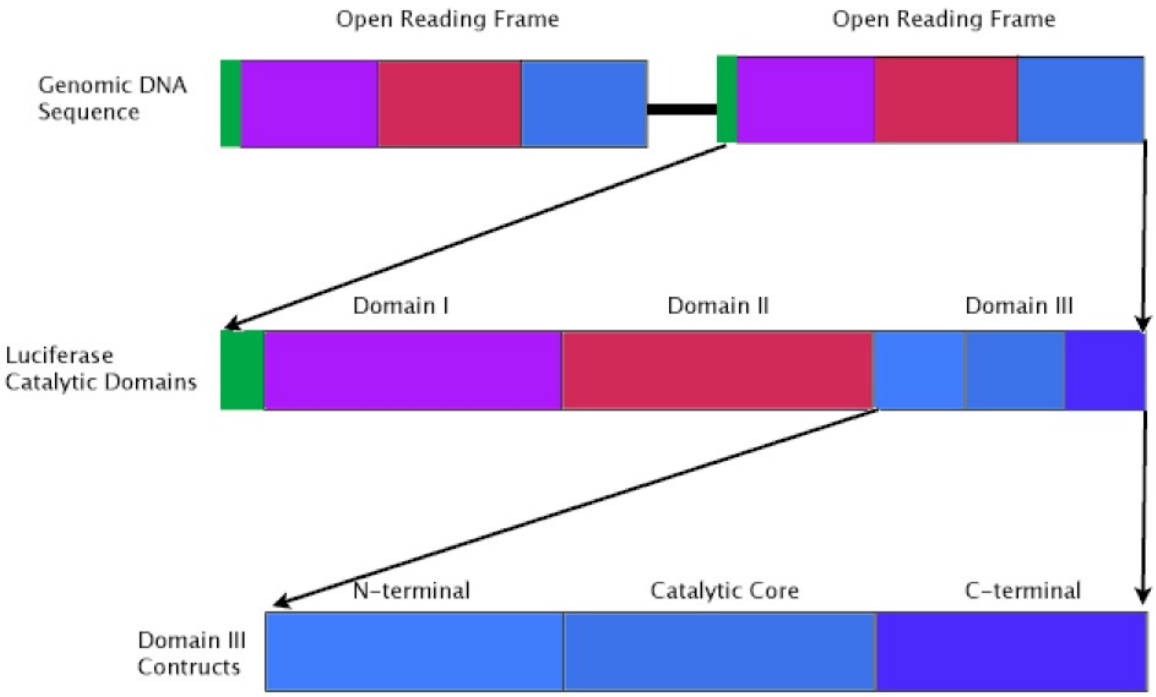


Figure 9. Open reading frame and domains of dinoflagellate luciferase in the genome of a typical dinoflagellate.

The open reading frame of the luciferase gene is composed of 3,723 base pairs that translates into an approximately 137 kDa enzyme [37]. As previously stated, the protein consists of three functional homologous domains that are ~42 kDa each and are all enzymatically active. Approximately 75% of the amino acids in each domain are identical, with catalytic cores and active sites that are ~95% identical and conserved across all three domains [37]. Interestingly, the individual domains are more homologous to the same domain from another species then they are to other domains of the same species [36].

In the genome, several open reading frames of the luciferase gene are arranged in tandem and show no introns [38]. Also, no obvious eukaryotic promoter sequence or TATA box is found within the intergenic space, although a conserved thirteen-nucleotide sequence was identified based on comparison with the peridinin chlorophyll protein gene that could serve as a promoter

[39], [21]. Located in the N-terminus of each domain are four histidines that are conserved. Although these residues are not required for catalysis, it is believed that they play a significant role in the pH regulation of luciferase activity [28, 40]. It is postulated that these particular residues regulate luciferase activity by inducing pH initiated conformational shifts that allow luciferin access to the active site [40]. Additionally, the C-terminus of each domain is thought to be essential for the bioluminescence reaction, although the structural basis for this requirement is not entirely evident from the crystal structure [41].

The three-dimensional structure of domain III from the luciferase of *Lingulodinium polyedrum* has been determined and reveals that anti-parallel β-strands are arranged in a barrel-like formation that makes up the catalytic core and contains the active site. The N-terminal histidine residues proposed to be responsible for pH regulation are a part of a bundle of three α-helices that acts as a lid to the barrel [28]. These histidines, which typically have a pK_a of ~6.0, are proposed to regulate the luciferase enzyme through a conformational change at pH values lower than ~6.5. The lower pH is likely to alter the protonation state of the histidines, and therefore their charge, and may lead to a change in the enzyme structure that controls substrate access to the active site. Specifically, under basic conditions, the lid may be closed preventing substrate access to the active site, while under acidic conditions, the lid may open allowing luciferin access to the active site catalyzing the bioluminescence reaction. The modulation of enzyme activity with pH is usually considered to be due to the variation of the protonation state of residues in the active site that are important for substrate binding or catalysis. The regulation mechanism exhibited by dinoflagellate luciferase may therefore be unique if it is indeed achieved by the protonation of residues outside the active site (Figure 10) [28, 42].

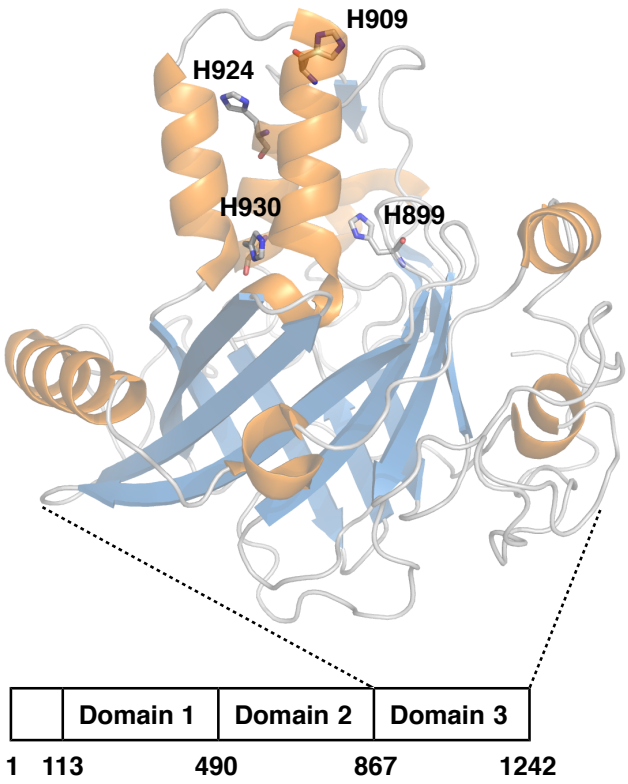


Figure 10. Tertiary structure of domain III of *Lingulodinium polyedrum* showing the helical bundle (the lid) and the catalytic core. α-Helices are colored in orange and β-strands are colored in blue (PDB ID: 1VPR).

The biosynthetic pathway of the luciferin substrate has yet to be determined. With only a few aspects of this pathway elucidated, we employ a strategy that involves incubating known compounds (e.g., chlorophyll *a*, pheoporbide *a*, and pyropheophorbide *a*) with dinoflagellate cell free extract to explore an alternative chlorophyll catabolism pathway that may lead to dinoflagellate luciferin. Proteome arrays are also utilized to identify the enzymes involved. Additionally, the conformational dynamics of the luciferase enzyme is uncharacterized. In this dissertation, we execute a new computational approach in which constant pH simulations are coupled with accelerated molecular dynamics in an effort to observe large conformational changes of the luciferase domains at microsecond-to-millisecond timescales. Moreover, the oxidation mechanism and identity of the light emitting excited state intermediate

(bioluminophore) remains a mystery. Using time-dependent long-range correct density functional theory, we calculate the first ten low-lying excited states of proposed reaction intermediates (the peroxy anion, hydroperoxide, *gem*-diol and *gem*-diolate intermediates) in an effort to identify the most consistent excited state energy and energetically feasible bioluminophore. Thus, this dissertation aims to advance knowledge of these areas of dinoflagellate bioluminescence utilizing an array of multi-disciplinary biochemical and computational approaches.

Chapter 1

The Biosynthetic Pathway of Dinoflagellate Luciferin

Identification and Characterization of Novel Chlorophyll Catabolites from the Bioluminescent Dinoflagellate *Pyrocystis fusiformis*

1.1 Background

Dinoflagellates are freshwater and marine microorganisms that are a causative agent of both red tides and the bioluminescence of the sea [2]. Utilizing dinoflagellate luciferase, dinoflagellate luciferin, and molecular oxygen in an enzymatic redox reaction, dinoflagellates are capable of dramatic displays of bright blue light emission with a λ_{max} at 475 nm [27]. Currently, the metabolic pathway involved in the biosynthesis of dinoflagellate luciferin is poorly understood and efforts to elucidate it have been underrepresented.

In examining the structure of dinoflagellate luciferin, it was observed to be structurally very similar to the ubiquitous photosynthetic pigment chlorophyll *a*. Therefore, it was postulated that the dinoflagellate bioluminescence substrate is a degradation product or catabolite of chloropyll *a* [26, 32]. Consistently, tracer experiments with ^{15}N-labeled glycine and glutamate showed identical labeling patterns for both compounds [26, 32]. However, the genes and corresponding enzymes involved in the breakdown of chlorophyll *a* to luciferin are unknown.

To date, only a single intermediate linking chlorophyll *a* and dinoflagellate luciferin has been purified and spectroscopically characterized from the cell-free extracts of the dinoflagellate *Pyrocystis lunula* [43]. Named P630 for its absorption λ_{max} of 630 nm, the precursor can be reduced to dinoflagellate luciferin by treatment with dithiolthreitol (DTT) or by the catalytic action of a NAD(P)H-dependent reductase [44].

However, neither the identity of the P630 NAD(P)H-dependent reductase nor the structure of P630 itself was reported. Interestingly, the absorption spectrum of the blue oxidation product of dinoflagellate luciferin bears a striking resemblance to the absorption spectrum P630 in all respects [43, 44]. Additionally, is has been reported that both P630 and dinoflagellate luciferin exist as a chromopeptide of approximately 5.4 kDa in size [45]. The nature of the peptide linkage to P630/luciferin and the sequence of the peptide are also unknown (Figure 11).

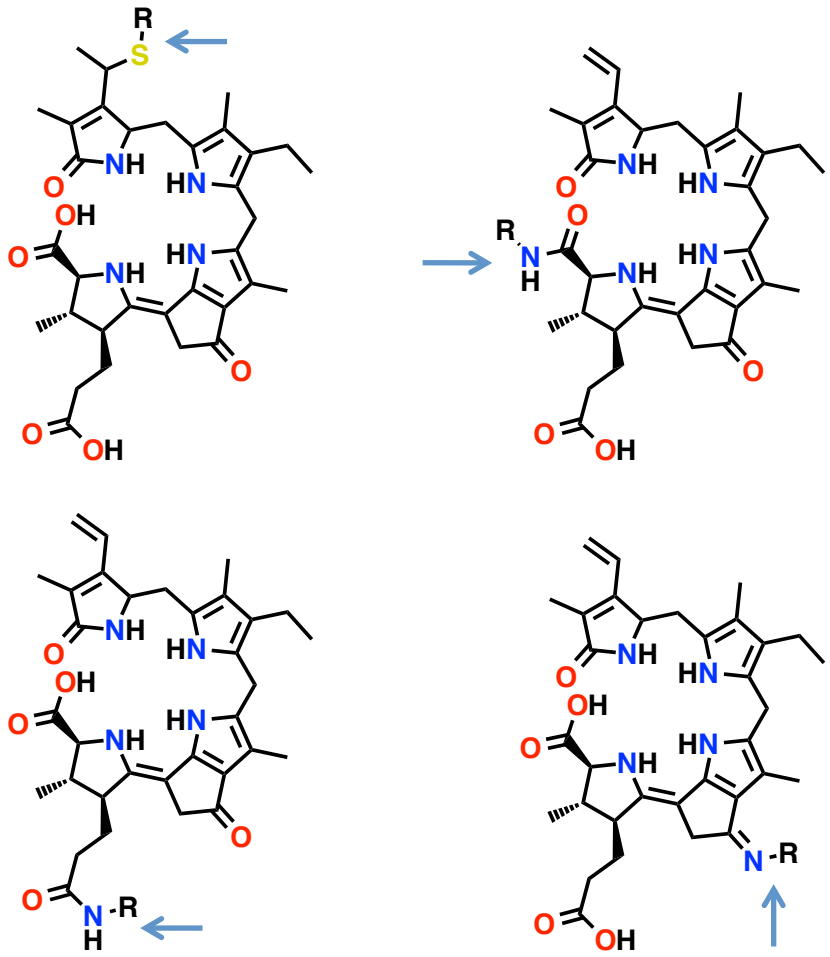


Figure 11. Possible chomopeptide linkages of dinoflagellate luciferin.

With only the penultimate intermediate identified in the biosynthetic pathway of dinoflagellelate luciferin, all the previous steps in its biosynthesis from chlorophyll *a* remain unknown. Chlorophyll degradation has been heavily studied in the senescent leaves and ripening fruits of higher plants, which results in the formation of an open-chain tetrapyrrole similar to luciferin (Figure 12) [34, 46]. Thus, it is likely that the chemistry involved in the aforementioned process is similar to that required to biosynthesize dinoflagellate luciferin [35].

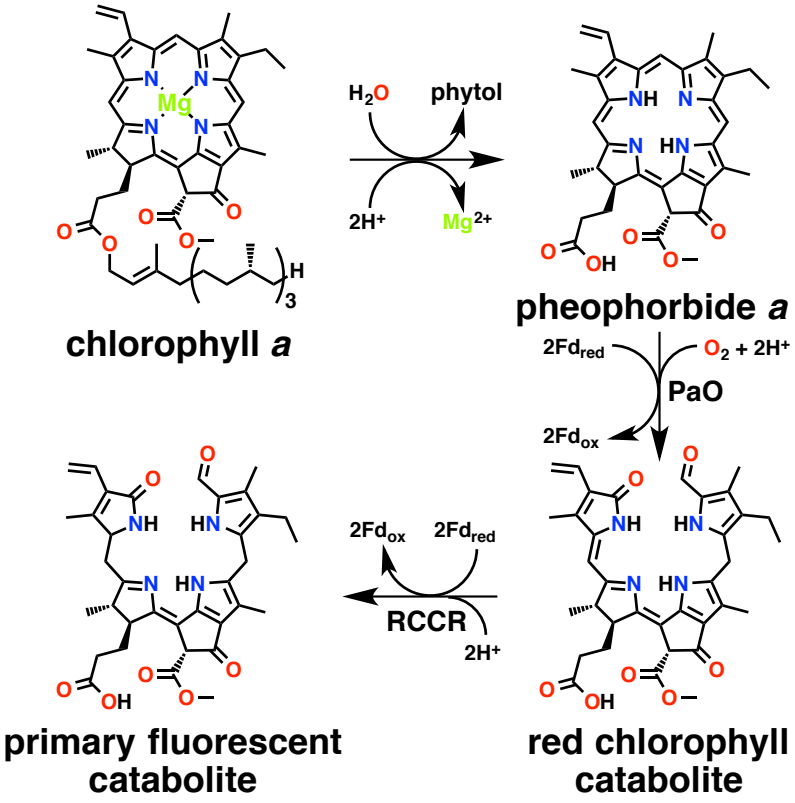


Figure 12. Chlorophyll *a* catabolism in higher plants.

The chlorophyll *a* degradation pathway begins with the removal of the phytyl chain by way of an ester hydrolysis with the enzyme chlorophyllase producing chlorophyllide *a* [33]. A metal dechelatase is then employed to dechelate the central magnesium ion from the chlorophyllide creating pheophorbide *a* [47]. The next step involves an oxidative cleavage of

pheophorbide *a* that opens the tetrapyrrole ring at the α bridge creating the red chlorophyll catabolite. This reaction is catalyzed by a pheophorbide *a* oxgenase (PaO) in higher plants. Chlorophyll *a* catabolism ends with a reduction of the red chlorophyll catabolite by the red chlorophyll catabolite reductase to yield the primary fluorescent catabolite [34]. An additional pathway also exists in which a second enzyme-catalyzed ester hydrolysis is carried out along with a decarboxylation to remove the methyl ester present on the exocyclic ring of chlorophyll *a*. In higher plants, a pheophorbidase enzyme is present to catalyze the ester hydrolysis, which produces pyropheophorbide *a* after spontaneous decarboxylation. In contrast, in algae, pyropheophorbide *a* is created directly from pheophorbide *a* through enzymatic catalysis with a demethoxycarbonylase (Figure 13) [48]. Since dinoflagellate luciferin also lacks the methyl ester functional group, it is possible that a similar enzyme is involved in the biosynthesis of luciferin and that pyropheophorbide *a* is an intermediate in the pathway. Subsequent oxidative cleavage of the δ bridge of pyropheophorbide *a* (in a manner analogous to the formation of the red chlorophyll catabolite) and attachment of the short peptide may lead to the aforementioned P630 precursor. Finally, the previously mentioned NAD(P)H-dependent reductase would catalyze the reduction of P630 to luciferin in a step analogous to the formation of the primary fluorescent catabolite of chlorophyll *a*.

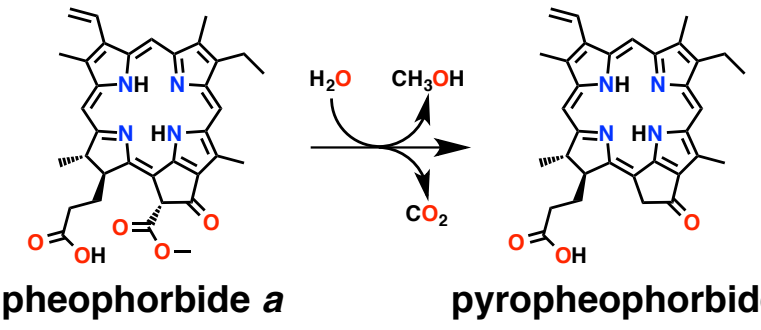


Figure 13. Ester hydrolysis and decarboxylation catalyzed by pheophorbidase (Phedase) in plants and pheophorbide demethoxycarbonylase (PDC) in algae to form pyropheophorbide *a*.

Postulating that the initial steps of luciferin biosynthesis are identical to that of the chlorophyll *a* degradation pathway and based on the previously discovered P630 intermediate and its reductase, this study aims to expand our understanding of the biosynthesis of dinoflagellate luciferin.

1.2 Methods

Chemicals

Chlorophyll *a*, pheophorbide *a*, and pyropheophorbide *a* were all purchased from Frontier Scientific, Inc. (formerly Porphyrin Products, Logan, UT). Unless otherwise noted, all other compounds and chemicals were purchased from Sigma Aldrich (St. Louis, MO).

Cultivation and Harvesting

Cultures of the dinoflagellate species *Pyrocystis fusiformis* were purchased from Sunnyside Sea Farms of (Santa Barbara, CA), while cultures of the species *Pyrocystis lunula* were purchased from the Experimentelle Phykologie und Sammlung von Algenkulturen der Universität Göttingen (EPSAG) in Germany. The dinoflagellates were then sub-cultured into seawater that was supplemented with F/2 Guillard's marine enrichment medium with the exclusion of silicates, both of which were purchased from the National Center for Marine Algae and Microbiota (East Boothbay, ME). The cultures were then grown on a reverse circadian cycle under 12 hours of fluorescent light during the day phase and then 12 hours of darkness for the night phase.

The dinoflagellates were harvested via Buchner funnel filtration. To obtain cell-free protein extracts, *Pyrocystis fusiformis* was harvested while in the night phase. In contrast, for the

extraction of tetrapyrroles, *Pyrocystis lunula* was used and harvested in the day phase. Generally, from starter culture to the point at which cell density is at a sufficient level for extraction, a cultivation period of approximately 1-2 months is recommended. This yields approximately 1.0 g of wet cell paste for every liter harvested.

Extraction of Proteins and Tetrapyrroles

Dinoflagellates used as cell-free protein extracts were immediately transferred from the filter paper to an ice cold 0.1 M sodium phosphate buffer, pH 6.8, supplemented with 20% glycerol and Amresco's plant protease inhibitor cocktail. The cell suspension was allowed to incubate on ice for 10 minutes. The suspension was then transferred to a Dounce homogenizer (12 mL Pyrex glass in glass) that had been chilled on ice and the dinoflagellates were immediately ground with the glass pestle labeled 'A' (loose pestle) three times while maintaining ice cold temperatures. The 'B' glass pestle (tight pestle) was then used three times on the slurry while maintaining ice-cold temperatures at all times. The dinoflagellate cell lysate was immediately centrifuged at 4 °C with a speed of 20,000 × g in a Beckman Optima XE-90 Ultra-centrifuge with Type 45 Ti rotor for a period of one hour to pellet the cell debris and insoluble fraction. The supernatant was then decanted and distributed into working aliquots and frozen with liquid nitrogen and stored at −80 °C for later use. The cell debris containing insoluble proteins was resuspended in an identical buffer with the addition of 1.0% Bio-Rad's triton X-100 to increase solubility. This suspension was then distributed into working aliquots and frozen with liquid nitrogen and stored at −80 °C for later use.

Those dinoflagellates that were earmarked for tetrapyrrole extraction were immediately transferred from the filter paper post harvesting to a boiling extraction buffer composed of 2 mM

potassium phosphate, pH 8.5, containing 5 mM 2-mercaptoethanol. Unless otherwise noted, the following procedure was carried out under anaerobic conditions in a MBraun Labmaster SP Glovebox. The suspension was allowed to boil for 5 to 10 minutes to ensure inactivation of enzymes that may be detrimental to the structural integrity of tetrapyrroles. The hot cell suspension was then immediately chilled on ice for 10 minutes. Subsequently, the suspension was transferred to a Dounce homogenizer (chilled in ice) (12 mL Pyrex glass in glass) and the dinoflagellates were ground in an identical manner as stated above. The cell lysate was then transferred to an air-tight ultracentrifuge tube and spun down at 20,000 × g for a period of one hour to pellet the cell debris. The centrifuge tube was returned to the anaerobic environment where the supernatant was decanted and transferred to an air-tight Labconco 250 mL lyophilization vessel. The supernatant was then removed from the anaerobic environment, flash frozen with liquid nitrogen, and subjected to lyophilization. The resultant powder was then fractionated into 40 mg portions and stored at −80 °C until later use.

Cell-free Extract Assays

Chlorophyll *a*, pheophorbide *a*, and pyropheophorbide *a* were individually treated with the cell-free extract and the insoluble fraction of *Pyrocystis fusiformis* and the effects were monitored using UV-visible spectroscopy. A typical reaction contained 500 μL of cell-free extract, 10 μL of a variable reductant and/or cofactor (250 μM final concentration), 1 μL of either chlorophyll *a*, pheophorbide *a*, or pyrophephorbide *a* in dimethyl sulfoxide (50 μM final concentration), and 0.1 M sodium phosphate buffer, pH 6.8, to bring the reaction volume up to 2 mL. The reductants/cofactors included NADH, NADPH, FMN, sodium dithionite, sodium

ascorbate, α-ketoglutarate, and magnesium chloride. The reactions were observed for 25 minutes over a UV-visible spectral range from 300 nm to 800 nm with scanning every 2 seconds.

Purification of P630

The lyophilized tetrapyrrole crude extract powder was dissolved in nanopure water at a concentration of 40 mg per 100 μL. This crude extract was then subjected to 100:1 methanol to crude solution extraction combining with 9.9 mL of methanol to a total volume of 10 mL. The resulting precipitate from the ethanol extraction was then pelleted by centrifugation at 10,000 × g in a Beckman Optima XE-90 Ultra-centrifuge with Type 45 Ti rotor. The supernatant was decanted while the pellet was discarded. Using a stream of nitrogen, the methanol was evaporated to concentrate the sample before HPLC analysis.

The sample was then injected on an Agilent 1260 Infinity Quaternary LC System equipped with a diode array detector (DAD) VL+ and an Agilent Poroshell 120 EC-C18 (4.6 × 150 mm, 2.7 μm) column. Agilent OpenLAB ChemStation Edition software was used for data analysis. The chromatographic method utilized for purification of P630 consisted of the following gradient elution of methanol (solvent A) and water (solvent B): 0-100% A over 15 minutes, 100% A for 15 minutes. The flow rate was 1.0 mL/min with a total run time of 30 minutes and the chromatogram was acquired with detection at 666 nm. Fraction collection was enabled on the instrument to pool all the fractions of P630 from multiple injections of the methanol extracted crude extract.

The reaction mixture containing P710 was initially extracted with methanol to precipitate the proteins in solution. The resulting precipitate was then pelleted via high-speed centrifugation. The supernatant was retrieved while the pellet was discarded. The supernatant was then dried

with a stream of nitrogen to concentrate the sample. The reaction products were then separated using thin layer chromatography (TLC) silica gel plates and ethyl acetate as the mobile phase and visualized using long wavelength UV light (365 nm) using a UVP UVGL-55 handheld UV lamp. Three distinct spots were apparent, which were then each individually isolated by scraping the silica off the plate at the location of each spot.

The reaction products were then extracted from the silica by scrapping off the TLC spot and extraction with ethyl acetate. The sample is then evaporated to dryness, and dissolved in methanol before being subjected to ultra-high pressure liquid chromatography tandem mass spectrometry (UPLC-MS/MS) to identify the mass of each component. Analysis was performed on a Waters Acquity UPLC/Q-TOF Premier Mass Spectrometer equipped with an identical Agilent Poroshell 120 EC-C18 column. For the mass spectrometric analysis, unless otherwise stated, all samples were directly injected into the UPLC-MS, bypassing the chromatography stage completely. The electrospray ionization (ESI) mass detector was configured to positive ion mode with scanning between 300-800 *m*/*z*.

1.3 Results

Reaction Assays and Kinetics

Chlorophyll *a* was incubated with cell-free extracts of *Pyrocystis fusiformis* to see if it could be directly converted to dinoflagellate luciferin. Upon introduction of chlorophyll *a* to the reaction mixture, absorption peaks appeared at 500 nm and 545 nm while the Soret band at 430 nm increased in intensity and became slightly more blue-shifted, with an isosbestic point observable at approximately 425 nm. The characteristic absorption peak of chlorophyll *a* at 660 nm also decreased in intensity over time (Figure 14). Supplementing the reaction with additional

reductants and cofactors (NADH/FMN, NAD(P)H/FMN, sodium dithionite, sodium ascorbate) did not have an appreciable effect on the kinetics of the observed spectral change.

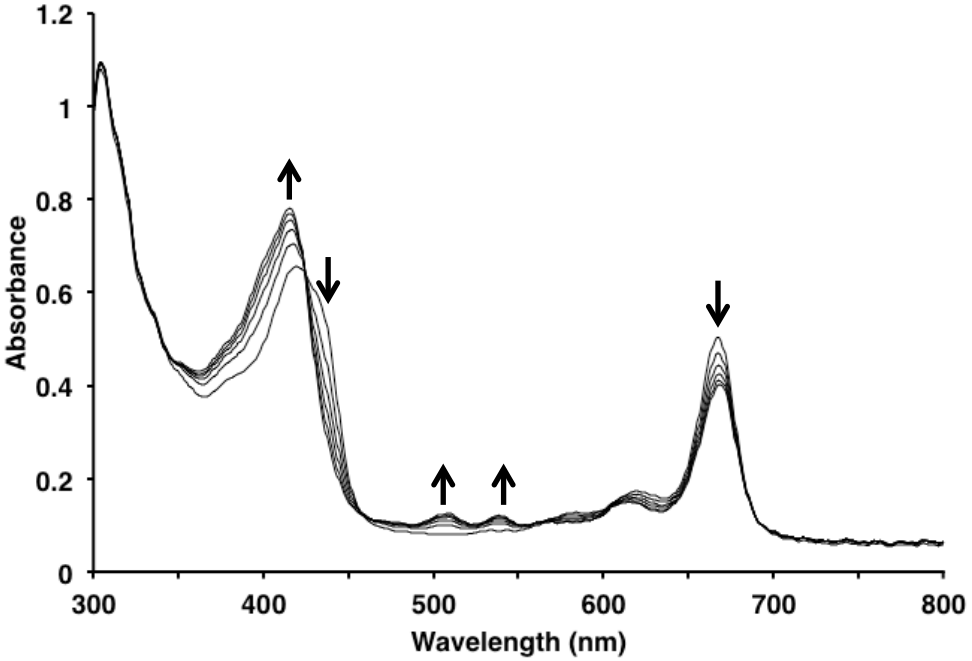


Figure 14. UV-visible spectral changes with increasing absorption at 500 nm, 545 nm, and the Soret band at 430 nm observed upon incubating chlorophyll *a* with *Pyrocystis fusiformis* cell-free extracts.

Pheophorbide *a* was then incubated with *Pyrocystis fusiformis* cell-free extracts. In this case, however, there were no discernable changes observed in the UV-visible spectrum of pheophorbide *a* during the course of the incubation, suggesting that no reaction occurred (Figure 15). The addition of cofactors or reductants also did not appear to have any effect. Strikingly, the UV-visible spectrum of pheophorbide *a* was identical to the spectrum obtained after incubating chlorophyll *a* with the cell-free extracts. This suggests that the extracts are converting chlorophyll *a* to pheophorbide *a*, but are not further converting pheophorbide *a* to downstream catabolites. Additionally, these reactions were carried out with the insoluble fraction resulting in no apparent reaction.

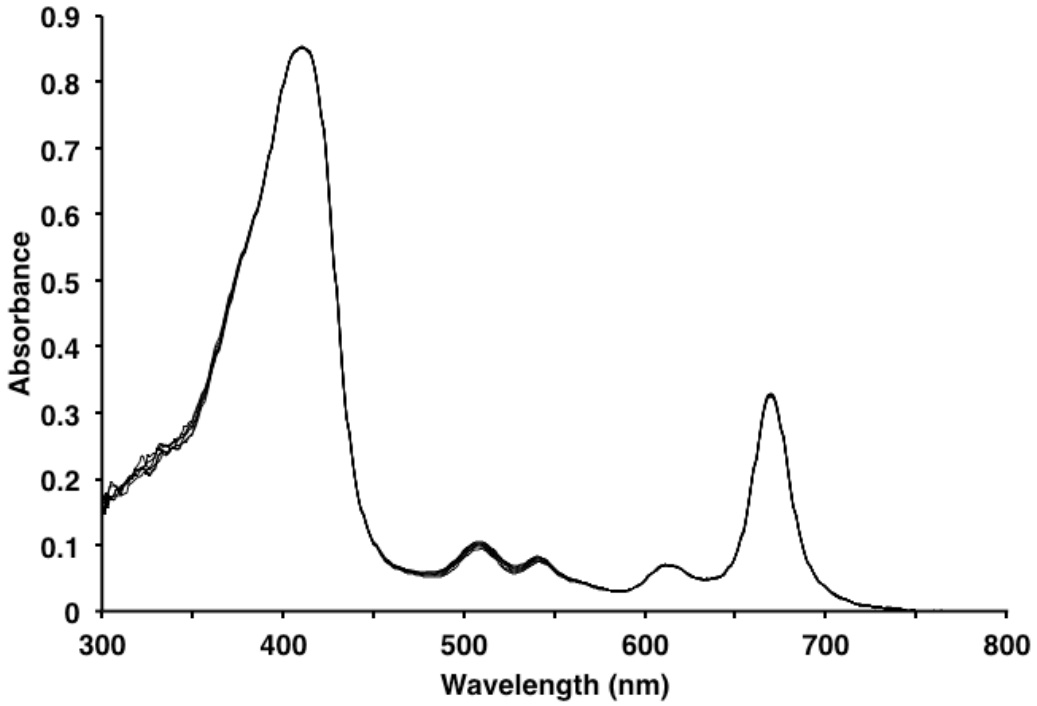


Figure 15. UV-visible absorption spectra of pheophorbide *a* after incubating with *Pyrocystis fusiformis* cell-free extracts showing no apparent reaction.

Failing to advance chlorophyll *a* catabolism beyond formation of pheophorbide *a*, pyropheophorbide *a* was treated with *Pyrocystis fusiformis* cell-free extracts. Almost immediately, a peak began to materialize at 710 nm with a corresponding isosbestic point at 690 nm. Simultaneously, the peak at 390 nm decreases while a new Soret band appears at 450 nm, generating a second isosbestic point at 430 nm (Figure 16).

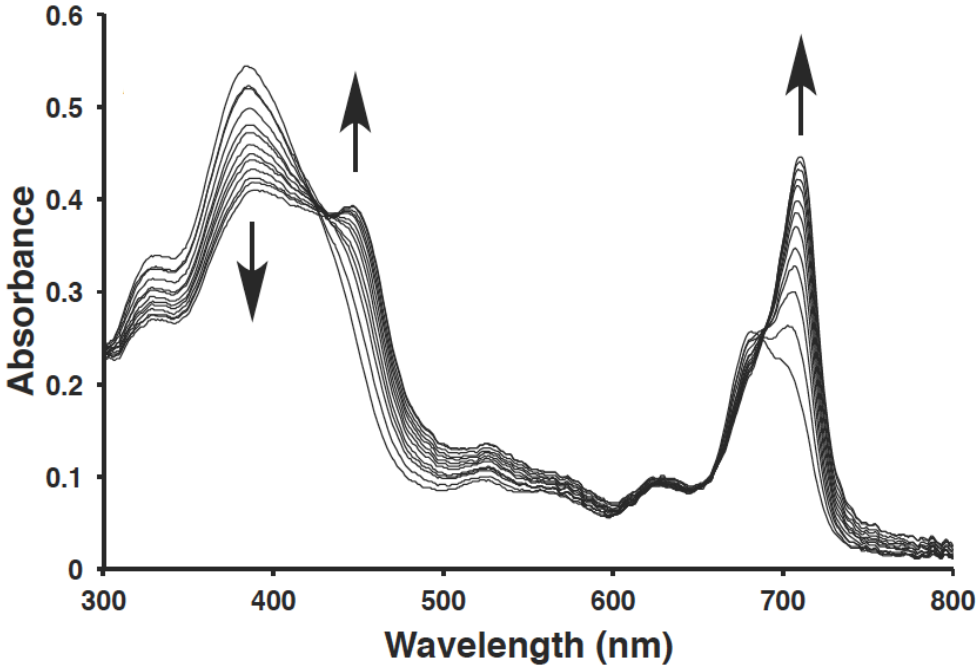


Figure 16. UV-visible spectral changes observed upon incubating pyropheophorbide *a* with *Pyrocystis fusiformis* cell-free extracts showing formation of a reaction product with a 710 nm absorption peak and a corresponding isosbestic point a 690 nm. Simultaneously, the peak at 390 nm decreases while a new Soret band appears at 450 nm, generating a second isosbestic point at 430 nm.

The resulting chromophoric compound, which was named P710 for its intense absorption in the near-infrared region at 710 nm, may be a novel catabolite in the chlorophyll degradation pathway in dinoflagellates. Interestingly, the reaction kinetics of P710 synthesis can be enhanced by supplementing the reaction mixture with either reducing equivalents (NADH/FMN, NAD(P)H/FMN, sodium dithionite, sodium ascorbate) or cofactors such as magnesium chloride. Although the mechanism for this enhancement in reaction velocity has not been determined, it was observed that a mixture of NADH/FMN and magnesium chloride has the greatest effect (Figure 17).

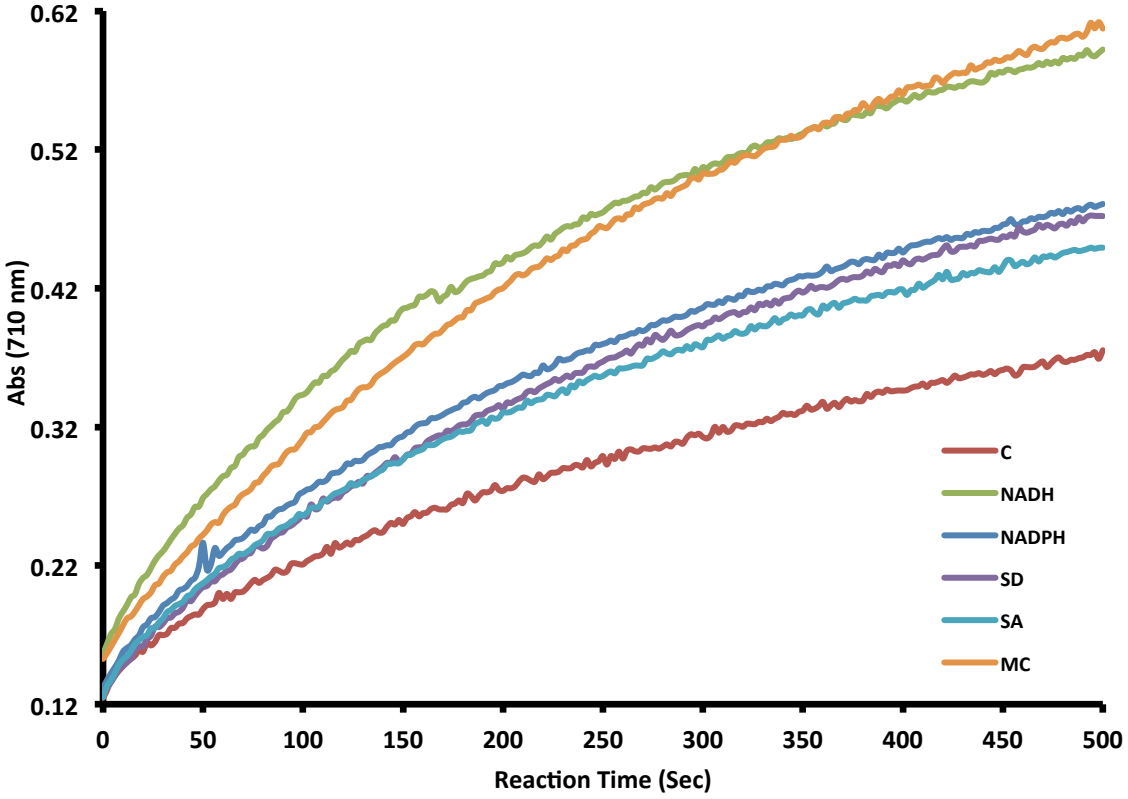


Figure 17. Comparative kinetic assays of pyropheophorbide *a* conversion to P710 by *P. fusiformis* cell-free extracts and in the presence of various cofactors/reducing agents. Legend abbreviations: control (C), nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADPH), sodium dithionite (SD), sodium ascorbate (SA), and magnesium chloride (MC).

It was also observed that the absence of white light had a detrimental effect on the reaction velocity. Although not seemingly required for its formation, directly shining light on the reaction enhanced the yield of P710 (Figure 18).

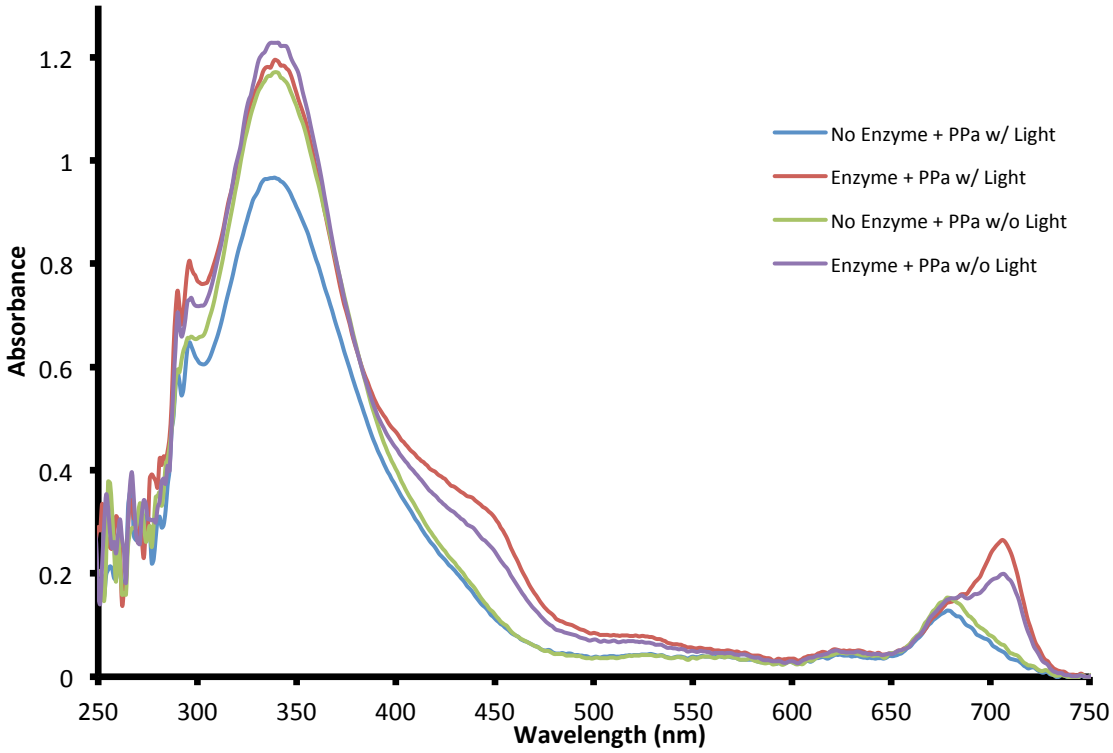


Figure 18. Reaction progress of P710 synthesis after a duration of 1500 seconds showing the effects of the presence and absence of light.

Characterization of P710

The predicted mass-to-charge ratio (*m/z*) of the protonated molecule $[M + H]^+$ of pyropheophorbide *a* is 535.26 *m/z*. When the pyropheophorbide *a* standard was analyzed via mass spectrometry, the mass spectrum contained a major peak at the predicted value.

When the crude pyropheophorbide *a* reaction mixture was analyzed by mass spectrometry, the spectrum yields a mixture of compounds but the three most prominent peaks are closely related based on the mass defect. The most abundant of the protonated molecule peaks possesses a *m/z* identical to that of the starting compound, pyropheophorbide *a*, at 535.26 *m/z*. Intriguingly, the next two most predominant peaks are found at 551.26 *m/z* (M + 16) and 567.26 *m/z* (M + 32), which represent the insertion of one and two oxygen atoms into the

pyropheophorbide *a* starting material, respectively (Figure 19). The sequential insertion of oxygen atoms into pyropheophorbide *a* is consistent with modifications that might lead to oxidative ring cleavage with the appropriate regiochemistry for the formation of dinoflagellate luciferin. However, at this point it is unclear which, if any of these secondary protonated molecule peaks correspond to P710, and the position of oxygen incorporation in pyropheophorbide *a*.

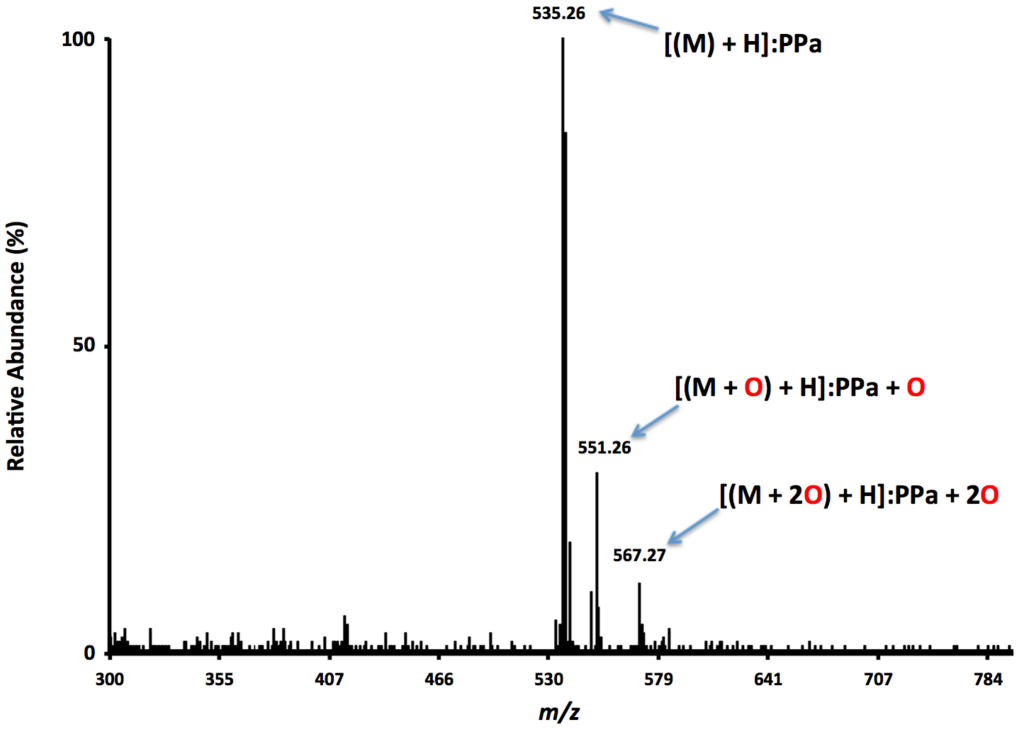


Figure 19. Mass spectrometric analysis of crude pyropheophorbide *a* (PPa) reaction mixtures displaying peaks corresponding to the starting material (535.25 *m*/*z*), pyropheophorbide *a* plus the insertion of one oxygen atom (552.26 *m*/*z*), and pyropheophorbide *a* plus the insertion of 2 oxygen atoms (567.26 *m*/*z*).

Upon separation via TLC, three resolved spots appeared. Thus, the isolated reaction products were subjected to analysis by both UV-visible spectroscopy and mass spectrometry (Figure 20). Consistent with the MS data of the crude reaction mixture, the largest spot on the TLC plate had a retention factor that was identical to the pyropheophorbide *a* standard. Subsequent analyses of the isolated compound by mass spectrometry (showing a single dominant protonated molecule peak at 535.26 *m/z*) and UV-visible spectroscopy confirmed that it was indeed pyropheophorbide *a* (Figure 21.)

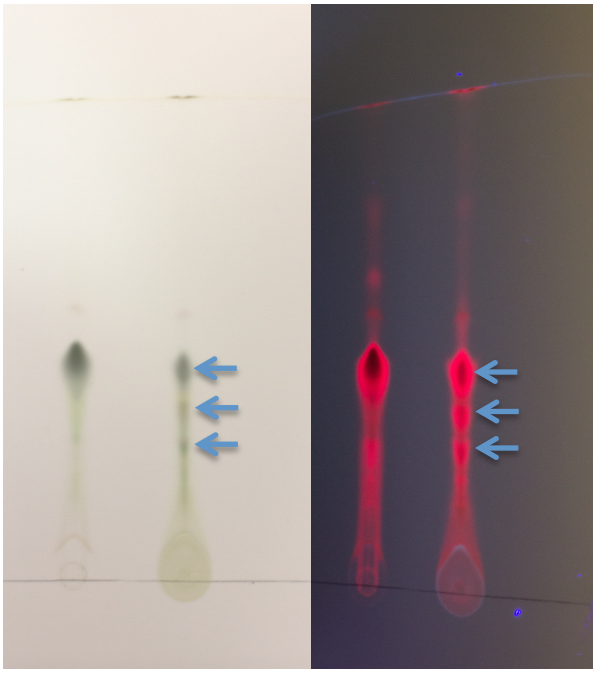


Figure 20. Comparative TLC of pyropheophorbide *a* standard and reaction mixtures with *P. fusiformis* cell-free extracts showing the formation of two distinct products. The TLC plate was photographed under ambient light (left panel) and long wavelength UV light (right panel).

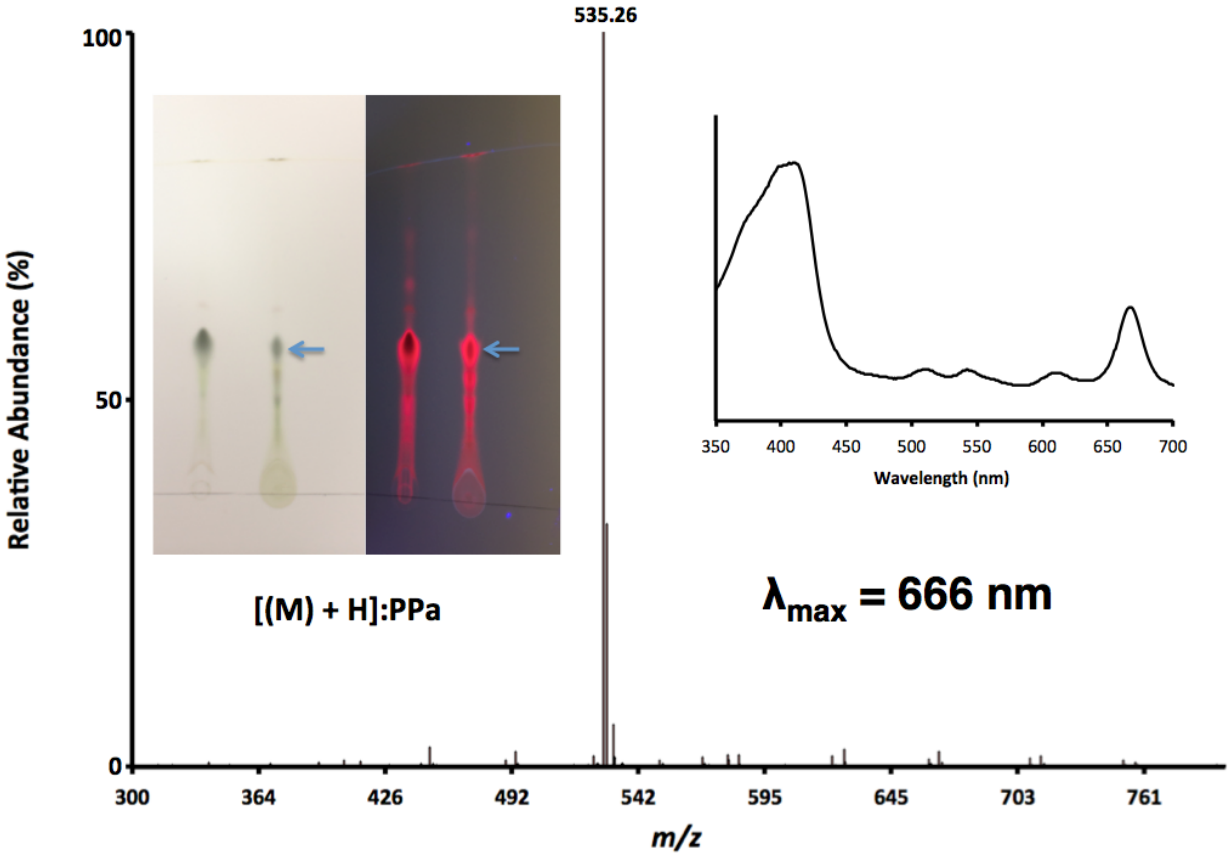


Figure 21. Mass spectrometric and UV-visible analysis of the first TLC spot identifying it as pyropheophorbide *a*.

The second TLC spot was then extracted and analyzed via UV-visible spectroscopy and the resulting spectrum was the same as that of P710. Further analysis by mass spectrometry showed a primary peak at 551.26 *m/z*, which is identical to the second most abundant peak in the mass spectrum of the crude reaction mixture (Figure 22). Thus, the structure of P710 is consistent with pyropheophorbide *a* modified with the insertion of a single oxygen atom.

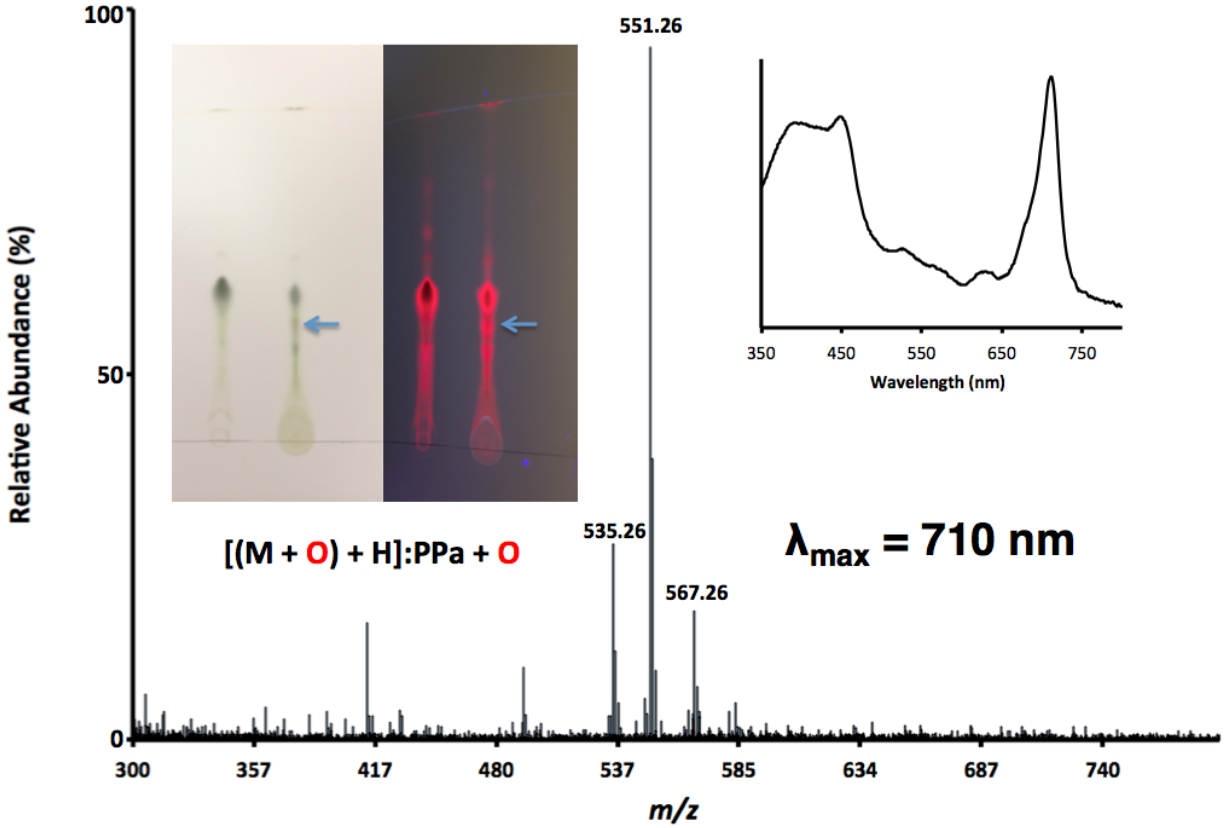


Figure 22. Mass spectrometric and UV-visible analysis of the second TLC spot revealing P710 to be a modified pyropheophorbide *a* containing an additional oxygen atom.

The presence of a third TLC spot, as well as a third prominent mass spectrum peak, was unexpected since the formation of an additional product was not obvious during the course of the reaction of pyropheophorbide *a* with dinoflagellate cell-free extracts. UV-visible analysis of the contents of the third TLC spot reveals a compound with a λ_{max} at 680 nm. Mass spectrometric analysis of this compound, which was named P680 for its light absorption properties, revealed an enriched protonated molecule peak at 567.26 *m/z*. This mass corresponds to the predicted molecular weight of pyropheophorbide *a* modified by the insertion of two oxygen atoms (Figure 23). Both P710 and pyropheophorbide *a* remain present in the sample, as evidenced by the mass spectrum of P680, possibly due to inefficient separation on the TLC plate.

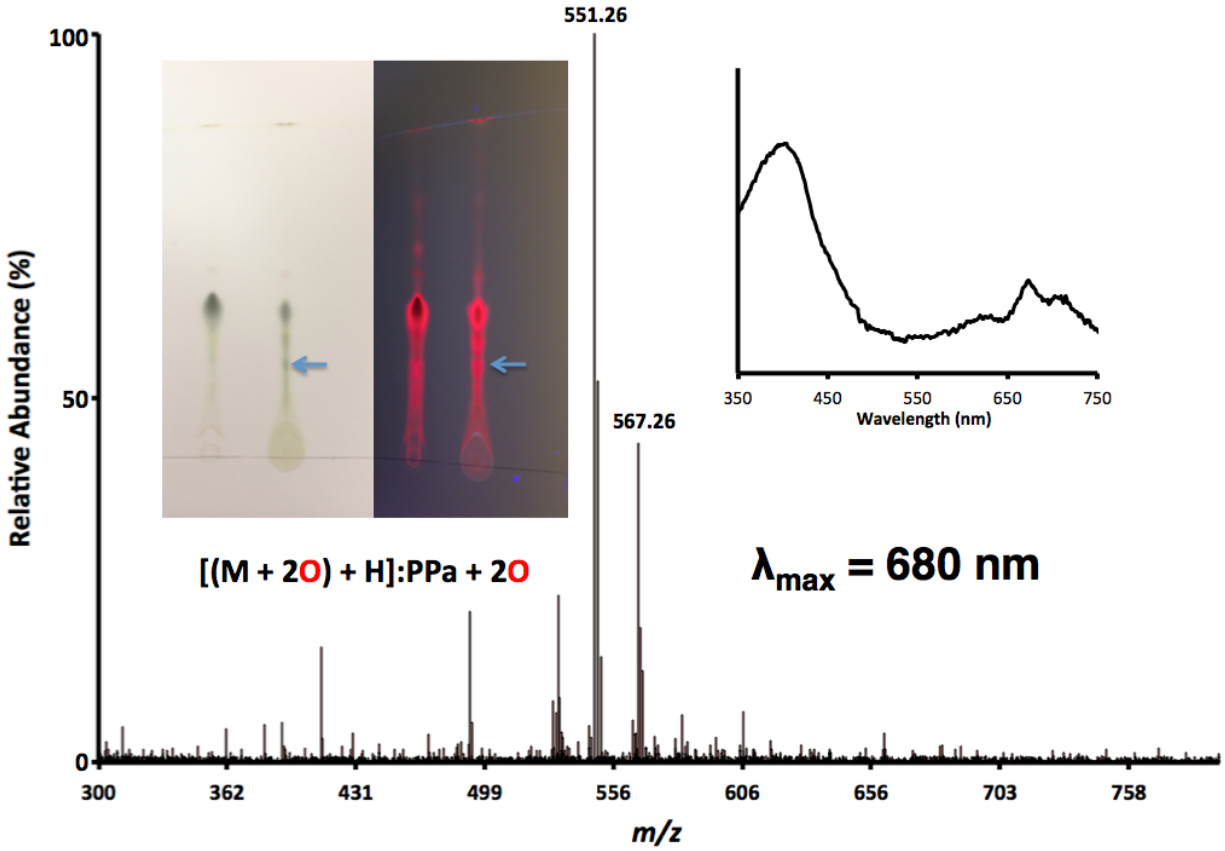


Figure 23. Mass spectrum and UV-visible analysis of third TLC spot featuring enrichment of the 567.26 *m/z* peak in comparison to the previous mass spectra of pheophorbide *a* and P710. This *m/z* is representative of a pyropheophorbide *a* molecule modified with the insertion of two oxygen molecules.

HPLC was also utilized in an attempt to purify sufficient quantities of P710 and P680 for spectroscopic analyses. The HPLC method was identical to that used for the purification of P630 described above. Subjecting the concentrated *P. fusiformis* reaction mixture to HPLC analysis yielded a chromatogram wherein the two most prominent signals were that of P710 and pyropheophorbide *a*, which have retention times of 0.5 min and 14.5 min, respectively (Figure 24). The short retention time of P710 suggests that it was not retained by the stationary phase of the C18 column and a peak for P680 was not detected on the chromatogram.

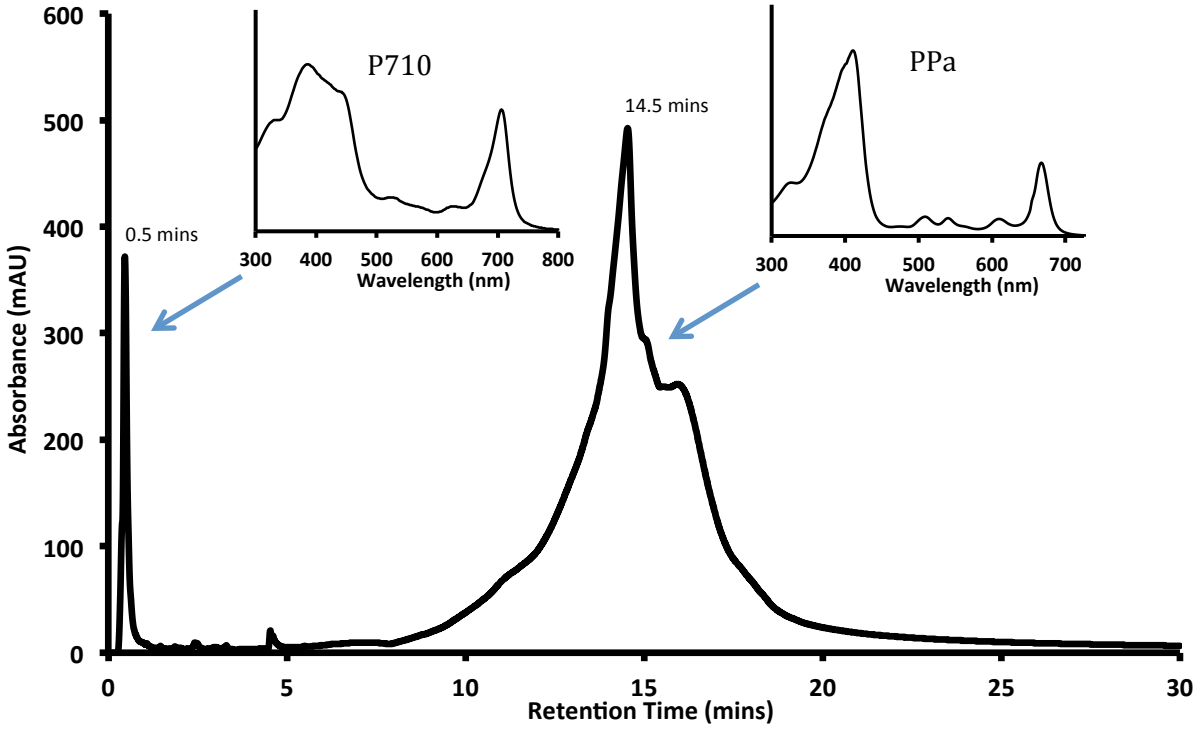


Figure 24. HPLC chromatogram of the *P. fusiformis* reaction featuring well defined peaks for both P710 and pyropheophorbide *a*.

Purified P710 was then analyzed using fluorescence spectroscopy to verify that its spectral properties were distinct from previously discovered chlorophyll derivatives, such as chlorophyll *f*. This red shifted chlorophyll, which differs from chlorophyll *a* by the presence of a formyl group at C2, was discovered in 2010 and has a very similar absorption spectrum as P710. A fluorescence excitation scan of P710 revealed excitation maxima at both 390 nm and 712 nm. Consequently, when P710 was excited at 390 nm, it fluorescenced with an emission maximum of 664 nm (Figure 25). This differs greatly from the fluorescence emission maximum of chlorophyll *f* at 731 nm [49]. Based on the spectroscopic and MS data obtained, our proposal for the structures of P710 and P680 are presented (Figure 26).

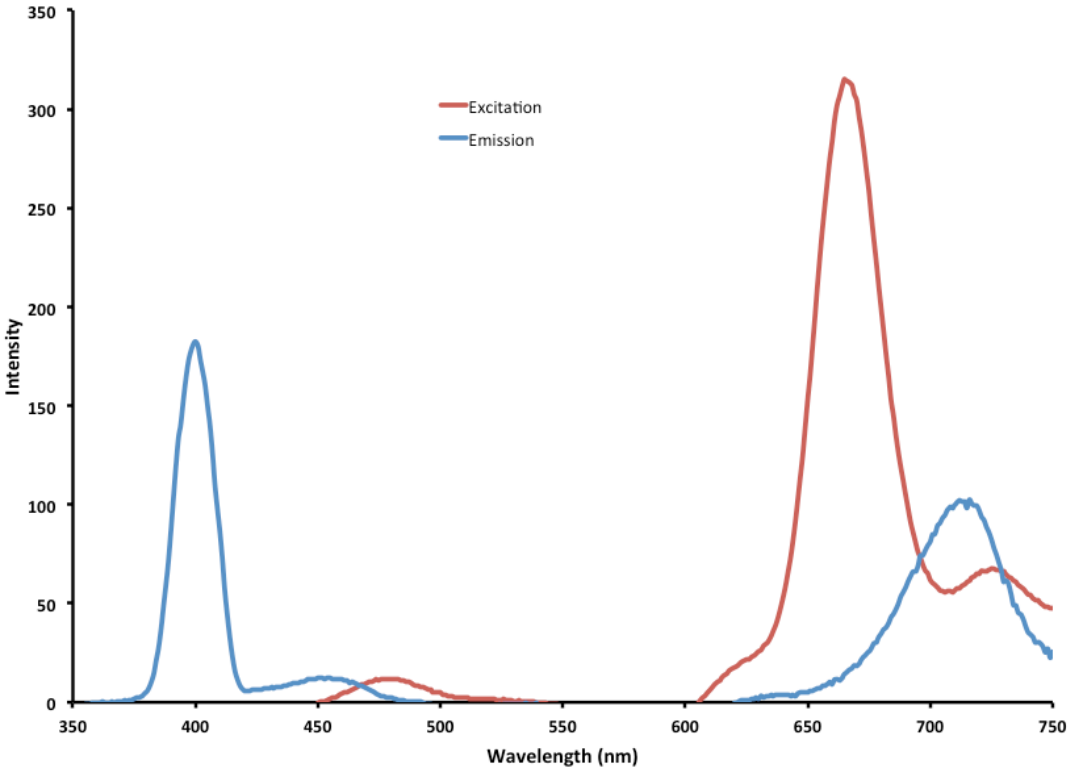


Figure 25. Fluorescence excitation and emission spectra of P710 in aqueous solution.

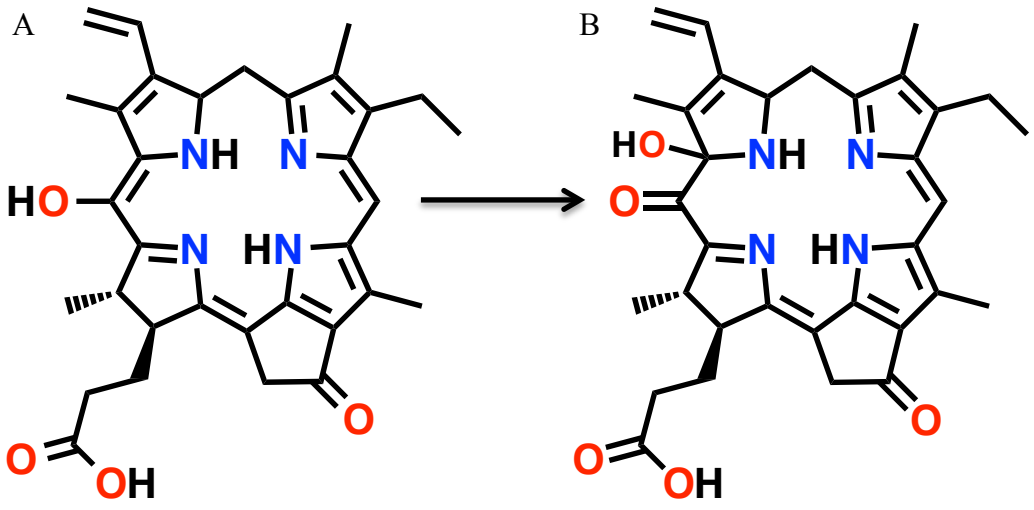


Figure 26. The proposed structures of (A) P710 and (B) P680 based on MS evidence.

Identification of P710 Synthase

In an effort to identify the enzyme responsible for the synthesis of P710, HPLC was used to fractionate the proteome of *Pyrocystis fusiformis*. The proteins were separated on a size exclusion column and separated into a 96-well plate. NADH/FMN and magnesium chloride were then added to each well to help maximize the yield of P710. Pyropheophorbide *a* was then introduced and the reaction progress monitored. Based on these results, P710 synthase activity was only detected in the first fraction. Using molecular weight standards, the calculated molecular weight of this putative P710 synthase is approximately 240 kDa (Figure 27).

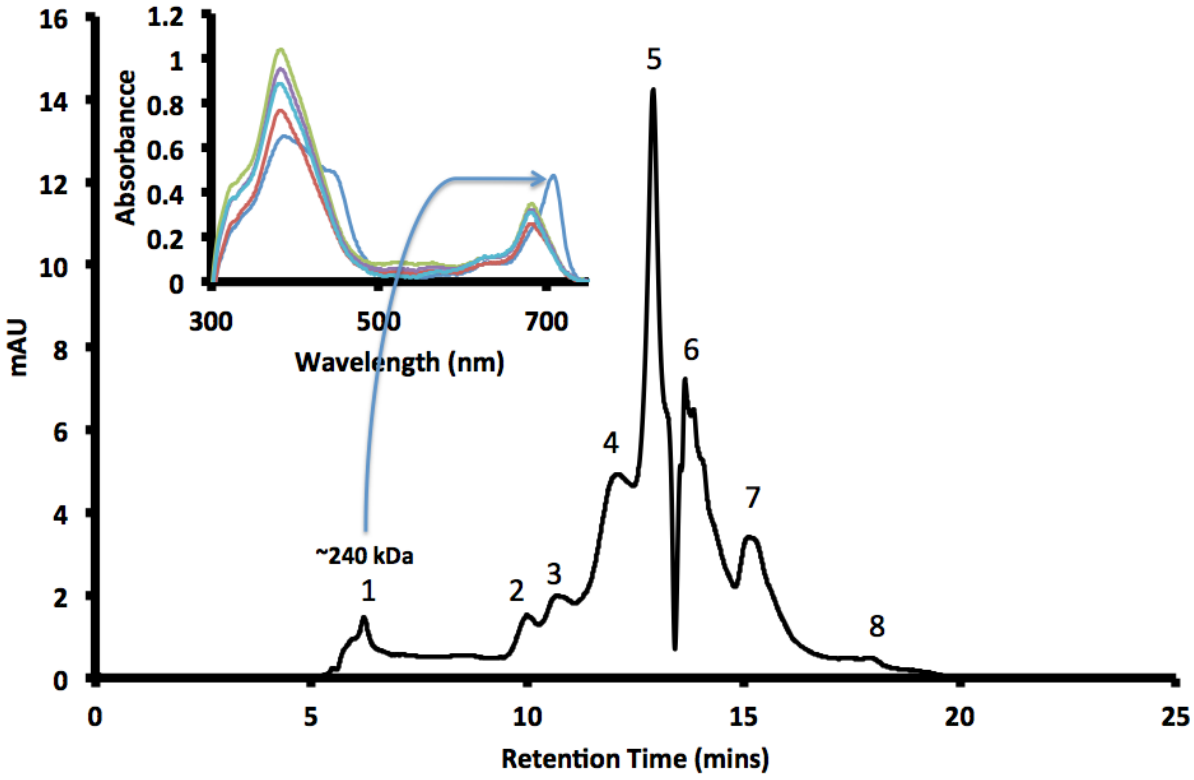


Figure 27. Size exclusion chromatography of the *P. fusiformis* proteome. P710 synthase activity was only detected in the first ~240 kDa.

Characterization of P630

HPLC purification was carried out on crude extracts of P630 according to the above stated gradient program. Using this method, P630 eluted with a retention time of 7.3 minutes (Figure 28). The pooled P630 fractions were concentrated using a stream of nitrogen gas. The concentrated solution of P630 exhibited a characteristic intense blue hue and a UV-visible spectrum strongly resembling that of the blue oxidation product of dinoflagellate luciferin.

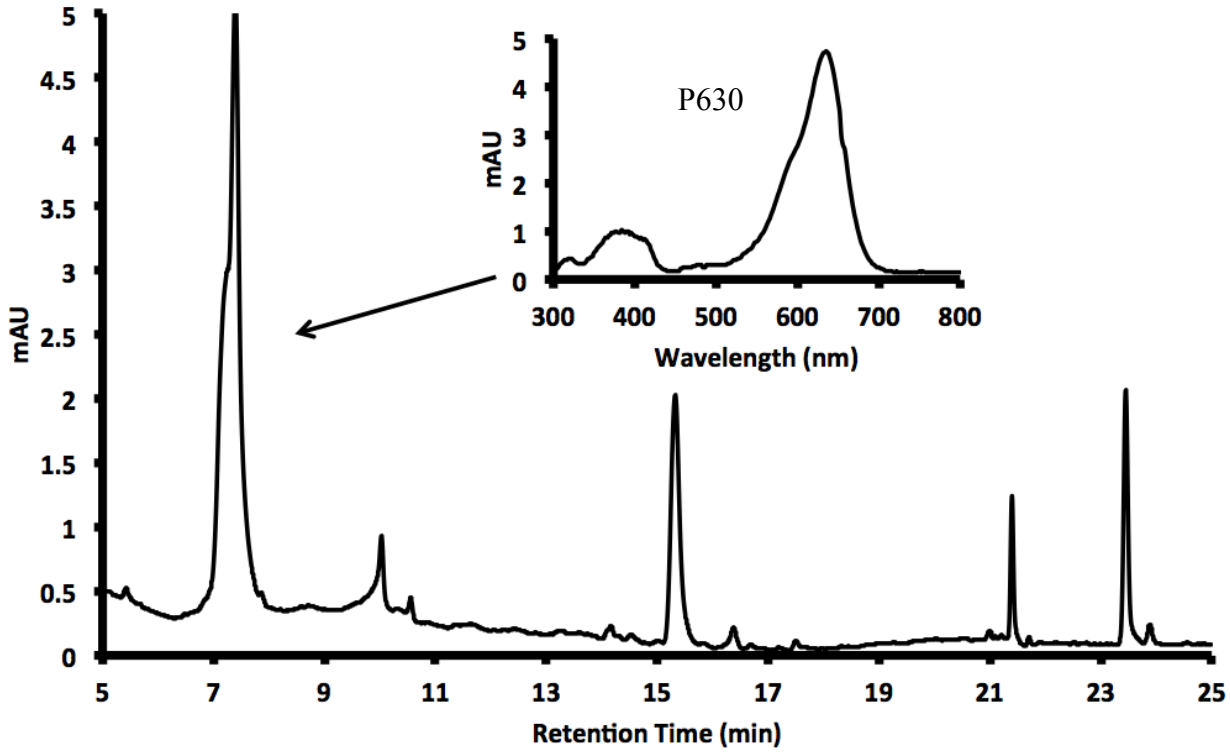


Figure 28. HPLC chromatogram of a crude dinoflagellate extract featuring a strong chromatographic peak for P630 at 7.3 minutes.

This purified P630 was then subjected to analysis via UHPLC-MS/MS utilizing an identical gradient program and mass analysis program as stated above for P710 and P680 (Figure 29). Additionally, the sample was treated with dithiolthreitol (DTT), and consistent with literature was converted to dinoflagellate luciferin (Figure 30).

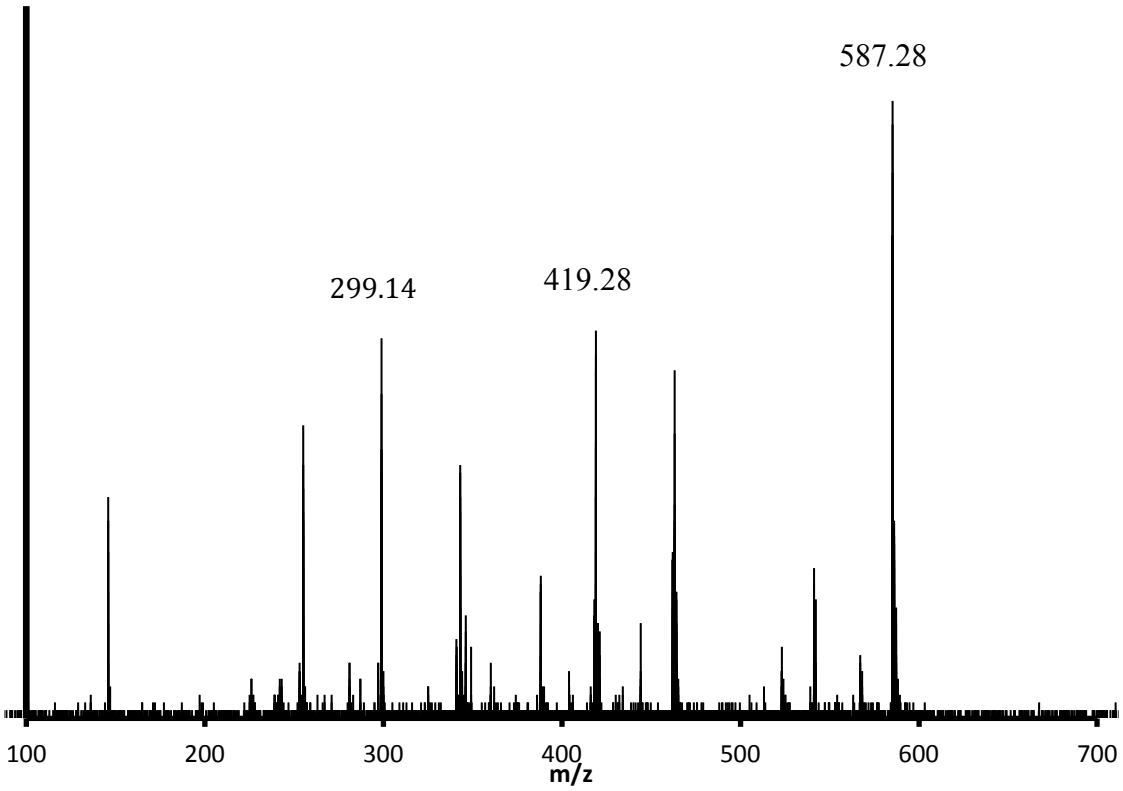


Figure 29. LC-MS data of P630.

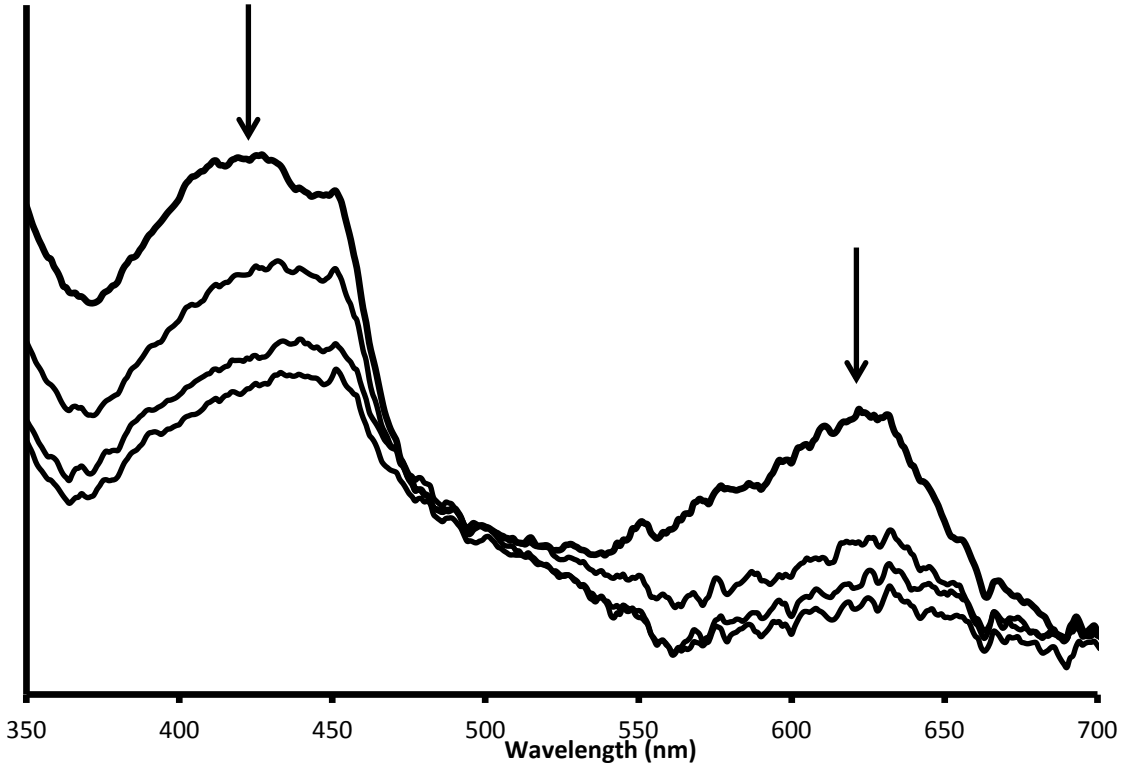


Figure 30. Reduction of P630 to dinoflagellate luciferin by DTT.

1.4 Discussion

With mass spectrometry indicating that P710 is the product of the addition of a single atom of oxygen, we propose that carbon 20 of the δ bridge of the tetrapyrrole is the site of oxygenation. Thus, we find that this structure is consistent with the type of reaction that would be required for the next step of the biosynthesis of luciferin after pyropheophorbide *a*. The use of tandem mass spectrometry and nuclear magnetic resonance can confirm our proposal. Oxidative cleavage of pyropheophorbide *a* leading to the open-chain tetrapyrrolic structure of dinoflagellate luciferin requires the net addition of three oxygen atoms [27]. When pheophorbide *a* is oxidatively cleaved to yield the red chlorophyll catabolite, there is only a net gain of two oxygen atoms at the site of ring opening, the α bridge. This reaction is catalyzed by a single

enzyme, pheophorbide *a* oxygenase [34]. Therefore, the triple oxygenation of pyropheophorbide *a* leading to ring opening for luciferin biosynthesis is likely to require multiple enzyme-catalyzed steps. Based on the findings in these experiments, it is possible that the novel chlorophyll catabolite P710 is the first intermediate in the reactions required for the tetrapyrrole ring opening. Additionally, the discovery of P680 further supports this assertion. Mass spectrometric analysis of P680 demonstrates that it is closely related to both pyropheophorbide *a* and P710 by their common mass defect. Also, its observed mass is indicative of pyropheophorbide *a* augmented with the insertion of two oxygen atoms (or equivalently P710 with the addition of one oxygen atom). Based on these data, we propose the following biosynthetic pathway for dinoflagellate luciferin (Figure 31).

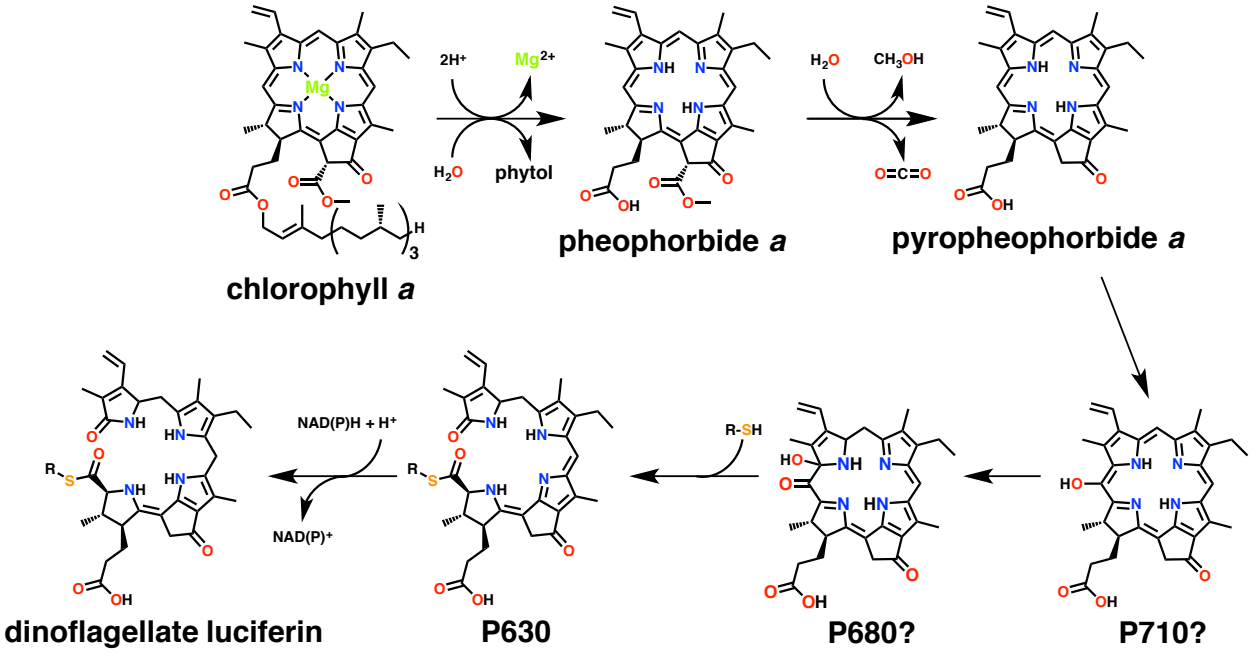


Figure 31. Proposed pathway for dinoflagellate luciferin biosynthesis from chlorophyll *a*.

Chromopeptide formation was included in our proposed biosynthetic pathway for dinoflagellate luciferin after the synthesis of P680 and concomitant with the formation of P630. This is consistent with the observation that both P710 and P680 were formed by incubating pyropheophorbide *a* with *P. fusiformis* cell-free extracts, while the reaction did not seem to progress further to produce P630. Therefore, the lack of observed P630 formation may be due to the lack of sufficient quantities of the peptide in the cell-free extracts needed to drive the reaction forward. The purpose of the peptide attached to P630 and dinoflagellate luciferin has yet to be determined. One possibility is that it plays the role of the luciferin binding protein in *Pyrocystis* species and stabilizes luciferin until needed for the bioluminescence reaction. Another possibility suggested by the current study is that chromopeptide formation is coupled to the ring opening reaction. Yet another possibility is that the chromopeptide itself may be involved in docking the substrate to the luciferase enzyme. In any case, elucidating the true nature of the chromopeptide is likely to provide great insight into the biosynthesis luciferin and the mechanism of the bioluminescence reaction.

During the initial purification of dinoflagellate luciferin from natural sources, Woodland Hastings and coworkers, identified a blue compound, which they interpreted as an air-oxidation product of luciferin [29]. Previously, they showed that dinoflagellate luciferin could be chemically converted to a blue oxidation product by treating it with 0.1% I_2 in ethanol. The Hastings group then characterized the structure of the blue oxidation product by high-resolution mass spectrometry and NMR spectroscopy and showed it to be two electrons more oxidized than luciferin [29]. Fresneau and coworkers at the Institut de Biochimie, Universite' de Paris-Sud in France also isolated a blue compound (P630) from dinoflagellates, which they characterized using UV-visible and fluorescence spectroscopy and determined was the direct biosynthetic

precursor to dinoflagellate luciferin. Specifically, a NAD(P)H-dependent reductase was identified that could convert P630 to dinoflagellate luciferin. However, the structure of P630 was not elucidated. Comparison of the UV-visible spectrum of the blue oxidation product reported by Hastings and coworkers with that of P630 (reported here or in the literature by Fresneau and colleagues) show that they are essentially identical. Furthermore, the mass spectrometric analysis of P630 reported here is also consistent with the structure reported for the blue oxidation product. We therefore assign the structure of P630 to be that reported for the blue oxidation product, and conversely, that the blue oxidation product is not an artifact of the aerobic purification of dinoflagellate luciferin but is its immediate biosynthetic precursor. This explains why exposure of dinoflagellate luciferin to oxygen does not form the blue oxidation product, but rather the non-enzymatic air-oxidation product.

In summary, two novel chlorophyll catabolites (P710 and P680) that are potential intermediates in the dinoflagellate luciferin biosynthetic pathway were identified by treating pyropheophorbide *a* with *P. fusiformis* cell-free extracts. P710 synthase activity was isolated to the ~240 kDa fraction of the *P. fusiformis* proteome. In addition, the structure of P630 was confirmed to be that of the blue oxidation product of dinoflagellate luciferin. Together, these data provide significant insight into the biochemistry of bioluminescent dinoflagellates.

Chapter 2

The pH Regulation of Dinoflagellate Luciferase

Probing the pH Regulation Mechanism of Dinoflagellate Luciferase Using Constant pH Accelerated Molecular Dynamics

2.1 Background

Dinoflagellates are a group of marine protozoa that are capable of fantastic displays of bioluminescence when physically agitated [50]. The chemical reaction responsible for this phenomenon involves the enzyme-catalyzed oxidation of a substrate, dinoflagellate luciferin, by an enzyme, dinoflagellate luciferase, to yield an oxyluciferin product and the release of electromagnetic radiation at 475 nm [51]. Dinoflagellate luciferin is a open-chain tetrapyrrole molecule derived from chlorophyll *a* that is conserved throughout all dinoflagellate species, though its origin is not yet completely understood [26]. Interestingly, luciferases share sequence identity of approximately 73% across the various species of dinoflagellates and, with exception of one bioluminescent species (*Noctiluca scintillans*), the majority contain three catalytically active domains on a single polypeptide [36, 52]. Additionally, several luciferase open reading frames are arranged as tandem repeats separated by non-coding intergenic space in the dinoflagellate genome (Figure 32) [37] .

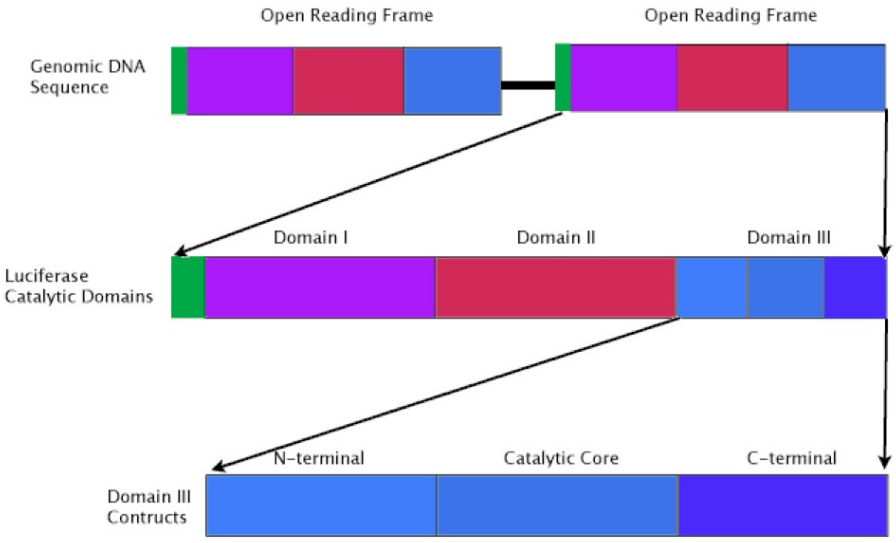


Figure 32. Schematic of the open reading frame of a typical dinoflagellate luciferase gene as it is arranged in the genome.

Full-length luciferase includes an approximately 110 amino acid N-terminal region in addition to the tandemly arranged triplicated domains of approximately 377 amino acids each, making its total molecular weight approximately 137 kDa [53]. Structurally, the N-terminus of a single domain consists of an α-helical bundle while the active site region resides within a β-barrel [28]. With the exception of two α-helices, the C-terminus adopts a largely random loop conformation in the crystal structure, yet seems to be required for bioluminescence activity [28, 41].

The N-terminal α-helical bundle is of great interest due to the presence of four conserved histidines that are believed to be involved in the regulation of the enzyme as a function of pH [40]. It has been previously established that luciferase is inactive at pH 8.0 and is activated when the pH drops to 6.3 [40]. This revelation was due in part by the fact that scintillons, intracellular organelles present in dinoflagellates as outpocketings that protrude from acidic vacuoles and contain all components required for bioluminescence, open voltage-gated proton channels upon

physical agitation. This allows protons to flood into the scintillons, thus activating luciferase and ultimately initiating bioluminescence [13, 14]. Based on the crystal structure of domain III of luciferase from *Lingulodinium polyedrum*, it is postulated that His899, His909, His924, and His930 are protonated when the pH is lowered, thus inducing a large conformational change in the enzyme that allows the substrate to access the active site (Figure 33) [28].

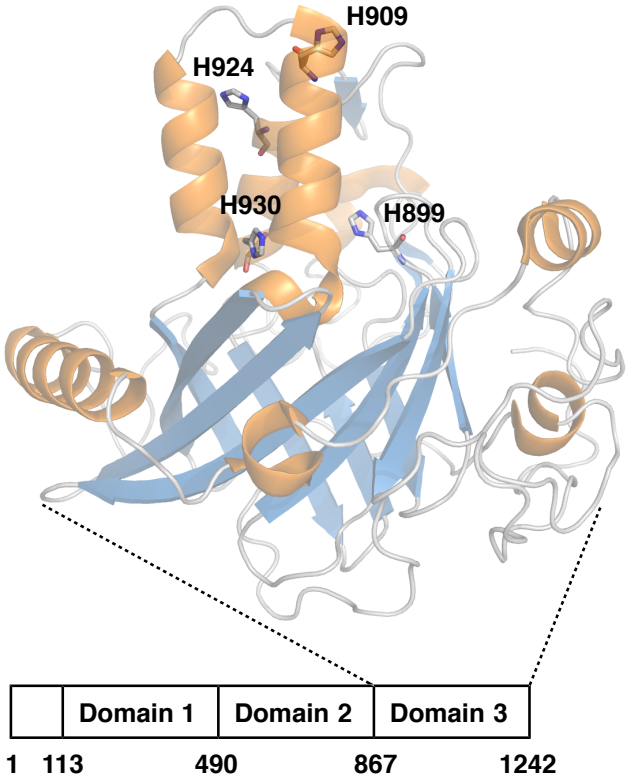


Figure 33. Crystal structure of domain III of luciferase from *Lingulodinium polyedrum* featuring the N-terminal α-helical bundle and the β-barrel encompassing the active site. H899, H909, H924, and H930 are also indicated.

Previous attempts to probe this unique pH regulation mechanism experimentally involved the site directed mutagenesis of the four aforementioned histidines into alanines, and by the complete removal of the α-helical bundle. These experiments demonstrated that these four histidines were responsible for the loss of activity in luciferase at pH 8.0 [40]. Molecular

dynamic simulations were also carried out on domain III of luciferase from *Lingulodinium polyedrum* in an effort to gain some theoretical insight subsequent to solving the crystal structure. These rudimentary calculations were performed first by mutating the histidines of interest in the α helical bundle with alanines and then were performed with each of the histidines artificially protonated. The simulations were allowed to run for a 10 nanosecond timescale. In these simulations, it was determined that the volume of the active site pocket expanded from approximately 174 $Å^3$ to 693 $Å^3$. This volume expansion included the opening of a solvent channel through the α-helical bundle to accommodate the luciferin substrate [28]. While these results are promising, the computational methodology can be improved.

The protonation of an amino acid residue is a dynamic process that transiently alters the electrostatics of the system [54]. Artificially protonating the four histidines of the α-helical bundle responsible for regulation bypasses a random sampling process and biases the protonation state of the residue to one extreme [55]. In addition, large conformational changes in proteins occur over timescales of microseconds and beyond [56]. Running a molecular dynamics calculation for only 10 ns may not be long enough to allow the trajectories to fully sample conformational space during the course of the simulation.

This study attempts to address these issues by employing and coupling two relatively new molecular dynamics methodologies: constant pH molecular dynamics (CpHMD) to address the artificial protonation bias and accelerated molecular dynamics (aMD) to address the timescale issue. In CpHMD, pH is treated as an external thermodynamic parameter. Thus, in a Generalized Born implicit solvent, a combined molecular dynamics/Monte Carlo approach is used at each step to sample the conformations and protonation states of titratable residues in a protein [55, 57]. A new protonation state is chosen at random and the decision to accept this change is based

on the Metropolis criterion in which the total transition energy, ΔG, is used. The calculation of ΔG in this case is described in equation 1:

$$\Delta G = k_B T(\text{pH} - \text{pK}_{a,\text{ref}})\ln 10 + \Delta G_{\text{elec}} - \Delta G_{\text{elec,ref}} \quad (1)$$

In this equation, k_B is the Boltzmann constant, T is the temperature, pH is the specified solvent pH, $pK_{a,ref}$ is the pK_a of the reference compound, ΔG_{elec} is the electrostatic energy component of the titratable residue, and $\Delta G_{elec,ref}$ is the electrostatic component of the transition energy. In the event that the Monte Carlo move has met the Metropolis criterion, the move is accepted and a new protonation state will be applied to the residue, and the simulation continues. If the Monte Carlo move is not accepted, the residue's protonation state remains unchanged and the simulation continues [54, 58].

As previously stated, biological processes that require large conformational changes occur on a time regime of microseconds and beyond. Using CpH simulations with conventional molecular dynamics only allows practical access to the 10-100 nanosecond timescales. Therefore, accelerated molecular dynamics was employed. Accelerated molecular dynamics works by applying a bias potential to the true potential. This enhances the sampling and escape rates from the potential wells, which allows various conformations of a protein to be accessed routinely even when they occur over millisecond time regimes [59, 60].

2.2 Computational Methodology

Using *Pyrocystis fusiformis* as the model organism, the tertiary structure of domain I, domain II, and domain III of luciferase was predicted with the I-TASSER (Iterative Threading ASSEmbly Refinement) protein structure and function prediction software from the Zhang Lab at the University of Michigan. The crystal structure of luciferase domain III from *Lingulodinium polyedrum* (PDB: IVPR) was used as the homology model in these calculations.

All molecular dynamics simulations were performed using AMBER (Assisted Model Building with Energy Refinement) 14 and AMBER Tools 15 installed at the Alabama Supercomputing Authority. The xLeap tool from AMBER Tools 15 was utilized to generate the topology and input files employing the ff99SB force field in a Generalized Born implicit solvent. The constant pH input file was then created using the cpinutil Python script also available from AMBER Tools 15. For the purposes of this study, the input file allowed for histidine and cysteine residues to be tiratable, since the pK_a of these two residues, 6.0 and 8.2 respectively, are the only ones expected to lie within 1 pH unit of the range of pH values used in the CpH simulations. pH values of pH 6 and 8 were chosen for the simulations because they are near the pH environments at which luciferase activity is maximal and near zero, respectively. Minimization, heating, and equilibration of each protein structure were carried out by Multisander with message passing interface (MPI) for parallel computing using a salt concentration (Debye–Hückel based) of 0.1 M, a cutoff for non-bonded interactions and the calculation of effective Born radii of 30 Å, a temperature of 300 K, and a time constant of 2 ps. SHAKE was also employed to constrain bonds that included hydrogen. The time step had a duration of 2 fs.

Prior to initiating the constant pH accelerated molecular dynamics hybrid simulation, a conventional molecular dynamics (cMD) simulation had to be carried out to obtain required values for an accelerated molecular dynamics simulation. These values were the average total potential energy and the average dihedral energy over 5000 steps and were used to calculate the EthreshP, alphaP, EthreshD, and alphaD parameters. The production simulations for the cMD were performed for 10 ns with identical settings as stated above for the equilibrium simulations. The AMBER program pmemd.cuda was used in lieu of Multisander for the cMD production simulations so that graphics processing units (GPUs) could be utilized at this stage of the calculation.

Constant pH molecular dynamics was then coupled with accelerated molecular dynamics by adding the flag iamd=3 to the input file. The EthreshP, alphaP, EthreshD, and alphaD values were also defined in the input file for each domain. All other settings remained the same as for the conventional molecular dynamics input file with one exception. The length of the simulation was lengthened to 1 μs of accelerated molecular dynamics (which is equivalent to 2 ms of unbiased molecular dynamics simulation) to enable the observation of large conformational shifts in luciferase during the course of the trajectory [56, 59]. All CpHaMD simulations were carried out on GPUs employing the pmemd.cuda program. Three trials of each of the three luciferase domains were performed at both pH 6 and pH 8. When examining the C-terminal region of the luciferase crystal structure, two pairs of cysteine residues with the potential to form disulfide bonds were observed (between Cys116 and Cys273 and between Cys127 and Cys141). To ascertain if a redox regulatory mechanism is also possible, a secondary set of identical simulations were performed on each of the three luciferase domains containing two disulfide bonds in the C-terminal region.

2.3 Results

The crystal structure of luciferase domain III used in the homology models were crystalized at pH 8, presumably in the inactive form. Thus, CpHaMD simulations performed on luciferase at this pH were expected to remain relative stable, displaying no major conformational changes. In contrast, collective motions representing the conformational activation of luciferase (e.g., in the α-helical bundle containing the proposed regulatory histidines) were anticipated in the CpHaMD simulations conducted at pH 6.

The I-TASSER predicted structure of *Pyrocystis fusiformis* domain III is presented in Figure 34. A root-mean-squared deviation (RMSD) analysis of the MD trajectories per residue indicates that the majority of the movement in the enzyme occurs in the N-terminal region situated in proximity to the proposed regulatory α-helical bundle (Figure 35). In examining the trajectories from the domain III simulations at pH 6 (with no disulfide bonds specified), there is an obvious conformation shift when observing the distance between His35 and His66 (equivalent to His899 and His930 in full-length luciferase) in the α helical region (Figure 36). After approximately 4 μs, the distance between these histidines increased from ~11 Å to ~16 Å. In contrast, the relative positions of these two histidines remained relatively stable, experiencing no statistically significant distance variation during the course of the trajectory at pH 8 (Figure 36). Similar results were observed in domain III simulations containing disulfide bonds (Figure 37). In all of the MD simulations, the inter-residue distance remained relatively constant after 20 μs. Thus, the data shown does not exceed 20 μs even though the trajectories were carried out for at least 2000 μs in all cases.

To assess whether this trend extended to luciferase domains of other dinoflagellate species, identical simulations were performed on luciferase domain III of *Lingulodinium*

polyedrum. This domain also behaved similarly, exhibiting a conformation change in the pH 6 environment while remaining in relatively stable and experiencing no major conformation shifts when the simulation was maintained at pH 8 (Figure 38). A visualization of the trajectory of the pH 6 simulation of domain III also shows an obvious conformation change of the loop containing His66 leading to a collective motion in the α-helical bundle that appears to be spreading apart, possibly for substrate accessibility (Figure 39).

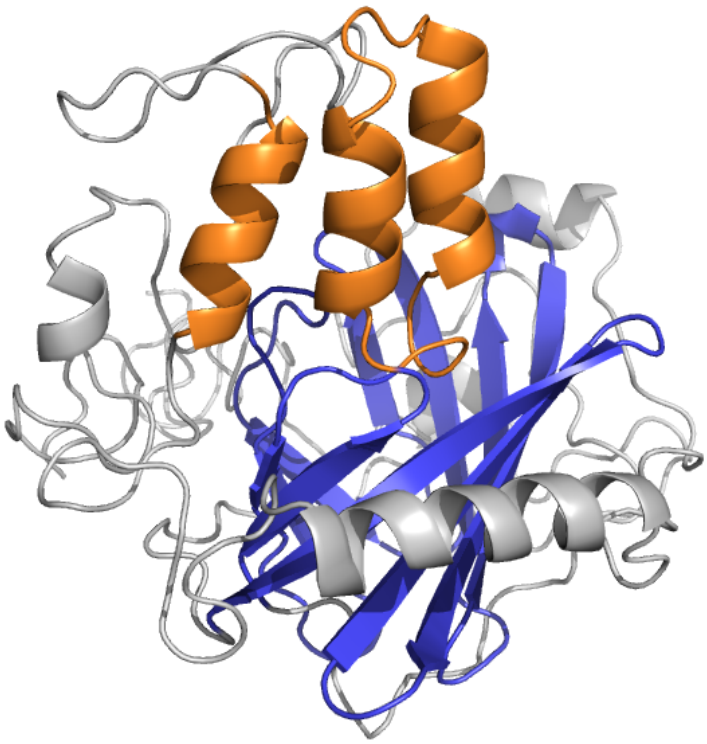


Figure 34. I-TASSER predicted structure of *Pyrocystis fusiformis* luciferase domain III.

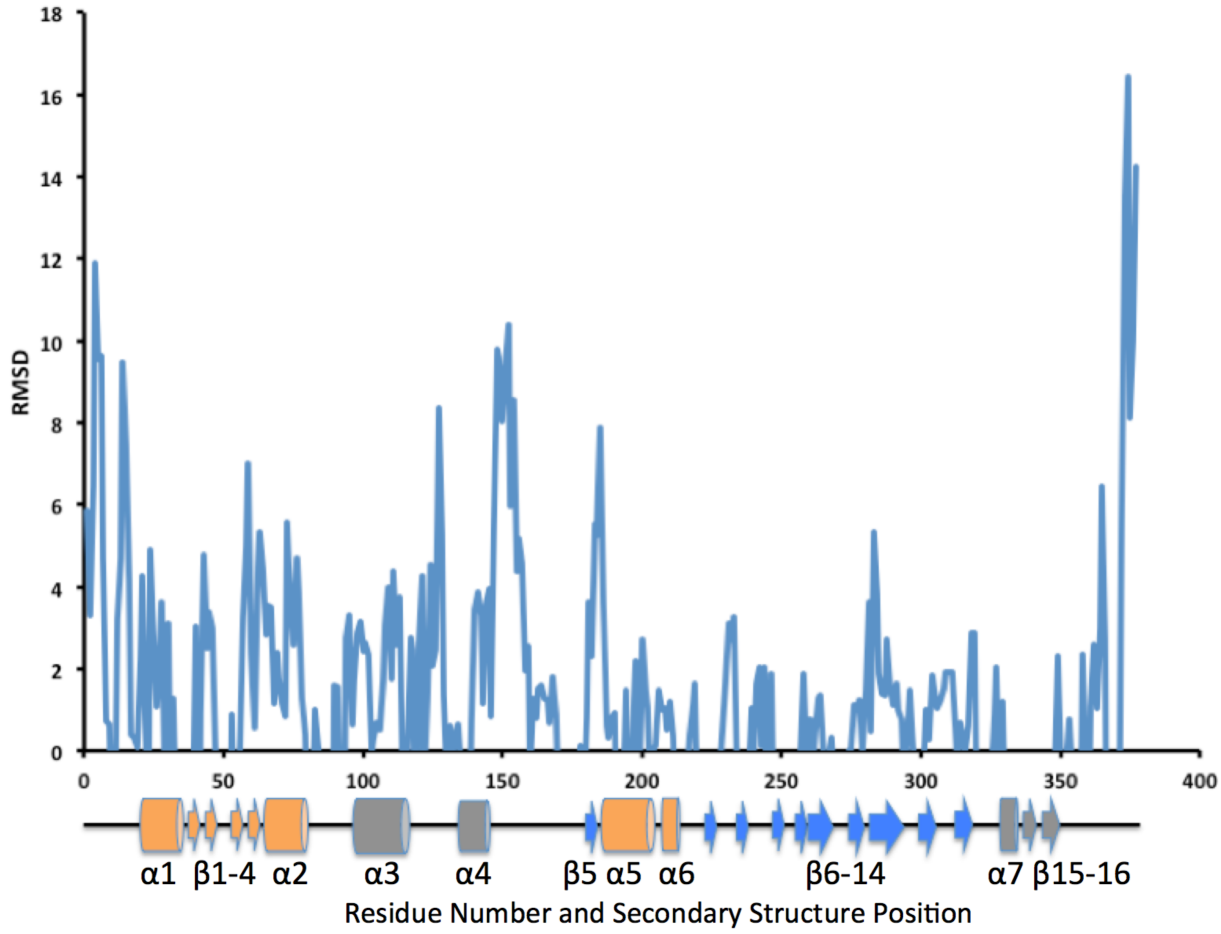


Figure 35. Per residue root-mean-squared deviation of *Pyrocystis fusiformis* luciferase domain III during the course of a MD trajectory with secondary structure specified.

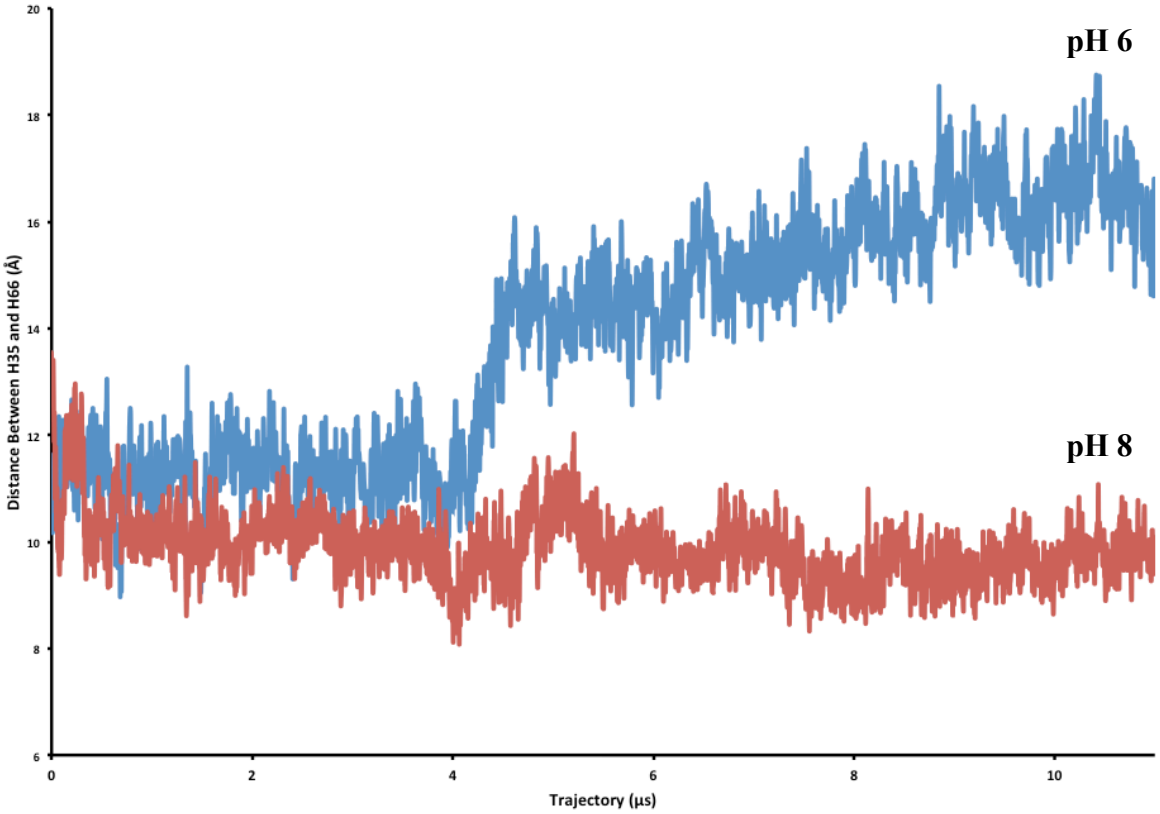


Figure 36. Residue-to-residue distance over the course of the trajectory of His35 and His66 in *Pyrocystis fusiformis* luciferase domain III featuring divergence of the two residues after ~4 µs at pH 6 and no divergence of these two residues at pH 8.

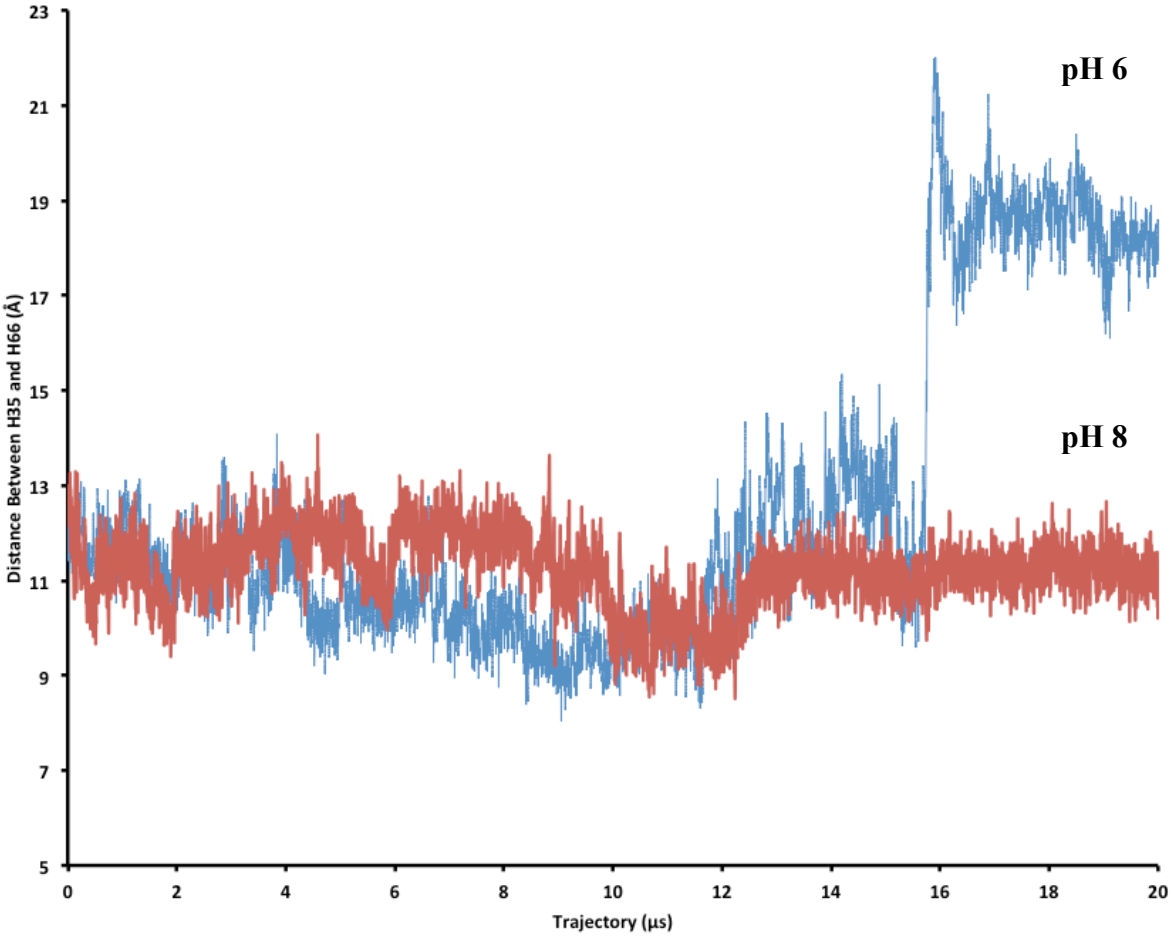


Figure 37. Residue-to-residue distance over the course of the trajectory of His35 and His66 in *Pyrocystis fusiformis* luciferase domain III with the inclusion of C-terminal disulfide bonds featuring divergence of the two residues after ~16 μs at pH 6 and no divergence of these two residues at pH 8.

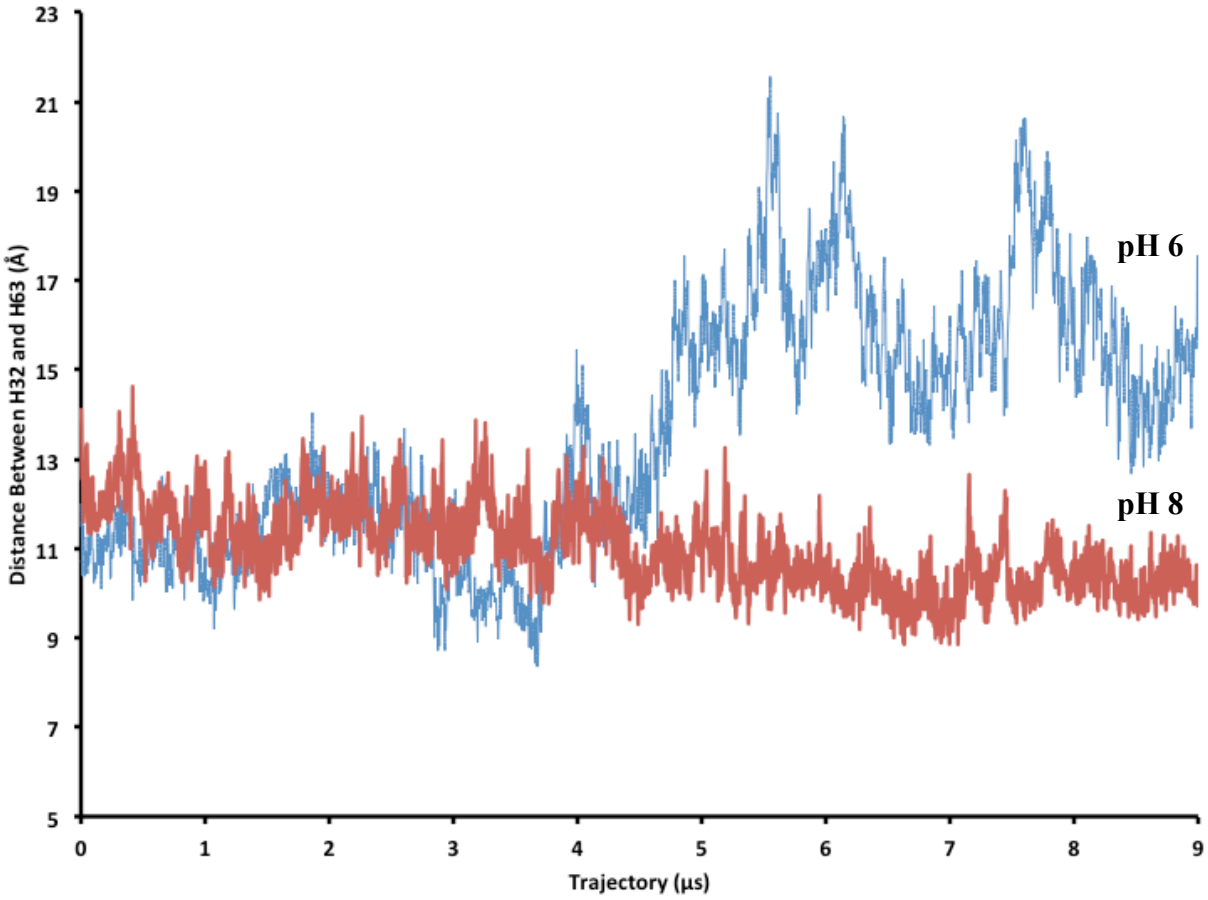


Figure 38. Residue-to-residue distance over the course of the trajectory of His32 and His64 in *Lingulodinium polyedrum* luciferase domain III featuring divergence of the two residues after ~5 µs at pH 6 and no divergence of these two residues at pH 8.

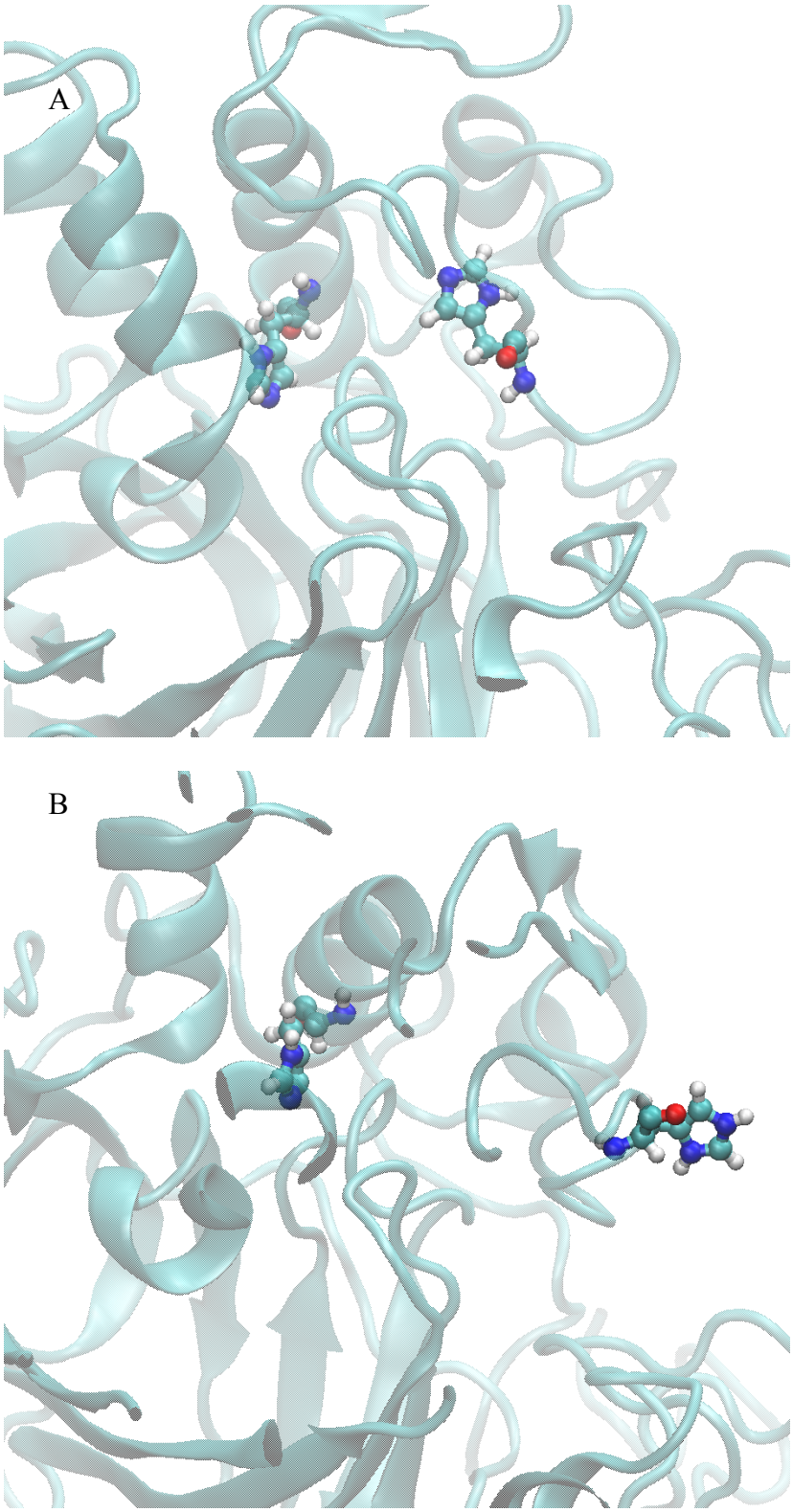


Figure 39. Visual representation of His35 and His66 before (A) and after (B) the conformational change of luciferase domain III at pH 6.

2.4 Discussion

The luciferase domain III behaved similarly with and without disulfide bonds and consistent with an enzyme regulated by pH. A structure crystalized at pH 8 indeed shifted in conformation when subjected to molecular dynamics at pH 6 and remained relatively static when the molecular dynamics simulation was maintained a pH 8. Based on our analysis of the trajectories, RMSD, and visualization of the actual conformation change that luciferase domain III exhibits at pH 6, an expansion of the volume in and around the proposed regulatory α-helical bundle is plausible for greater substrate accessibility during activation of the enzyme. Visual evidence of His66 altering its conformation locally prior to a collective motion of the α-helix that it is part of the proposed regulatory α-helical bundle provides additional theoretical support for the necessity of the histidine residues in the mechanism of regulation. It is likely that the protonation of both His66 and the nearby His35 on the neighboring α-helix induces a repulsive effect that is at least partially responsible for the local conformation change in His66 that leads to a shift of its parent α-helix away from the α-helix containing His35. Based on this theoretical evidence, we deduce that the mechanism of pH activation of dinoflagellate luciferase involves protonation of the proposed regulatory histidines and the resulting repulsive forces induces a conformational change in the α-helical bundle, allowing the substrate access to the active site.

The data collected from the CpHaMD simulations of domains I and II of *Pyrocystis fusiformis* luciferase were not quite as conclusive as those of domain III. No significant variation in the residue-to-residue distance between His36 and His67 (analogous to His35 and His66 in domain III) of domain II was observed during the course of the 2 ms simulation at pH 6. The same observation was made with the domain II simulation that was run at pH 8. However, in this case, the two residues of interest began the simulation separated by approximately 20 Å. This

was highly unexpected considering that all of the other simulations began the post-equilibrium stage with His36 and His67 between 10 Å and 12 Å apart.

The outcome with domain I was also unexpected (Figure A2S2). During the 2 ms CpHaMD simulation, an inverse trend was observed in which the distance between His36 and His67 of domain II decreased at both pH 6 and 8. Upon further analysis of this inter-residue distance, the observed decrease was delayed by ~5 μs at pH 6 versus pH 8.

The structures obtained at the end of the CpHaMD simulations were then compared to the original. Both domain I and domain II underwent significant, large-scale conformational changes (Figure A2S3 and Figure A2S4). While the structure of the β-barrel of each domain seemed to be maintained throughout the course of the simulations, the α-helices in both the N- and C-terminal regions adopted distinct conformations with no apparent pattern between domains.

It is possible that the reason for this discrepancy revolves around the homology models. The *Lingulodinium polyedrum* crystal structure used to predict the tertiary structures of all three domains of *Pyrocystis fusiformis* was in fact a domain III structure. This luciferase domain shares 82% sequence identity with domain III from *Pyrocystis fusiformis*. In contrast, domain I and domain II share only 74% and 75% sequence identity with *Lingulodinium polyedrum* domain III, respectively. Therefore, the likely lower quality of the homology models of domains I and II (compared to domain III) may explain the inconsistent behavior of these domains during the CpHaMD simulations. The application of molecular dynamics to the predicted structures of the first two domains may therefore require longer simulations to achieve a local minimum that is more representative of their actual structures at each pH. On the other hand, it is possible that the unique behaviors observed during the CpHaMD simulations may reflect intrinsic differences between the domains, which behave in concert to activate full-length luciferase at pH 6.

We have found that CpHaMD is a suitable if not preferable methodology for probing large conformation changes and collective motions in proteins that are induced by a shift in the pH environment. Thus, using CpHaMD, we were are able to confirm theoretically that domain III does undergo a conformational change during a drop in pH from 8 to 6 and that this structural change may be directly related to the pH regulation mechanism of dinoflagellate luciferase. In future work, we intend to quantify the volume change by employing isothermal-isobaric (NPT) molecular dynamics simulations and compare the results with experimental values obtained using photothermal methods [61-64].

Chapter 3

The Mechanism of Dinoflagellate Bioluminescence

Computational Investigation of the Mechanism of Dinoflagellate Bioluminescence: Chemically Initiated Electron Exchange Luminescence or Twisted Intramolecular Charge Transfer?

3.1 Background

Dinoflagellates are an important group of eukaryotic microorganisms found in fresh water and marine environments [65]. Certain species of dinoflagellates are photosynthetic and capable of bioluminescence induced by physical agitation [66]. The bioluminescence reaction, which is regulated on a circadian rhythm, involves the oxidation of an open-chain tetrapyrrolic substrate, dinoflagellate luciferin (LH_2), by the enzyme dinoflagellate luciferase (LCF) as illustrated in Figure 40 [29, 67]. Of all the major classes of luciferase (i.e., firefly, bacterial, jellyfish and dinoflagellate), least is known about the mechanism of light production by dinoflagellate luciferase [31].

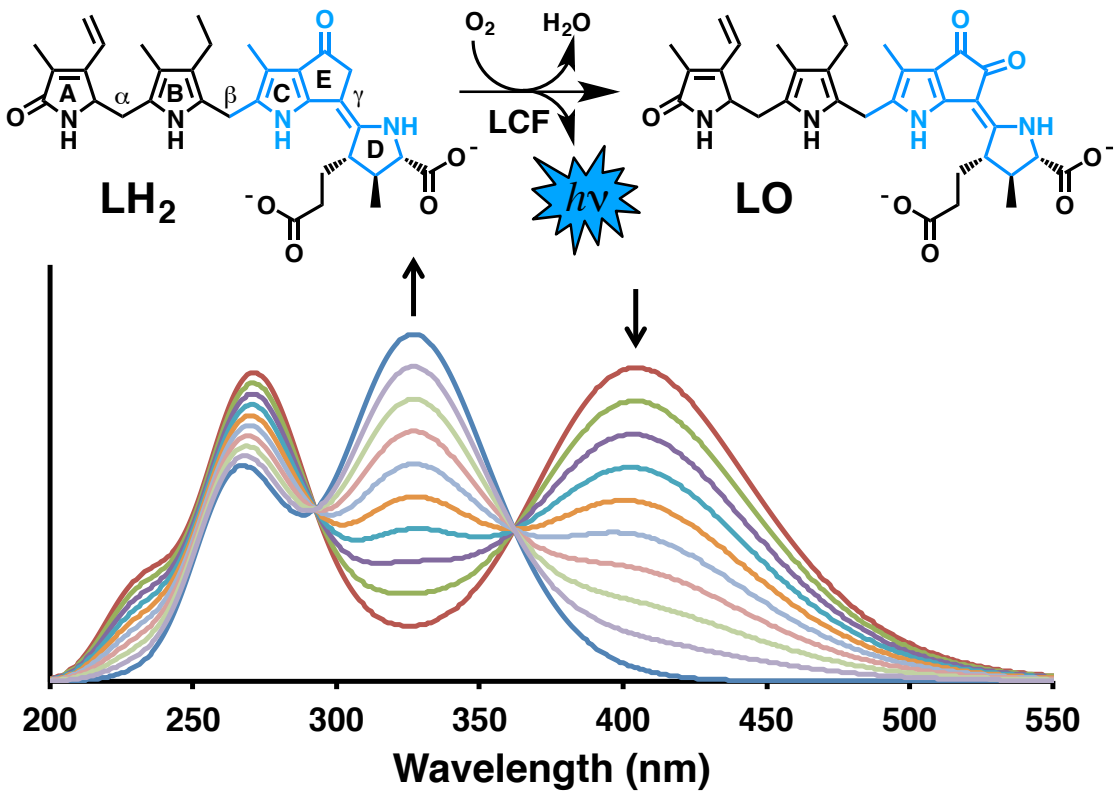


Figure 40. Calculated UV-visible spectral changes for the conversion of the chromophore (highlighted in blue) of dinoflagellate luciferin (LH_2) to oxyluciferin (LO).

3.2 Computational Methodology

In an effort to assess the feasibility of several proposed mechanisms of dinoflagellate luciferase (LCF) catalysis, the ground and first ten low-lying excited states of hydroperoxide (LHOOH), peroxy anion ($LHOO^-$), *gem*-diol (LOHOH), and *gem*-diolate ($LOHO^-$) intermediates were calculated and compared with those of dinoflagellate luciferin (LH_2) and oxyluciferin (LO) using the *Gaussian 09* software package [65]. Geometry optimizations and excited state energy calculations were performed with time-dependent long-range corrected density functional theory (TDLCDFT) using the dispersion-corrected hybrid functional ωB97X-D and Pople's diffuse polarized triple-ζ 6-311++G(d,p) basis set [66, 67]. The polarizable continuum model (PCM) of solvation was applied to all calculations using either the default parameters for water or setting the dielectric constant to 4 to emulate the interior of a protein [29, 31]. To obtain energies for the

LCF reaction coordinate, analogous calculations were performed on singlet ($^1\Delta_g$) oxygen, water, acetate, and acetic acid (the latter two molecules serving as mimics of the proposed active site base and its conjugate acid, respectively). For LCF reaction energies starting with triplet ($^3\Sigma_g^-$) oxygen, 22.5 kcal/mol was subtracted from the corresponding value calculated with singlet oxygen [68]. UV-visible spectra were generated by adding together Gaussian functions centered at the calculated transition wavelengths and scaled by the corresponding oscillator strengths. Molecular orbitals were visualized using the natural transition orbital (NTO) functionality of the Chemissian v4.0 software package.

3.3 Results

To gain insight into the mechanism of dinoflagellate bioluminescence, a computational investigation utilizing time-dependent long-range corrected density functional theory (TDLCDFT) was employed [68]. Ground and excited state energies for key intermediates in several proposed mechanisms of LCF catalysis were calculated using the dispersion-corrected hybrid functional ωB97X-D and the 6-311++G(d,p) basis set as implemented in the *Gaussian 09* software package [69]. This computational methodology was chosen because of its excellent performance in the calculation of low-lying excited states of organic molecules [70].[8]

To verify the accuracy of the chosen method while reducing computational cost, calculations were performed on the first 10 excited states of the LH_2 chromophore synthesized by Stojanovic & Kishi (Figure 40) [30, 71]. This chromophore (λ_{max} = 389 nm), which is a luminescent substrate of LCF, exhibits remarkably similar spectral properties to that of LH_2 (λ_{max} = 388 nm) [29, 30, 71].

The configuration of the double bond connecting the E and D rings in LH_2 has not been conclusively assigned; thus, calculations were performed on both isomers of the chromophore. LH_2 is a catabolite of chlorophyll that has the *Z*-configuration about this bond [35]. However, the relative stability of the two isomers of the synthesized chromophore (3:1), and comparison of their ^{13}C chemical shifts with purified LH_2, suggests the bioluminescent substrate contains the *E*-configuration [30, 71].

Similarly, the calculated UV-visible spectra of the *E*-isomers of the LH_2 (λ_{max} = 405 nm) and oxyluciferin (LO, λ_{max} = 327 nm) chromophores (Figure 40) match the experimental spectra for the enzymatic reaction (λ_{max} = 388 nm and 340 nm, respectively, in 0.2 M phosphate buffer, pH 6.3) more closely then those of the *Z*-isomers (415 nm and 325 nm, respectively) [29, 30, 71]. In these calculations, the solvent (water) was implicitly included using the polarizable continuum model (PCM) [72]. These data further support the assignment of the *E*-configuration for the C15-C16 double bond of LH_2. However, the possibility that LH_2 exists as the *Z*-isomer *in vivo* and isomerizes to the more stable *E*-form during purification cannot be ruled out. The data also provide a benchmark for the anticipated errors in the calculated excitation wavelengths (± 15 nm) for intermediates in the proposed mechanisms of LCF (the microscopic reverse of which serves as a mimic of the bioluminescence emission process).

It has been proposed that the LCF catalytic cycle involves deprotonation of the E ring of LH_2 by an active site base (e.g., E1105 in the structure of LCF from *Lingulodinium polyedrum*) to generate an enolate anion that reacts with molecular oxygen to form a hydroperoxide intermediate [28-30, 71]. Formation of this intermediate may proceed through an excited state hydroperoxide, which could serve as the bioluminophore by relaxing with radiative emission of blue light at 474 nm. However, several alternative reaction mechanisms can be envisioned

wherein the light emitting species is an excited state peroxy anion, *gem*-diol, or *gem*-diolate intermediate (Figure 41).

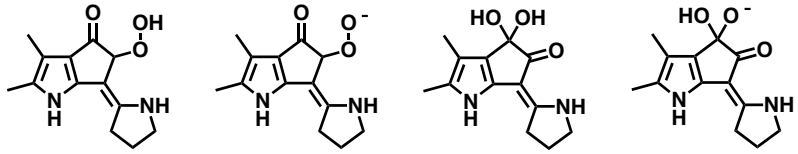


Figure 41. Chromophores of possible intermediates in the dinoflagellate luciferase (LCF) catalytic cycle, whose excited states may serve as the bioluminophore.

To ascertain the feasibility of each of these mechanisms and shed light on the structure of the potential bioluminophore, the energies of the ground and low-lying excited states for these intermediates were calculated as described above and compared with the ground state energies of the LH_2 substrate (+ O_2) and LO product (+ H_2O). The energy of each species along the reaction coordinate also included acetate/acetic acid as a mimic of the putative active site acid-base residue as appropriate. The effects of surroundings (aqueous vs. proteinaceous environment, the latter modeled using a dielectric constant of 4), configuration about the C15-C16 double bond (*E* vs. *Z*), and stereochemistry/conformation of the O_2-derived functional groups (*R* vs. *S*/eclipsed vs. staggered, as applicable) on the excitation wavelengths of the proposed intermediates were also examined (see Supporting Information for full computational results).

In the case of the peroxy anion intermediate, the calculated excitation spectrum typically contains a strong transition between 364 nm – 397 nm and a weak transition between 548 nm – 610 nm, each of which lie on average ~100 nm from the bioluminescence emission maximum of LCF (474 nm). In certain conformations of the *E*-isomer of the peroxy anion intermediate, the pyrrole nitrogen of the D ring donates a hydrogen bond to the terminal oxygen of the peroxide moiety, which increases the intensity of the long wavelength transition and shifts it to between

501 – 528 nm. This highlights the potential of the enzyme's active site to tune the bioluminescence emission wavelength/intensity using hydrogen bonding interactions. Moreover, although formation of the ground state peroxy anion intermediate from LH_2 is on average energetically favorable (-39.6 ± 3.5 kcal/mol with singlet O_2 or -17.1 ± 3.5 kcal/mol with triplet O_2), formation of the excited state peroxy anion is not (11.6 ± 2.0 kcal/mol and 34.1 ± 2.0 kcal/mol, respectively). Given that the catalytic cycle is likely to begin with molecular oxygen in its triplet ground state, the excited state peroxy anion may be thermodynamically inaccessible, and is thus unlikely to serve as the bioluminophore.

The blue shift in excitation wavelength of the peroxy anion intermediate towards the bioluminescence emission maximum induced by intramolecular hydrogen bonding suggests that an excited state hydroperoxide intermediate may be a good candidate for the bioluminophore. Indeed, the excitation spectrum of the hydroperoxide intermediate contains a strong transition between 438 – 482 nm. In particular, the excitation wavelength of the *Z*-isomer containing the *R*-configuration of the hydroperoxide moiety in proteinaceous medium is 475 nm, nearly identical to the bioluminescence emission maximum. In terms of energetics, formation of the ground state hydroperoxide intermediate is more favorable than that of the peroxy anion (-53.8 ± 2.8 kcal/mol with singlet O_2 or -31.3 ± 2.8 kcal/mol with triplet O_2). However, the excited state energies for the hydroperoxide intermediate are similar to those of the peroxy anion (8.6 ± 2.2 kcal/mol and 31.1 ± 2.2 kcal/mol, respectively), and thus similar concerns apply to the feasibility of its formation.

The excitation spectrum of the *E*-isomer of the *gem*-diol intermediate is similar to the hydrogen bonded peroxy anion in that it contains a long wavelength transition between 507 – 512 nm, which is on average ~35 nm longer than the bioluminescence emission maximum. In

contrast, the transition closest to the bioluminescence emission maximum for the *Z*-isomer of the *gem*-diol intermediate was distantly located between 355 – 359 nm. For both the *E*- and *Z*-isomers of the *gem*-diol intermediate, the energy of formation of the ground state (-93.6 ± 0.3 kcal/mol and -64.2 ± 0.9 kcal/mol with triplet oxygen, respectively) and the excited state (0.4 ± 0.3 kcal/mol and 16.1 ± 0.9 kcal/mol, respectively) is much more favorable then that of the peroxy anion or hydroperoxide intermediate. Interestingly, the large discrepancy in the calculated ground and excited state energies of the two isomers of can be attributed to the fact that, during the geometry optimization, the *E*-isomer maintained its planar conformation, while an unexpected rotation occurred about the C15-C16 bond of the *Z*-isomer such that the E and D rings became perpendicular to one another.

Similar planar and twisted geometries were also observed with the optimized structures of the *E*- and *Z*-isomers of the *gem*-diolate intermediate, respectively. The excitation spectrum of the *E*-isomer contains a strong transition between 433 – 446 nm, which is on average ~35 nm shorter than the bioluminescence emission maximum, and both the ground and excited states are stable relative to LH_2 and triplet oxygen (-90.8 ± 0.4 kcal/mol and -25.6 ± 0.7 kcal/mol, respectively). Furthermore, the excitation wavelengths of the *E*-isomers of the *gem*-diol and *gem*-diolate intermediates symmetrically bracket the bioluminescence emission maximum, suggesting that a hydrogen bonded excited state *E*-*gem*-diol(ate) intermediate is a strong candidate for the bioluminophore. The calculations also support the *Z*-isomer of the *gem*-diolate as an intermediate in the LCF catalytic cycle, due to the fact that its calculated excitation spectrum contains a transition between 457 – 486 nm, with an average (473 nm) that is closest to the bioluminescence emission maximum of all the intermediates considered. In addition, the ground and excited states of the *Z*-*gem*-diolate intermediate are both likely to be thermodynamically

accessible using triplet oxygen as the oxidant (-51.1 ± 0.7 kcal/mol and 9.4 ± 0.8 kcal/mol, respectively).

The data are therefore most consistent with a mechanism of LCF catalysis in which an excited state *gem*-diol(ate) intermediate serves as the bioluminophore. If the LH_2 substrate contains the *E*-configuration about the C15-C16 double bond, planarity of the E and D rings is likely maintained throughout the catalytic cycle and the bioluminophore is predicted to form a strong hydrogen bond with an active site residue. Alternatively, if LH_2 has the *Z*-configuration, the reaction likely involves a rotation about the C15-C16 bond and the formation of a twisted excited state *gem*-diolate intermediate.

The molecular orbitals (MOs) participating in the relevant transitions for the *E*-isomer of the *gem*-diol and the *Z*-isomer of the *R*-*gem*-diolate intermediates in aqueous solution are shown in Figure 42. In the former case, an electron is promoted from the highest occupied molecular orbital (HOMO) of the delocalized π-system of the chromophore to an orbital centered predominantly on the E ring carbonyl. The MOs for the analogous transition in the *E*-*gem*-diolate are similar (see Supporting Information). Thus, if a hydrogen bonded excited state *E*-*gem*-diol(ate) intermediate is to serve as the bioluminophore, an open-shell singlet containing an unpaired electron in each of these orbitals must be formed during the catalytic cycle, and photon emission occurs with intramolecular electron transfer from the carbonyl to the conjugated π-system.

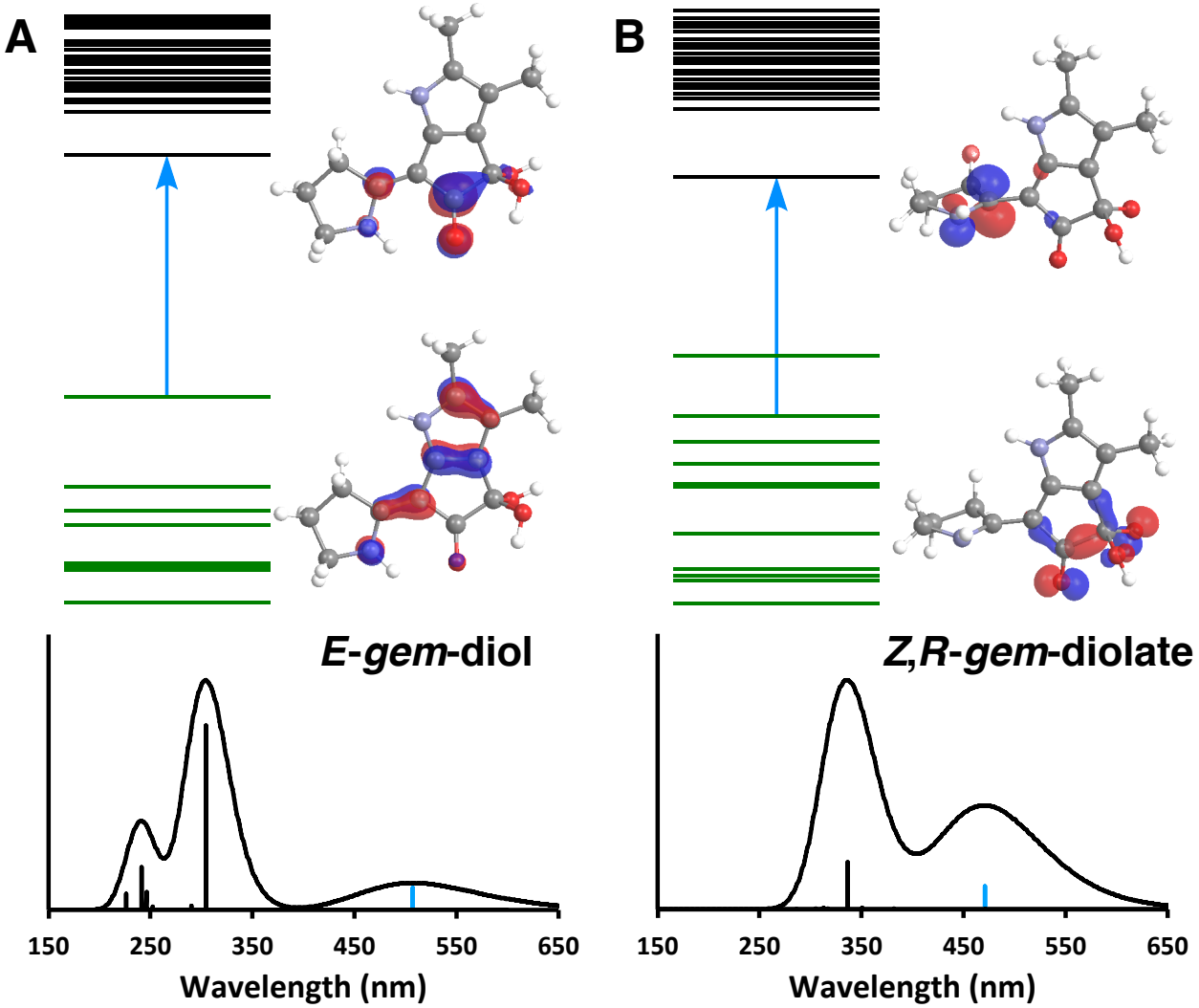


Figure 42. Molecular orbitals participating in (A) the 507 nm transition of the *E-gem*-diol intermediate and (B) the 471 nm transition of the *Z,R-gem*-diolate intermediate in aqueous solution.

In contrast, the transition for the *Z,R-gem*-diolate intermediate involves promotion of an electron from the second-highest molecular orbital (HOMO-1) centered on the E ring to the lowest unoccupied molecular orbital (LUMO) residing on the orthogonal D ring. Therefore, if an excited state *Z-gem*-diolate is the bioluminophore, an open-shell singlet with one unpaired

electron on each ring of the twisted intermediate is formed during the reaction and light is emitted with charge transfer from the D to the E ring.

3.4 Discussion

Mechanisms of LCF catalysis consistent with the above information are shown in Figure 43. Starting from either isomer of LH_2, the catalytic cycles involve deprotonation of the E ring by the active site base to generate an enolate anion that reacts with molecular oxygen to form a dioxetanolate intermediate. At this point the two mechanisms diverge. For the *E*-isomer, dioxetanolate ring-opening is initiated by electron transfer from the conjugated π-system. Subsequent deprotonation of the remaining E ring hydrogen by the resulting *gem*-diolate and back electron transfer leads to photon emission. Such a mechanism describes a chemically initiated electron-exchange luminescence (CIEEL) process, which has been invoked to explain all other known bioluminescent reactions, most notably that catalyzed by firefly luciferase [31, 73, 74].

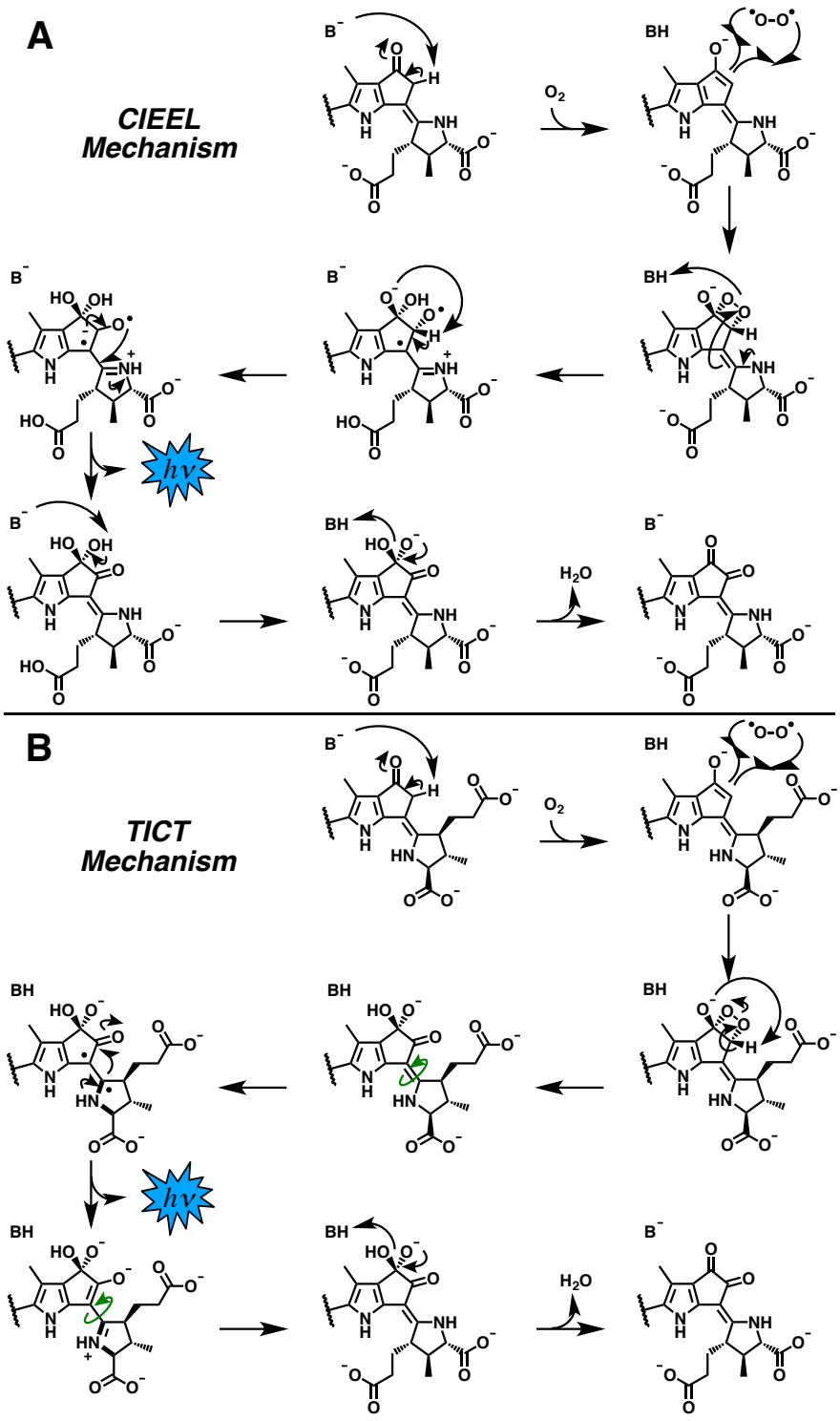


Figure 43. Proposed mechanisms of dinoflagellate luciferase catalysis involving (A) chemically initiated electron exchange luminescence (CIEEL), and (B) twisted intramolecular charge transfer (TICT).

For the *Z*-isomer, deprotonation of the E ring leads to O-O bond cleavage and induces a rotation about the C15-C16 bond, generating the open-shell singlet excited state. D-to-E ring charge transfer is then accompanied by photon emission. This mechanism describes the formation of a twisted intramolecular charge transfer (TICT) state. TICT states have been implicated in a number of photochemical processes, but have yet to be observed in a biological system [75].

In summary, TDLCDFT was used to calculate the ground and excited state energies of luminophores from several proposed mechanisms of LCF catalysis. Comparison with the experimental bioluminescence emission wavelength and an evaluation of the thermodynamically feasibility of forming each species favors a *gem*-diol(ate) intermediate over a peroxy anion or hydroperoxide as the bioluminophore. If the LCF catalytic cycle begins with the *E*-isomer of LH_2, the reaction is likely to involve a CIEEL mechanism and a hydrogen bonded bioluminophore, since the emission wavelengths of the excited state *E*-*gem*-diol and -diolate intermediates symmetrically bracket the bioluminescence emission maximum. However, if LH_2 has the *Z*-configuration, an excited state *Z*-*gem*-diolate intermediate (whose average emission wavelength most closely matches the bioluminescence emission maximum) is likely to serve as the bioluminophore and LCF would catalyze a biologically unprecedented TICT reaction.

Conclusion

Dinoflagellate bioluminescence remains one of the most understudied bioluminescence systems, despite the uniqueness of its components and regulatory mechanism. Pragmatically, elucidating unknown aspects of dinoflagellate luciferin and luciferase, of which there are many, advances the knowledge required to implement this versatile bioluminescence system as both a cellular imaging agent and reporter gene [25]. The vast majority of cellular imaging agents used in conjunction with fluorescence and laser-scanning confocal microscopes rely on fluorescent proteins that require excitation by an external light source [1, 76]. Luciferases represent a departure from this imaging paradigm wherein luciferase or a fusion protein of interest can be visualized on a regular dark field microscope [5, 77]. However, the use of dinoflagellate luciferase in such an application requires readily available access to its substrate, dinoflagellate luciferin. Thus, elucidating the dinoflagellate luciferin biosynthetic pathway is instrumental in this effort. Additionally, understanding the mechanism of the dinoflagellate bioluminescence reaction will allow us to probe the possibilities of varying the wavelength of energy emitted by the bioluminophore, thereby increasing the colorimetric diversity of future cellular imaging agents [77-79]. To this end, the work presented in the preceding chapters represent an important step forward in our understanding of dinoflagellate bioluminescence.

PART II

Heterologous Expression of *Holo* Methyl-coenzyme M Reductase (MCR)

Introduction

The biological production of methane is due to methanogenic archaea originating from a period of ancient earth. The key enzyme utilized by these and related organisms in both methanogenesis and anaerobic methane consumption is methyl-coenzyme M reductase (MCR) [80]. Heterologous expression of active *holo* MCR has never been achieved, in large part due to the lack of knowledge regarding the enzymes required for the biosynthesis of coenzyme F430, an essential cofactor of MCR [81, 82].

Methanogenesis can proceed using a wide range of electron acceptors, including carbon dioxide, acetate, methanol, and methylamines, that are ultimately reduced to methane in the terminal step of the pathway catalyzed by MCR [83]. Coenzyme B and methyl-coenzyme M in the presence of MCR undergo a reaction in which a heterodisulfide product is formed and methane is released (Figure 44) [84].

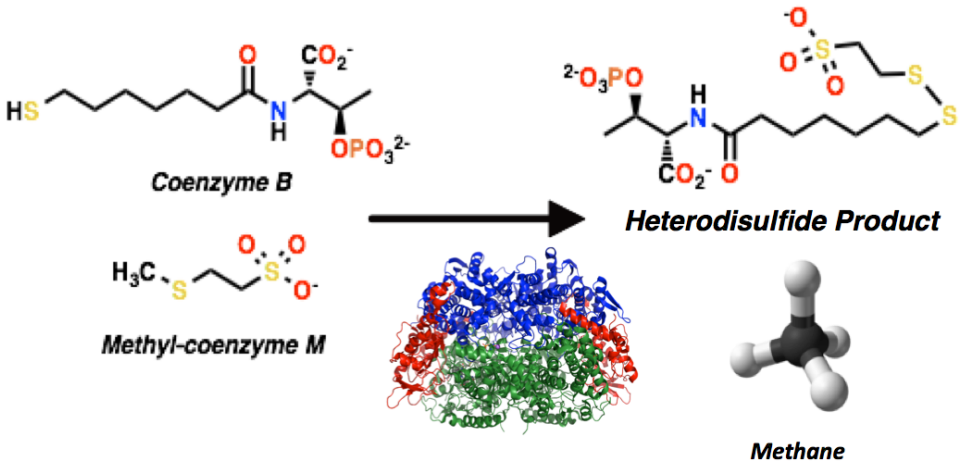


Figure 44. Final step in methanogensis featuring the MCR-catalyzed conversion of coenzyme B and methyl-coenzyme M into a heterodisulfide and methane.

MCR is composed of an $\alpha_2\beta_2\gamma_2$ heterohexamer of approximately 270 kDa [85]. Two active sites are housed within each symmetrical half of the enzyme and contain the essential cofactor of methanogenesis, coenzyme F430 [86]. Interestingly, the active site of MCR contains five novel post-translational modifications (PTM) whose origin and purpose are both unknown. Of the five PTMs, thioglycine, 5-(*S*)-methylarginine, 1-*N*-methylhistidine, *S*-methylcysteine, and 2-(*S*)-methylglutamine, four are conserved throughout the majority of methanogen species (Figure 45) [87]. Interestingly, *in vivo* labeling with L-(methyl-D_3)-methionine demonstrated that the methyl groups in four of the modified residues are likely introduced by *S*-adenosyl-L-methionine-dependent methyltransferases [87].

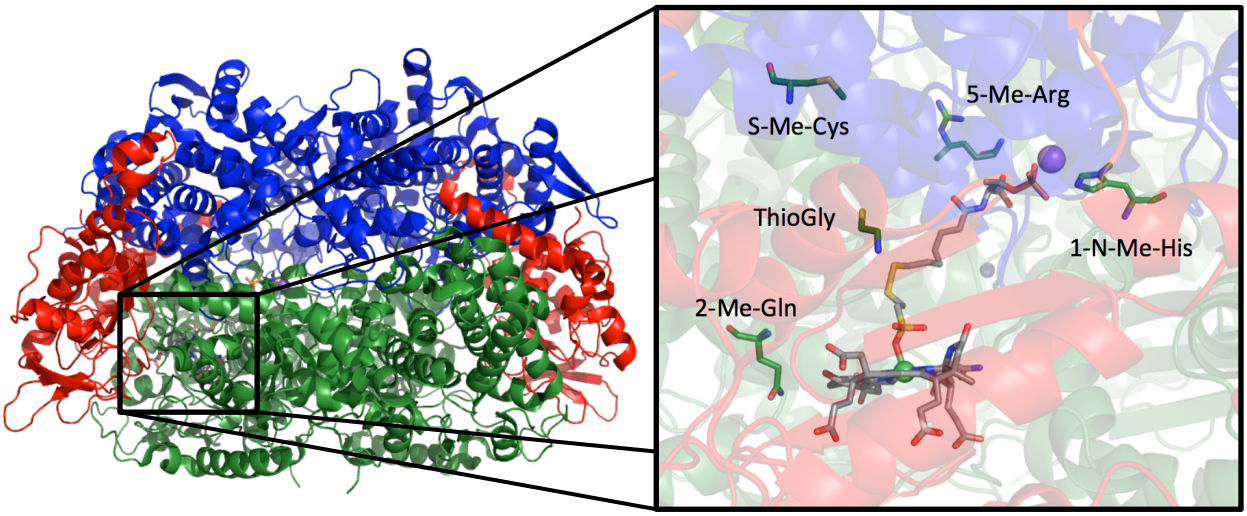
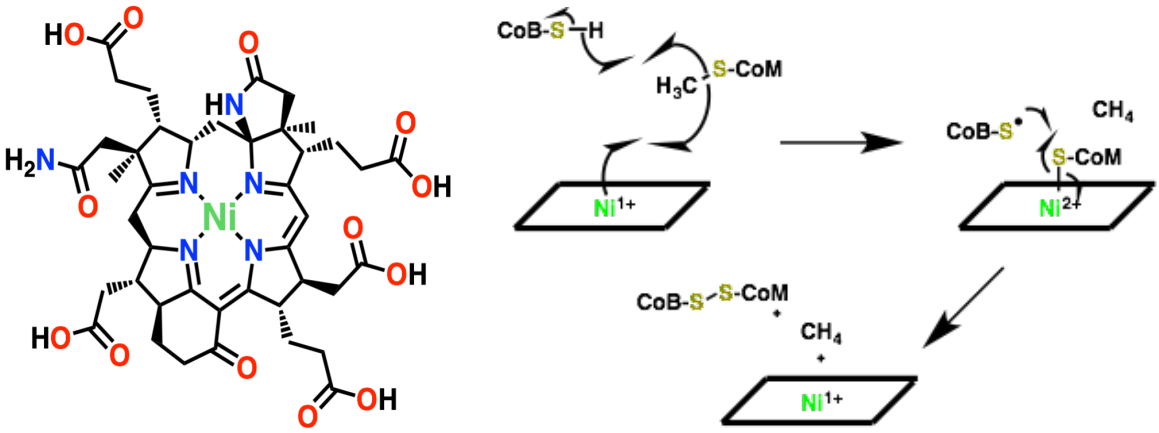


Figure 45. The MCR $\alpha_2\beta_2\gamma_2$ complex featuring an active site housing coenzyme F430 and the thioglycine, 5-(*S*)-methylarginine, 1-*N*-methylhistidine, *S*-methylcysteine, and 2-(*S*)-methylglutamine PTMs (PDB ID: 1HBM).

It is thought that an activating complex is required to reduce the nickel center of coenzyme F430 to Ni^{1+} in order for MCR to become active [88]. This complex consist of several components, including a heterodisulfide reductase that contains an electron bifurcation system that directs electrons to regenerating coenzyme M and coenzyme B while simultaneously sending electrons to the A3a multi-enzyme complex that ultimately reduces coenzyme F430 with the help of the A2 adenosine triphosphate (ATP)-binding protein [88]. Once MCR is activated, the reaction is postulated to proceed by way of a radical mechanism that is highly exergonic, exhibiting a biochemical standard Gibbs free energy of -30 kJ/mol (Figure 46) [84, 86, 89].



Figurre 46. Coenzyme F430 (left panel) and proposed radical mechanism of methane formation (right panel).

Until very recently, the biosynthetic pathway of coenzyme F430 was poorly understood, but Mansoorabadi and coworkers used a comparative genomics investigation to identify the coenzyme F430 biosynthesis (*cfb*) genes and characterized the functions of the encoded enzymes.

As established by our group and others, sirohydrochlorin is the last common precursor in the biosynthesis of all tetrapyrroles in methanogens. The biosynthesis of sirohydrochlorin from the ubiquitous tetrapyrrolic intermediate uroporphyrinogen III in *Methanosarcina acetivorans*

C2A was demonstrated using heterologously expressed SirA, a uroporphyrinogen III methyltransferase, and SirC, a precorrin 2 dehydrogenase. It was established in this study that most divalent metal ions (e.g., Fe^{2+}, Co^{2+}, Cu^{2+}, and Zn^{2+}) could spontaneously chelate sirohydrochlorin except for Ni^{2+}, the metal center of coenzyme F430. No Ni-specific chelatase has ever been reported, but present in the *cfb* cluster is a cobaltochelatase (*cbiX*) homolog. CbiX is one of only two class II chelatases present in the genome of *M. acetivorans* and was previously thought to only catalyze the cobalt chelation of sirohydrochlorin in cobalamin biosynthesis. Its ability to catalyze Ni^{2+} chelation was confirmed [90].

Additionally, cobyrinic acid *a*,*c*-diamide synthetase, CbiA, is involved in cobalamin biosynthesis, for which a homolog is also present in the *cfb* gene cluster. In cobalamin biosynthesis, CbiA catalyzes amidation reactions on the carboxylates of the *a*- and *c*-acetate side chains of cobyrinic acid with the assistance of glutamine and ATP. Identical carboxylates undergo amidation reactions in the biosynthesis of coenzyme F430. Thus, a CbiA homolog is a suitable enzyme candidate to carry out this step of F430 biosynthesis to produce a novel intermediate, Ni-sirohydrochlorin *a*,*c*-diamide, and this was demonstrated to be the case [90].

A six-electron reduction is required to achieve the highly reduced state of coenzyme F430. A nitrogenase homolog, *nifD*/*nifH*, is present in the *cfb* gene cluster. Nitrogenase (NifHDK) and a homolog from the chlorophyll biosynthetic pathway, the light-independent protochlorophyllide oxidoreductase (DPOR), catalyze a triple bond reduction in diatomic nitrogen to ammonia in a key step of nitrogen fixation and double bond reduction of protochlorophyllide *a* to chlorophyllide *a*, respectively. The NifD homolog, also known as methanogenesis marker 13 (Mm13), was thought to perform similar reductions in the

biosynthesis of F430 yielding a previously discovered intermediate, 15,17^3-seco-F430-17^3-acid and this was also confirmed [90].

The final step in the biosynthesis of coenzyme F430 requires an ATP-dependent intra-molecular cyclization of the *g*-propionate side chain to form the F carbo-exocyclic ring. Mur-ligases, a homolog of which is present in the *cfb* gene cluster, use ATP to activate a carboxylate in a reaction that catalyzes peptide bond formation non-ribsomally in the synthesis of peptidoglycan of bacterial cell wall. The carboxylate of the *g*-propionate side chain can be activated in a similar fashion by the Mur-ligase homolog to form the F carbo-exocyclic ring. This was verified, producing F430 in an *in vitro* system for the first time, albeit inefficiently. Coincidentally, Eduardus (Evert) Duin and coworkers overexpressed MCR in a methanogen and discovered that McrD, a member of the *mcr* gene cluster and a protein of unknown function, was co-purified bound to *apo* MCR (personal communication). Speculating that McrD was a chaperone protein that could bind and deliver F430 to MCR, it was included in the *in vitro* F430 biosynthesis system and was found to significantly improve the coenzyme yield (presumably via the alleviation of product inhibition of the Mur-ligase homolog).

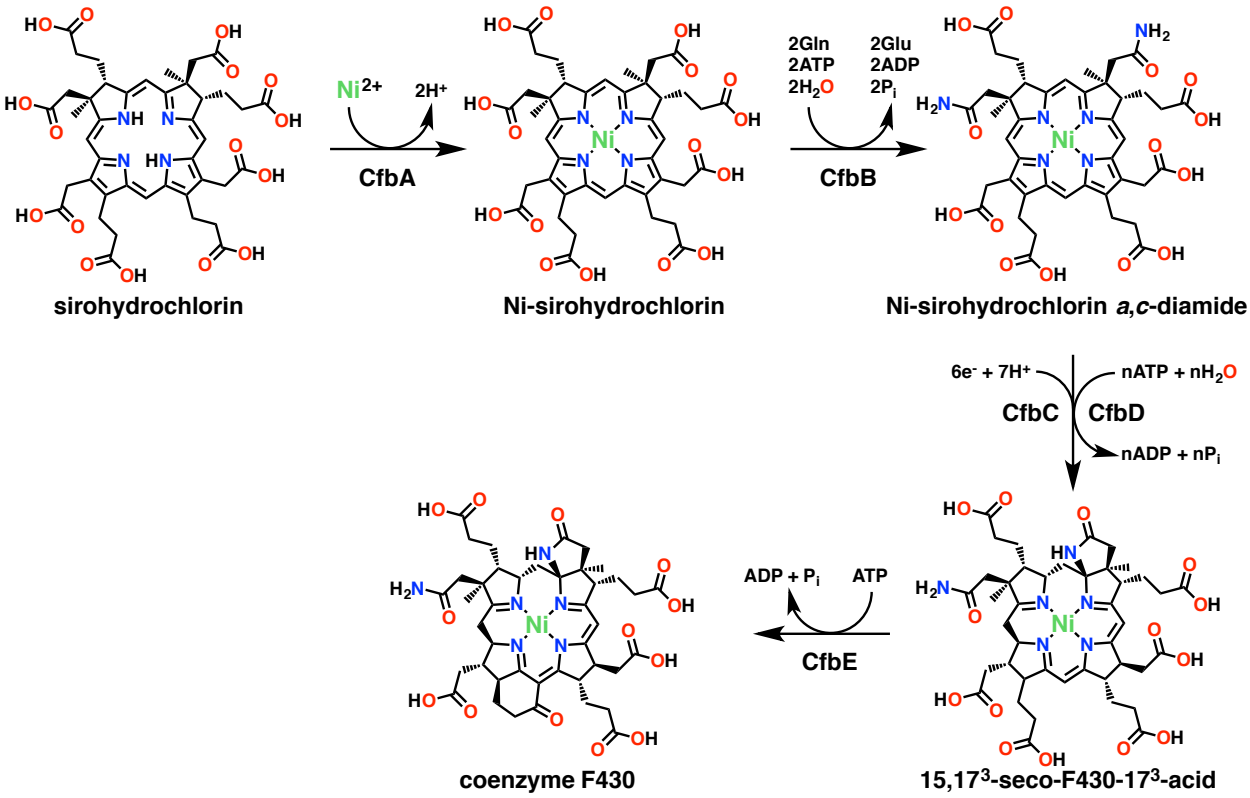


Figure 47. The biosynthetic pathway of coenzyme F430.

While a proof of concept has been established for the *in vitro* biosynthesis of F430, *in vivo* formation of coenzyme F430 will most likely be necessary for the construction of *holo* MCR. Additionally, proper assembly of the MCR subunits may require the incorporation of the active site PTMs. This dissertation aims to further advance the ultimate goal of constructing *holo* MCR by utilizing established molecular biology techniques in conjunction with novel methodologies to both establish a proof of concept for the *in vivo* biosynthesis of coenzyme F430 and to elucidate the origin and function of the PTMs present in the MCR active site.

Chapter 4

Coexpression of Coenzyme F430 Biosynthetic Genes

The *in vivo* Biosynthesis of Coenzyme F430 through Heterologous Expression of Multiple Plasmids in *Escherichia coli*

4.1 Background

Methanogenic archea are responsible for the production of vast amounts of methane released in the atmosphere [91]. These methanogens produce methane by metabolizing carbon dioxide and other C1 compounds in a conserved methanogenesis pathway that ends with methyl-coenzyme M reductase (MCR) catalyzing a reaction that combines methyl-coenzyme M and coenzyme B into a heterodisulfide that also yields methane [92]. MCR is a large 270 kDa complex consisting of an $\alpha_2\beta_2\gamma_2$ heterohexamer that contains two active sites [82]. Tightly bound with in each active site is a unique and essential Ni prosthetic group known as coenzyme F430 [86].

Recently, there has been great interest in engineering a non-methanogen host, such as *Escherichia coli*, to express active *holo* MCR for use in methane bioconversion strategies. Natural gas is abundant in shale formations, which are often located in remote areas. Thus, transporting natural gas can be both expensive and hazardous. In fact, excess methane is often flared or vented into the atmosphere and the practice has increased in the past decade [93]. Therefore, interest in MCR revolves around the fact that it has been demonstrated that MCR from both methanogens and anaerobic methanotrophs are capable of activating methane for oxidation and conversion into liquid fuel. This eliminates the need to transport the more costly and dangerous natural gas [94-99].

The genes and corresponding enzymes involved in coenzyme F430 biosynthesis (*cfb*) were recently discovered by our group [90]. Previous attempts to synthesize F430 *in vitro* have been successful, but attempting the biosynthesis *in vivo* in a non-methanogen host presents new challenges. For instance, *in vitro*, excess reducing agents (i.e., sodium dithionite), cofactors, and ATP can be supplemented in the reaction. In an *in vivo* system, there is no guarantee that the host can uptake additional supplements in the event that intracellular concentrations are insufficient. Moreover, the key step in the biosynthesis of F430, which is catalyzed by the CfbCD complex, requires reducing equivalents supplied by an external reductase (e.g., ferredoxin or flavodoxin), which may not be compatible in a heterologous host such as *E. coli*. If the endogenase reductase systems of *E. coli* are not compatible, methanogens possess several homologs that are possible candidates for the activation of the CfbCD complex.

In this study, the *cfb* genes plus *sirA* (uroporphyrinogen III methyltransferase) and *sirC* (precorrin 2 dehydrogenase) were cloned into a series of compatible coexpression plasmids, pETDuet, pRSFDuet, pACYCDuet and pCDFDuet. These plasmids were then sequentially co-transformed into an *E. coli* expression host and subsequently induced in an attempt to biosynthesize F430 *in vivo*.

4.2 Methods

Plasmid construction

The *cfbABCDE* and *sirAC* genes were amplified by polymerase chain reaction from the genome of *Methanosarcina acetivorans* C2A purchased from DSMZ in Germany (DSM-2834). Primers were synthesized by Sigma-Aldrich and their sequences are provided in the Supplemental Information. Phusion High-Fidelity DNA Polymerase (New England Biolabs) was utilized for all PCR reactions in accordance with the manufacturer's protocol.

All plasmids were purchased from Novagen EMD Millipore of Germany. The *sirC* gene was ligated into the first multiple cloning site of pETDuet (Figure 48). The *cfbA* and *cfbE* genes were ligated into the first and second multiple cloning sites of pCDFDuet, respectively (Figure 49). The *sirA* and *cfbB* genes were ligated in the first and second multiple cloning sites of pACYCDuet, respectively (Figure 50). The *cfbD* and *cfbC* were ligated in the first and second multiple cloning site of pRSFDuet, repectively (Figure 51).

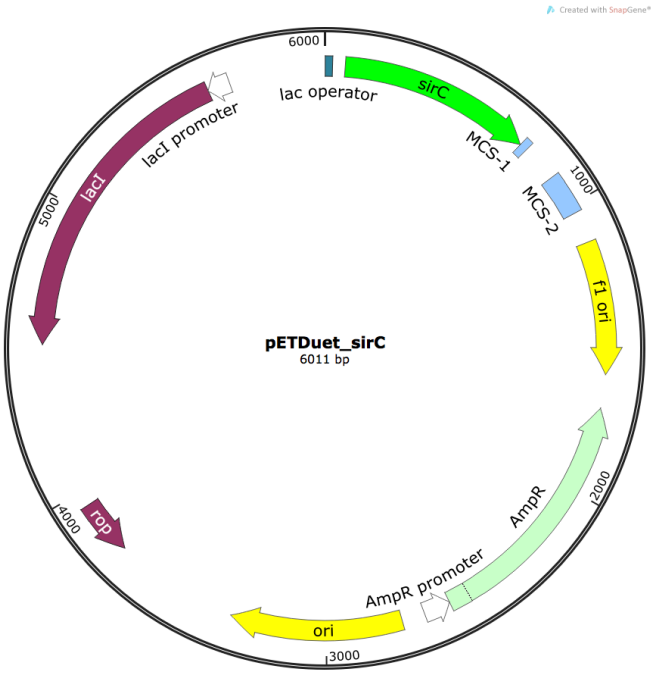


Figure 48. Ampicillin resistant pETDuet expression vector containing *sirC*.

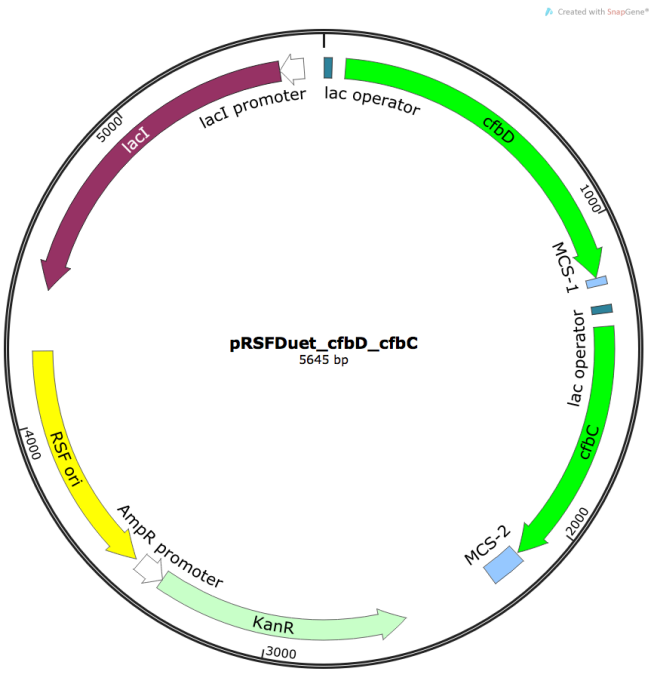


Figure 49. Kanamycin resistant pRSFDuet expression vector containing *cfbD* and *cfbC*.

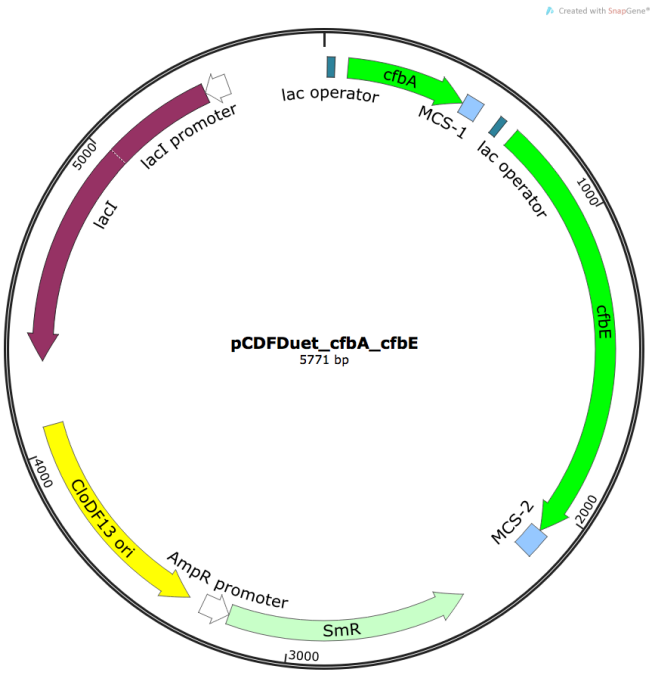


Figure 50. Spectinomycin resistant pCDFDuet expression vector containing *cfbA* and *cfbE*.

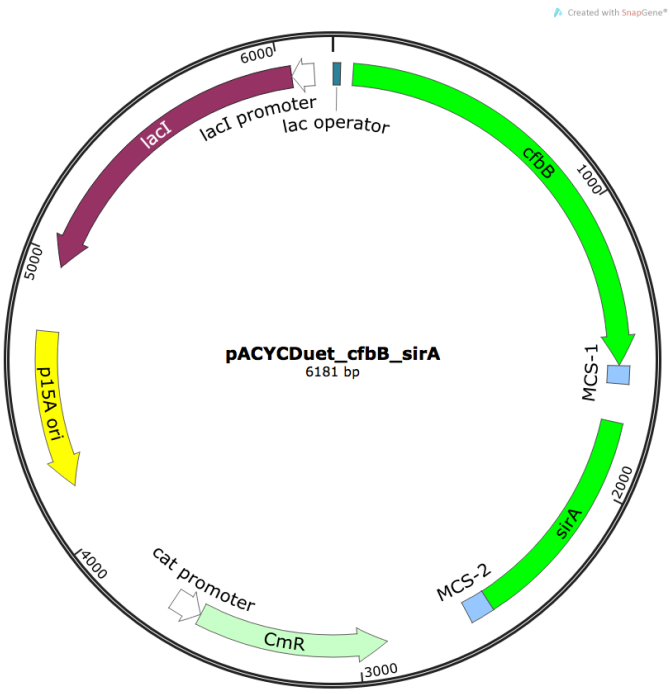


Figure 51. Chloramphenicol resistant pACYCDuet expression vector with *cfbB* and *sirA*.

Coexpression of Coenzyme F430 Biosynthesis Genes and F430 Extraction

Constructed plasmids were sequenced and verified to be correct before being sequentially transformed into *E. coli* BL21 (DE3) cells (New England Labs, Ipswitch, MA) using selection plates supplemented with the appropriate antibiotics. Once all four plasmids have been transformed into a single BL21 (DE3) cell, it will be resistant to four different antibiotics: ampicillin, kanamycin, spectinomycin, and chloramphenicol.

Terrific broth (TB) was prepared as the expression media and was supplemented with a working concentration of nickel chloride (200 µM final concentration), ferrous sulfate (200 µM final concentration), cobalamin (2 µM final concentration), riboflavin (50 µM final concentration), methionine (50 µM final concentration), cysteine (200 µM final concentration), histidine (100 µM final concentration), glutamine (100 µM final concentration), aminolevulinic acid (10 µM final concentration), porphobilinogen (10 µM final concentration), ampicillin (100

μg/mL final concentration), kanamycin (50 μg/mL final concentration), spectinomycin (100 μg/mL final concentration), and chloramphenicol (85 μg/mL final concentration). Expression of the coenzyme F430 biosynthesis genes was carried out at 37 °C until the OD_{600} = 0.6, after which the temperature was lowered to 18 °C. The expression cultures were then induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.4 μM and incubated for 16 h with shaking.

At the end of the incubation period, the cells were harvested by centrifugation of the expression culture at 8,000 rpm for 30 minutes. The cell pellet was then resuspended in 0.1 M Tris buffer, pH 8 for boiling and sonication to release intracellular tetrapyrroles in an anaerobic environment. Anaerobic conditions were maintained while the suspension was then subjected to ultracentrifugation at 20,000 ×g to pellet the cell debris. The supernatant was retrieved while the pellet was discarded. The supernatant was flash frozen with liquid nitrogen and subjected to lyophilization. The resulting powder was dissolved in a small amount of nanopure water and extracted with methanol. Precipitates were pelleted by centrifugation at 8,000 rpm and the methanol supernatant was retrieved and concentrated under a stream of nitrogen. The sample was then analyzed by high-performance liquid chromatography (HPLC) and liquid chromatography mass spectrometry (LC-MS).

HPLC and LC-MS Analysis

HPLC analysis was performed on an Agilent 1260 Infinity Quaternary LC System equipped with a Diode Array Detector (DAD) VL+ and an Agilent Poroshell 120 EC-C18 (4.6 × 150 mm, 2.7 μm) reverse phase column. The Agilent OpenLAB ChemStation Edition software was used for data analysis. The chromatographic method utilized for characterization of the coenzyme F430 biosynthetic reactions consisted of the following gradient of water (solvent A)

and acetonitrile (solvent B), each containing 0.5% formic acid: 0% B for 2 min, 0-20% B over 3 min, 20% B for 5 min, 20-25% B over 5 min, 25% B for 5 min, 25-30% B over 5 min, and 30-100% B over 5 min. The flow rate was 1.0 mL/min and the chromatogram was acquired with detection at 400 nm.

LC-MS analysis was performed on a Waters Acquity UPLC/Q-TOF Premier Mass Spectrometer equipped with the identical Agilent Poroshell 120 EC-C18 column. The Waters MassLynx MS software was used for data analysis. The LC method consisted of the same solvent system and gradient as described above. The electrospray ionization (ESI) mass detector was configured to positive ion mode with scanning between 0-1100 *m*/*z*. The inline Tunable UV (TUV) detector was set to 400 nm to match the peaks observed in the mass chromatograms to those observed by HPLC analysis.

4.3 Results

HPLC Analysis of Cell-Free Extracts

Based on initial HPLC analysis of the cell-free extracts of *E. coli* expressing the pETDuet-*sirC*, pRSF-*cfbC*-*cfbD*, and pACYC-*cfbB*-*sirA* plasmids (without pCDF-*cfbA*-*cfbE*), sufficient quantities of sirohydrochlorin could be achieved *in vivo*. This was evident by the presence of a chromatographic peak in high abundance with identical retention time and UV-visible spectrum to literature reports of sirohydrochlorin (Figure 52).

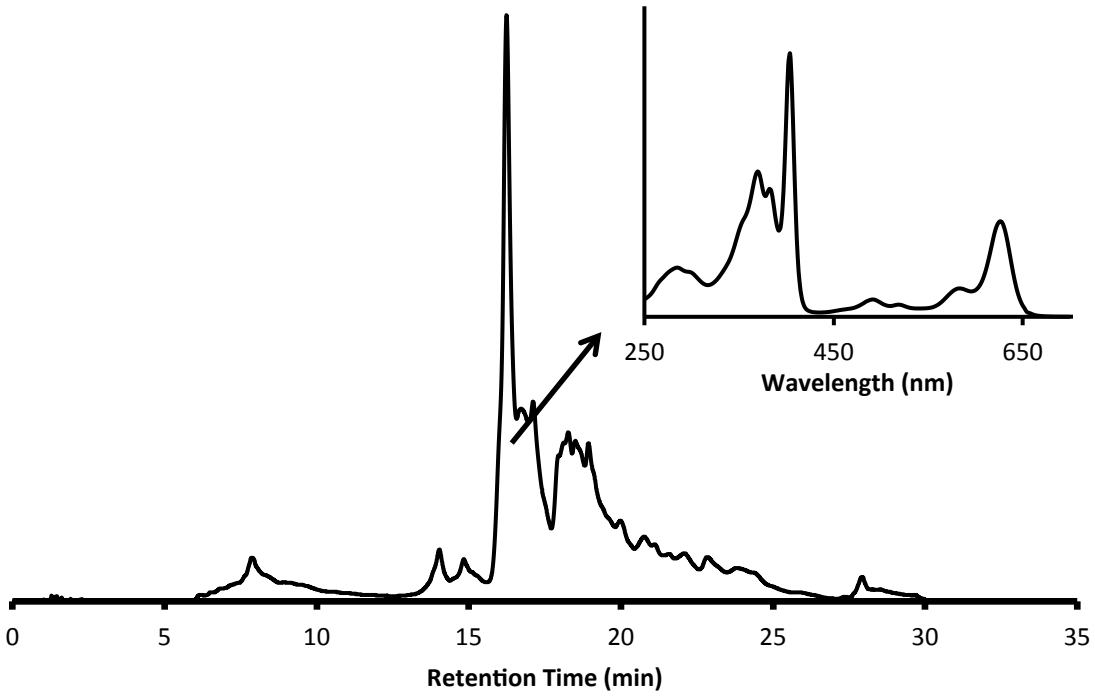


Figure 52. HPLC chromatogram of cell-free extracts containing only the pETDuet-*sirC*, pRSF-*cfbC*-*cfbD*, and pACYC-*cfbB*-*sirA* plasmids (without pCDF-*cfbA*-*cfbE*) showing high levels of sirohydrochlorin production.

LC-MS Analysis of Cell-Free Extracts

Cell-free extracts of *E. coli* cells expressing all required enzymes for the *in vitro* biosynthesis of coenzyme F430 from uroporphyrinogen III (the pETDuet-*sirC*, pRSFDuet-*cfbC*-*cfbD*, pACYCDuet-*cfbB*-*sirA* and pCDFDuet-*cfbA*-*cfbE* plasmid) were analyzed by LC-MS. As observed before in the cell-free extracts of *E. coli* missing *cfbA* and *cfbE,* sirohydrochlorin (863.3 *m/z*) was in high abundance (Figure 53).

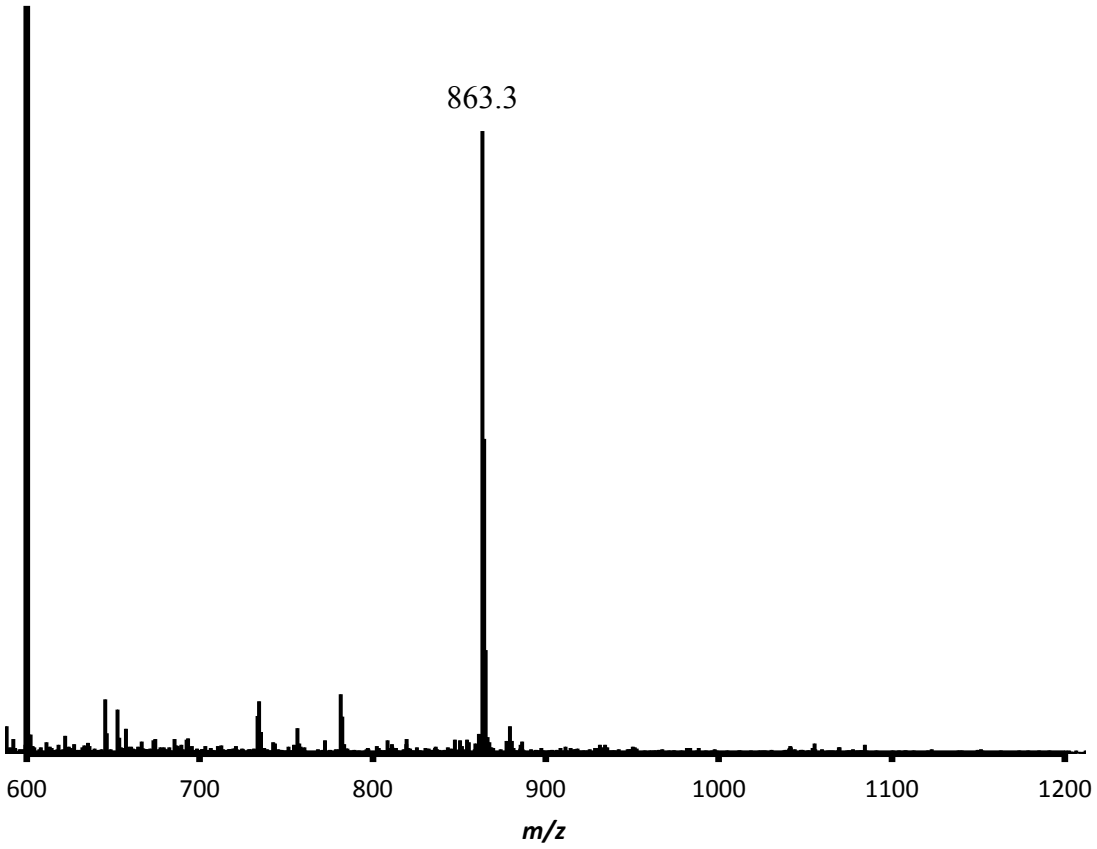


Figure 53. LC-MS spectrum data of sirohydrochlorin present in cell free extracts of *E. coli* expressing the pETDuet-*sirC*, pRSF-*cfbC*-*cfbD*, and pACYC-*cfbB*-*sirA* (without pCDF-*cfbA*-*cfbE*) plasmids.

Present in the sample of cell free extract from cells expressing all the *cfb* genes along with *sirA* was a definitive mass peaks for Ni-sirohydrochlorin *a*,*c*-diamide (917.2 *m/z*) (Figure 54). No discernable traces of either 15,17^3-seco-F430-17^3-acid or coenzyme F430 were detected.

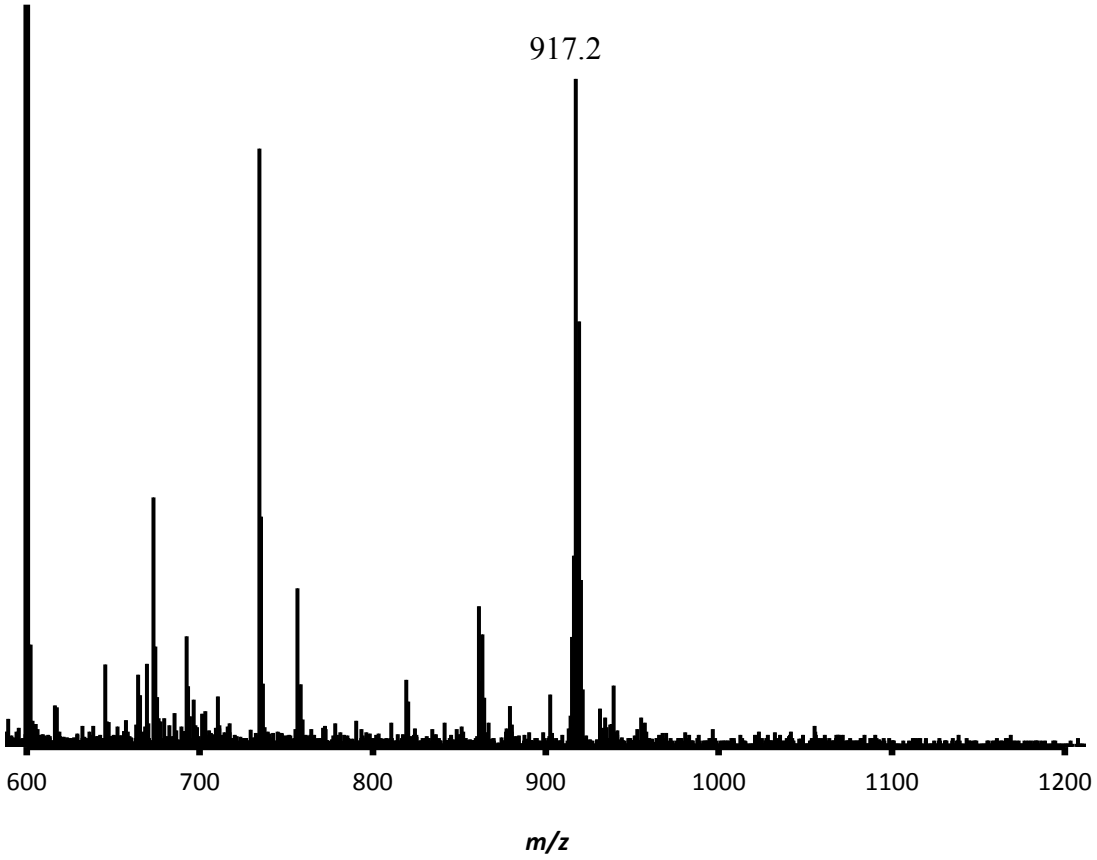


Figure 54. LC-MS data of Ni-sirohydrochlorin *a*,*c*-diamide (917.2 *m*/*z*) present in cell free extracts of *E. coli* expressing the pETDuet-*sirC*, pRSF-*cfbC*-*cfbD*, pACYC-*cfbB*-*sirA*, and pCDF-*cfbA*-*cfbE* plasmids.

4.4 Discussion

Certain considerations must be taken into account when attempting to carryout the biosynthesis of coenzyme F430 *in vivo* that may not have been necessary during the successful biosynthesis of F430 by the Mansoorabadi lab *in vitro*. The most significant of these considerations involve the CfbCD complex. This nitrogenase homolog is activated by sodium dithionite in previous *in vitro* assays performed by the Mansoorabadi lab [90]. Reduction of the CfbCD complex *in vivo* requires an endogenous *E. coli* reducing system, such as flavodoxin, which may or may not be compatible with the CfbCD complex. Lack of evidence of either

coenzyme F430 or the penultimate intermediate in our MS analysis seems to suggest that a heterologous (methanogenic) reducing system is required to activate the CfbCD complex. Additionally, analysis of the cell-free extracts reveals appreciable concentrations of sirohydrochlorin, with smaller amounts of Ni-sirohydrochlorin, and Ni-sirohydrochlorin *a*,*c*-diamide, suggesting that nickel uptake and incorporation is also partially limiting the yield of F430 biosynthetic intermediates in *E. coli*.

Another concern in producing F430 *in vivo* revolves around the host's propensity to not biosynthesize high levels of tetrapyrroles. Thus, we included the pre-sirohydrochlorin tetrapyrrole biosynthetic genes *sirA* and *sirC* in the expression plasmids to increase tetrapyrrole production and supplemented the expression media with 5-aminolevulinic acid (ALA) and porphobilinogen. Further enhancement of the production levels may be achievable by the inclusion of additional early stage tetrapyrrole biosynthetic genes, such as porphobilinogen synthase (*hemB*), hydroxymethylbilane synthase (*hemC*), and uroporphyrinogen III synthase (*hemD*).

This study attempted to produce coenzyme F430 *in vivo* using all the previously reported coenzyme F430 biosynthetic enzymes. Although we were unable to complete the synthesis of F430 in our *E. coli* host with the established *cfb* genes and their corresponding enzymes, we were able to produce the first two dedicated biosynthetic intermediates, Ni-sirohydrochlorin and Ni-sirohydrochlorin *a*,*c*-diamide, and identified several potential bottlenecks in this process. This work further advances the knowledge required to achieve the ultimate goal of constructing active *holo* MCR in a heterologous host.

Chapter 5

The Post-translational Modifications of Methyl-coenzyme M Reductase

The Origin and Purpose of Novel Post-translational Modifications Present in the Active Site of Methyl-coenzyme M Reductase

5.1 Background

Methyl-coenzyme M reductase (MCR) is the enzyme in methanogens that performs the terminal step in methanogenesis. It is responsible for catalyzing the reaction in which methyl-coenzyme M (the methyl source) combines with coenzyme B such that the products are methane and a heterodisulfide [100]. MCR itself, a 270 kDa complex, is an $\alpha_2\beta_2\gamma_2$ heterohexamer that contains two active sites. Additionally, MCR's active site is extensively post-translationally modified with a thioglycine, 5-(*S*)-methylarginine, 1-*N*-methylhistidine, *S*-methylcysteine, and 2-(*S*)-methylglutamine (Figure 55) [82]. Incidentally, a recent study demonstrated that not all five post-translational modifications are conserved in all methanogens [87]. In particular, the 2-(*S*)-methylglutamine is not present is the *Methanosarcina* genus. Instead, *Methanosarcina spp.* possess a compensating alanine-to-threonine substitution in close proximity, which positions an ordered water molecule in the vicinity of the site occupied by the 2-methyl group [85, 87]. The genes and the corresponding enzymes responsible for the biosynthesis of these post-translational modifications and their function are undetermined, yet interest in elucidating these unknowns has gained much momentum recently in hopes of heterologously expressing active *holo* MCR. In this study, *Methanosarcina acetivorans* MCR was analyzed and thus lacks the 2-(*S*)-methylglutamine modification.

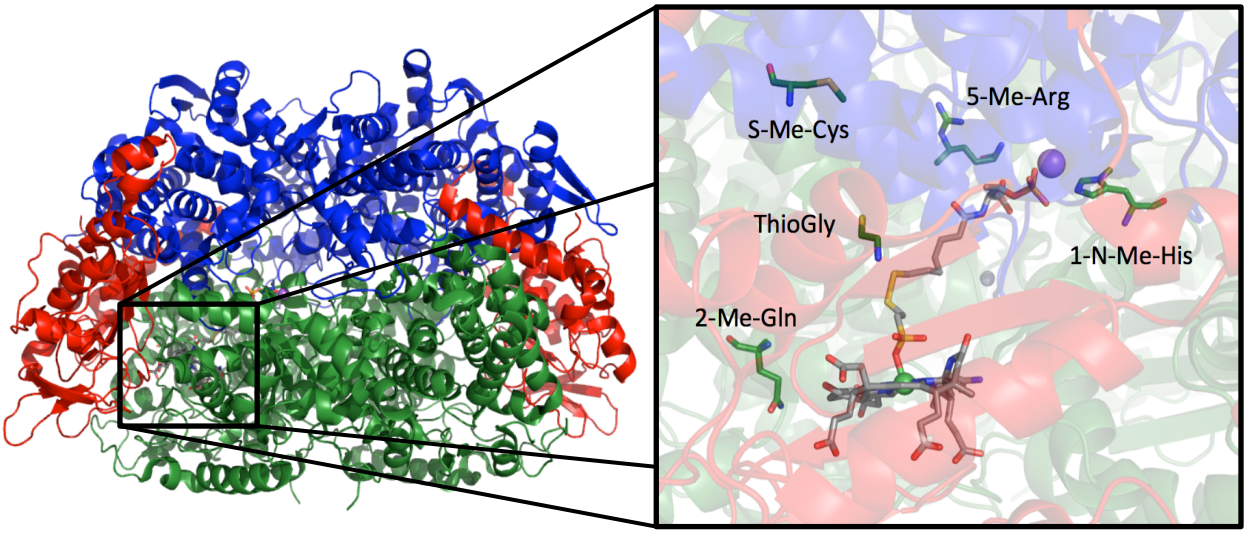


Figure 55. The MCR $\alpha_2\beta_2\gamma_2$ complex featuring an active site housing coenzyme F430, and the thioglycine, 5-(*S*)-methylarginine, 1-*N*-methylhistidine, *S*-methylcysteine, and 2-(*S*)-methylglutamine PTMs (PDB ID: 1HBM).

In deducing the chemistry required to biosynthesize the 1-*N*-methylhistidine and *S*-methylcysteine post-translational modifications, it was determined that a canonical *S*-adenosyl-L-methionine (SAM)-dependent methyltransferase reaction would be sufficient. Interestingly, flanking the *mcr* gene cluster in *Methanosarcina spp.* is an open-reading frame encoding a putative SAM-dependent methyltransferase (Figure 56). A Basic Local Alignment Search Tool (BLAST) analysis of this gene revealed that it is homologous to protein methylation gene A (*prmA*), which catalyzes the methylation of the N-terminal α-amino group and the ε-amino groups of two lysine residues in ribosomal protein L11, a key component of the large subunit of the ribosome. Thus, the *prmA* gene and its encoded enzyme is a prime candidate for the formation of the 1-*N*-methylhistidine and *S*-methylcysteine residues. On the other side of the *mcr* gene cluster resides methanogenesis marker 10 (*mm10*). BLAST analysis of *mm10* reveals that it is a member of the radical SAM superfamily. This is significant because radical SAM enzymes

are capable of catalyzing unusual methylation reactions, such as that required for the biosynthesis of the 5-(*S*)-methylarginine modification, using a highly reactive 5′-deoxyadenosyl radical. Additionally, the C-terminal end of *mm10* is homologous to a domain of unknown function that has weakly homology to glutathione reductase and contains a cysteine-rich motif that is consistent with the presence of an auxiliary iron-sulfur cluster.

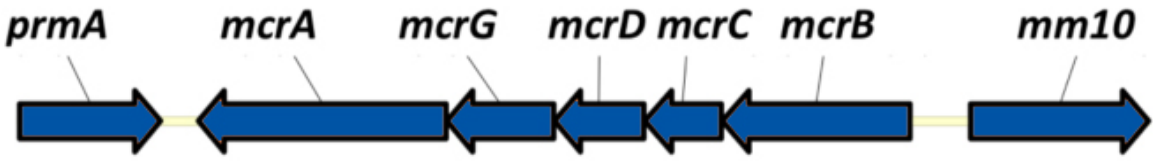


Figure 56. The *mcr* gene cluster from *Methanosarcina acetivorans*.

An extensive comparative genomics investigation yielded three viable candidate genes that may be responsible for the thioglycine post-translational modification. A similar modification is observed in thioviridamide, a thiopeptide containing natural product [101]. The *tva* gene cluster, which is responsible for the biosynthesis of thioviridamide, contains a TfuA associated YcaO homolog that has been implicated in the thioamide modifications [102]. Methanogenesis marker 1 (*mm1*) encodes an enzyme that is a member of the YcaO superfamily. Additionally, *tfuA* has been demonstrated to be important in the production of trifolitoxin, a thiazole containing peptide, though its function is unknown [103]. A *tfuA* homolog is usually found clustered with *mm1* in methanogen genomes. Finally, a *thiI* homolog, a sulfur carrier protein involved in thiamine biosynthesis, is present in the genomes of methanogens [104]. It is postulated herein that the products of these three genes (*mm1*, *tfuA*, and *thiI*) act in concert with each other to install the thioglycine PTM.

In this study, the *mcr* gene cluster and the aforementioned MCR modification genes were cloned from the *Methanosarcina acetivorans* genome into the compatible expression vectors pETDuet, pRSFDuet, pACYCDuet and pCDFDuet. A novel molecular biology strategy was then employed to combine these vectors together for the expression of multiple genes from a single plasmid, each with their own promoter and ribosome-binding site. The resulting plasmids were then cotransformed into *E. coli* for the simultaneous overexpression of the *mcr* gene cluster with each of the putative MCR modification genes to investigate the post-translational modification of MCR *in vivo*.

5.2 Methods

Plasmid Construction

The *mcrABCDG, mm1*, and *mm10* genes and the *prmA*, *tfuA*, and *thiI* homologs were amplified by polymerase chain reaction from the genome of *Methanosarcina acetivorans* C2A purchased from DSMZ in Germany (DSM-2834). Primers were synthesized by Sigma-Aldrich. The sequences of these primers are provided in the Supplemental Information. Phusion High-Fidelity DNA Polymerase (New England Biolabs) was utilized for all PCR reactions in accordance with the manufacturer's protocol.

All plasmids were purchased from Novagen EMD Millipore of Germany. The *mcrA* gene was ligated into the second multiple cloning site of pETDuet with a C-terminal 6×His-tag incorporated to create the pETDuet-*mcrA*-CTH plasmid (Figure 57). The *mcrG* and *mcrB* genes were ligated into the first and second multiple cloning sites of pETDuet, respectively. The *mcrC* and *mcrD* genes were ligated in the first and second multiple cloning sites of pRSFDuet, respectively. Once these plasmids were verified by sequencing, they were used as building

blocks for a novel plasmid construction strategy in which a series of multiple cloning sites from the Duet vectors are ligated together into a single plasmid, allowing for the opportunity of the constituent genes to be simultaneously overexpressed. This is accomplished by introducing a SwaI blunt restriction site approximately 100 bp upstream of the multiple cloning site of each plasmid by site directed mutagenesis. In this case, the base vector is SwaI-pRSFDuet-*mcrC-mrcD*. This plasmid was digested with the SwaI restriction enzyme. The multiple cloning sites containing the *mcrG* and *mcrB* genes were then amplified via PCR using phosphorlylated primers from the SwaI-pETDuet-*mcrG-mcrB* plasmid and ligated into the SwaI digested SwaI-pRSFDuet-*mcrC-mrcD* plasmid, forming the SwaI-pRSFDuet-*mcrG-mcrB-mcrC-mcrD* plasmid (Figure 58).

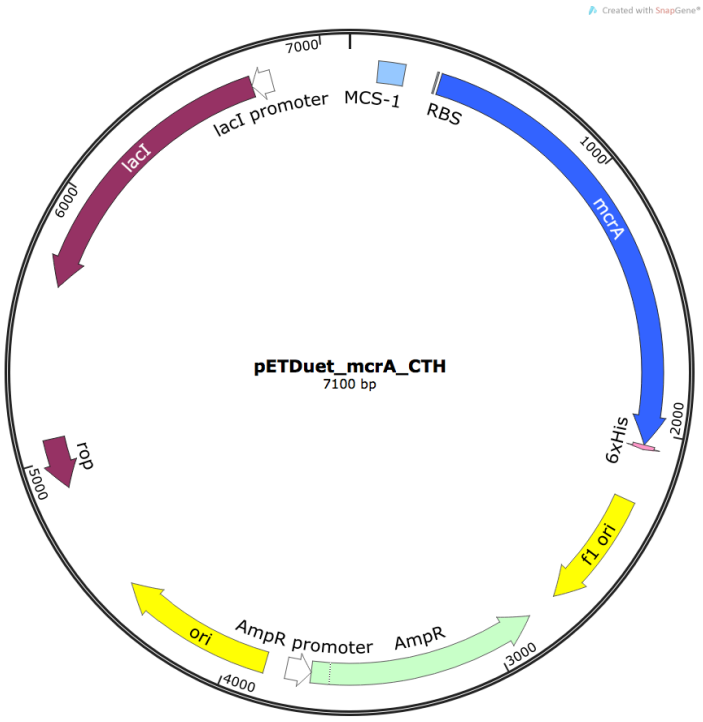


Figure 57. Ampicillin resistant pETDuet expression vector with C-terminal 6×His-tagged *mcrA*.

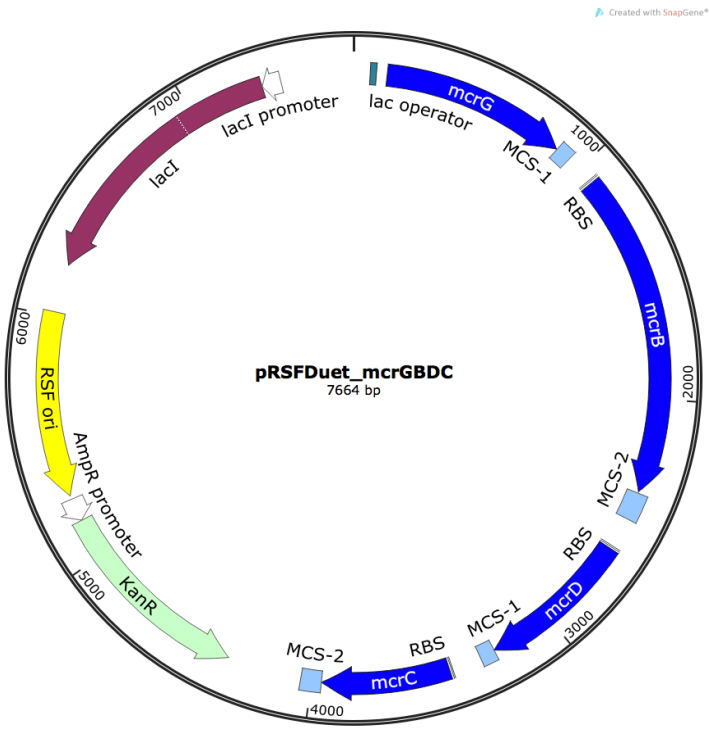


Figure 58. Kanamycin resistant pRSFDuet expression vector containing the *mcrGBDC* genes.

The *mm1* and *mm10* genes were ligated in to the first and second multiple cloning site of pACYCDuet, respectively (Figure 59). The *tfuA* homolog and *prmA* genes were ligated in the first and second multiple cloning site of pCDFDuet, respectively (Figure 60). To test whether the *thiI* gene is required as a sulfur transferase during the course of the thioglycine formation, a second pCDFDuet plasmid was constructed containing *tfuA* in the first multiple cloning site and *thiI* in the second.

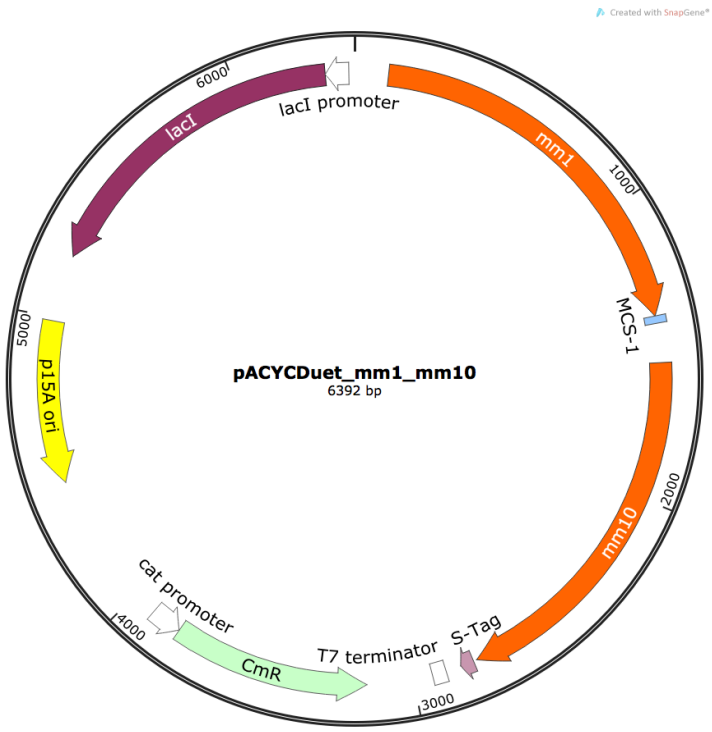


Figure 59. Chloramphenicol resistant pACYCDuet expression vector with *mm1* and *mm10*.

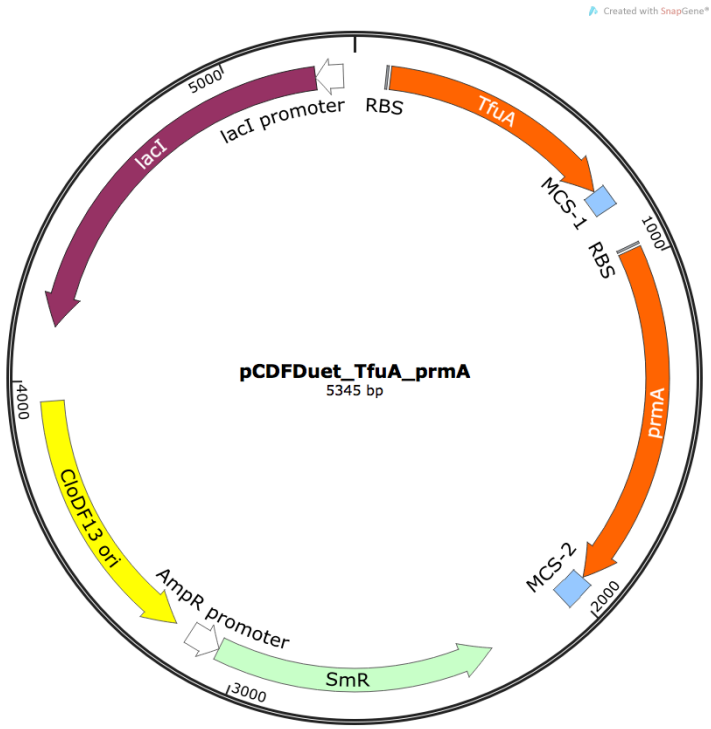


Figure 60. Spectinomycin resistant pCDFDuet expression vector containing *tfuA* and *prmA*.

Coexpression of MCR and Modification Proteins

Constructed plasmids were sequence verified to be correct before being sequentially transformed into *E. coli* BL21 (DE3) cells (New England Labs, Ipswitch, MA) with selection plates supplemented with the appropriate antibiotics. Once all four plasmids had been transformed into a single *E. coli* cell, it was resistant to four different antibiotics: ampicillin, kanamycin, spectinomycin, and chloramphenicol. The first of the resulting strains contained the SwaI-pRSFDuet-*mcrG*-*mcrB*-*mcrC*-*mcrD*, pETDuet-*mcrA*-CTH, pACYC-*mm1*-*mm10*, and pCDF-tfuA-prmA plasmids. The second resulting strain contained the SwaI-pRSFDuet-*mcrG*-*mcrB*-*mcrC*-*mcrD*, pETDuet-*mcrA*-CTH, pACYC-*mm1*-*mm10*, and pCDF-*tfuA*-*thiI* plasmids.

Terrific broth (TB) was prepared as the expression media and was supplemented with a working concentration of ferrous sulfate (200 μM final concentration), cobalamin (2 μM final concentration), riboflavin (50 μM final concentration), methionine (50 μM final concentration), cysteine (200 μM final concentration), ampicillin (100 μg/mL final concentration), kanamycin (50 μg/mL final concentration), spectinomycin (100 μg/mL final concentration), and chloramphenicol (85 μg/mL final concentration). Expression of each of the strains was carried out at 37 °C until the OD_{600} = 0.6, after which the temperature was lowered to 18 °C. The expression cultures were then induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.4 μM and incubated for 16 h with shaking.

To purify the MCR α subunit, the expression cultures were then centrifuged to pellet the cells, which were then resuspended in lysis buffer containing 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, lysozyme (1 mg/ml), Ameresco's Protease Inhibitor Cocktail, and 20% sterile glycerol at pH 8.0. The cell suspension was then sonicated and centrifuged at 30,000 ×g for 1 hour. The supernatant was then applied to a Bio-Rad Econo-Pac column packed with Ni^{2+}-

charged Profinity IMAC Resin. The column was then washed with 50 mM sodium phosphate, 300 mM NaCl, 5 mM imidazole, pH 8.0 buffer. The protein was then eluted with a 50 mM sodium phosphate, 300 mM NaCl, 500 mM imidazole, pH 8.0 buffer and the buffer was exchanged with 100 mM Tris-HCl, pH 8.0 buffer supplemented with a 20% sterile glycerol stock solution (80%).

Peptide Sequencing to Identify Post-translational Modifications

Purified MCR α subunit was then subjected to SDS-PAGE analysis and the appropriately sized band of 62 kDa was excised from the gel (see Supplemental Information). A trypsin digest was then performed on the gel pieces and the resultant pepetide fragments were analyzed by matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) for peptide and post-translational modification identification. MALDI-TOF MS analysis was performed on a Bruker Autoflex MALDI and the data was submitted to the publicly available Matrix Science software package accessing both the MASCOT Peptide Mass Fingerprint server and the University of Georgia Proteomics and Mass Spectrometry server for peptide/PTM identification.

5.3 Results

MADLI-TOF MS of Trypsin Digested Peptides

After digestion with trypsin, the MCR α subunit that was coexpressed with PrmA, was proteolyzed into several peptide fragments of interest that may contain the post-translational modifications of interest. Peptide fragment H271-R285 consisting of HAALVSMGEMLPARR is expected to contain the 1-*N*-methylhistidine modification. The unmodified fragment yields a monoisotopic mass-to-charge ratio (*m*/*z*) of 1638.8 on the MALDI-TOF mass spectrum. Inclusion of the 1-*N*-methylhistidine modification increases the *m*/*z* to 1652.8. The latter fragment was identified to be present and consistent with the presence of the 1-*N*-methylhistidine residue (Figure 61)

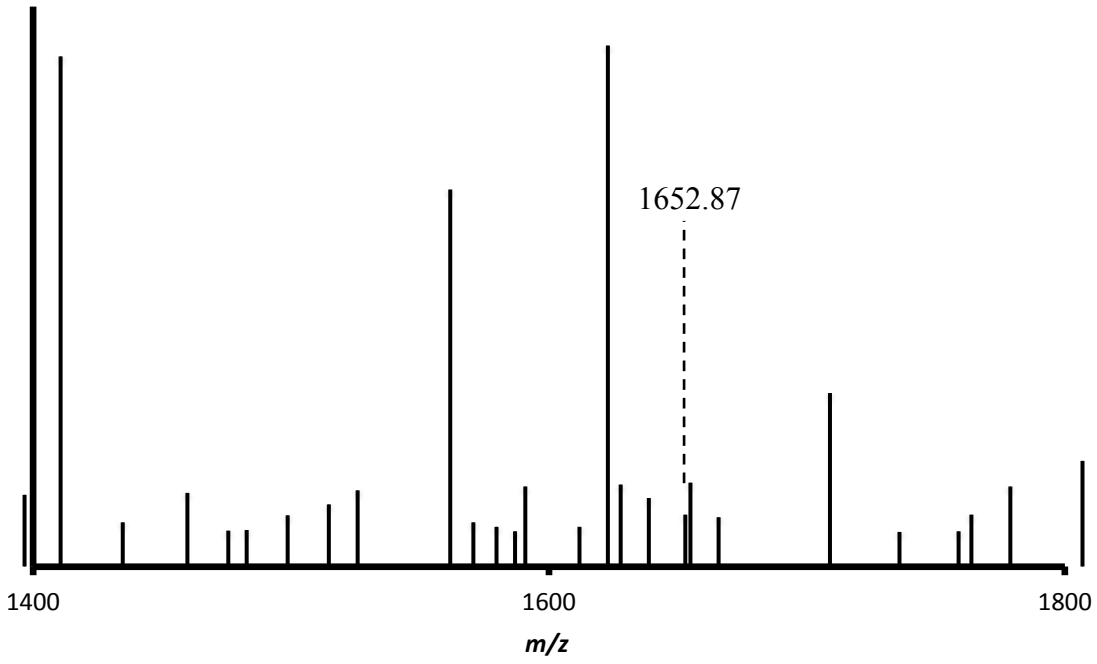


Figure 61. MALDI-TOF MS spectrum of trypsin-digested MCR α-subunit producing the HAALVSMGEMLPARR peptide fragment containing the 1-*N*-methylhistidine modification at 1652.87 *m/z* (see Supplemental Information for full peptide fragment analysis).

The *S*-methylcysteine modification is expected to be contained on a peptide fragment beginning with L461 and ending with R491, consisting of the sequence LGFFGFDLQDQCGATNVLSYQGDEGLPDELR. The unmodified peptide has a monoisotopic *m*/*z* of 3404.6, while the presence of the modification yields a peptide fragment of 3418.6 *m*/*z* in the MALDI-TOF mass spectrum. The latter fragment was identified and is consistent with the presence of the *S*-methylcysteine residue (Figure 62)

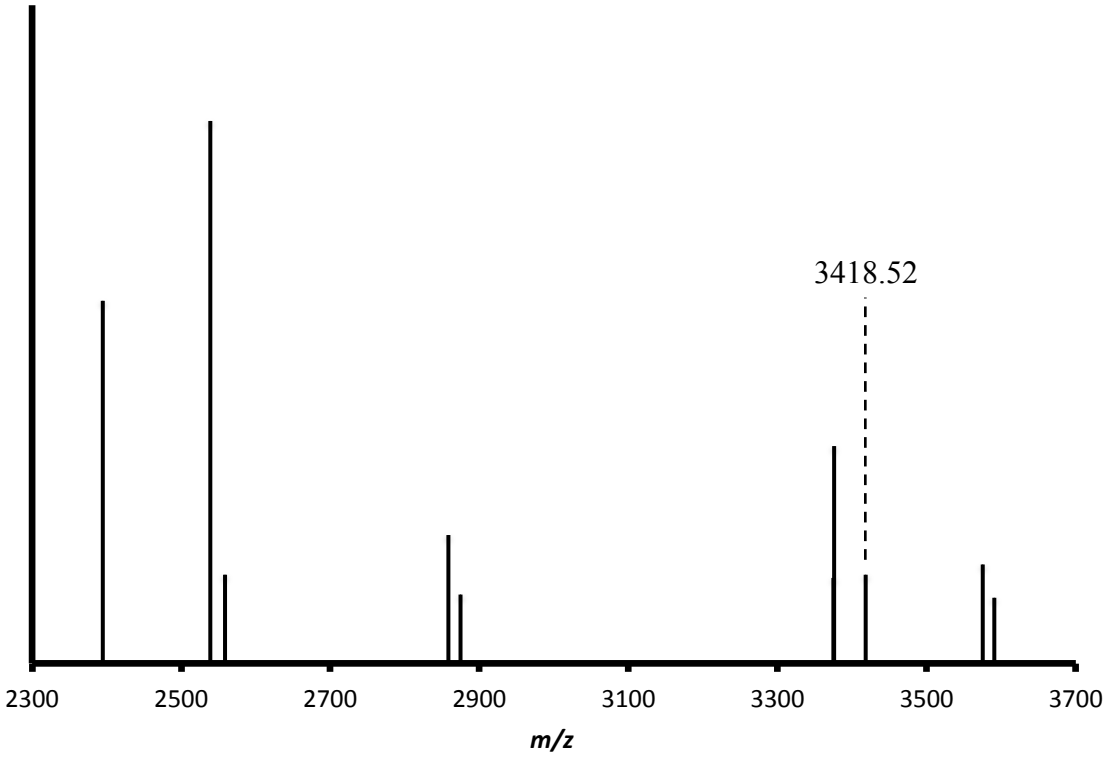


Figure 62. MALDI-TOF MS spectrum of trypsin digested MCR α subunit producing the LGFFGFDLQDQCGATNVLSYQGDEGLPDELR peptide fragment that contains the *S*-methylcysteine modification at 3418.5260 *m/z* (see Supplemental Information for full peptide fragment analysis).

5.4 Discussion

MCR α subunits expressed in the presence of PrmA yielded peptide fragments with charge-to-mass ratios consistent with the presence of both the 1-*N*-methylhistidine and *S*-methylcysteine modifications. These peptide fragments were not observed in MCR α-subunits expressed in the absence of PrmA, suggesting that PrmA is responsible for the biosynthesis of both the 1-*N*-methylhistidine and *S*-methylcysteine. Neither the 5-(*S*)-methylarginine nor the thioglycine post-translational modification were observed during the course of this study.

Collaborator William (Barny) Whitman (University of Georgia) and coworkers recently performed gene knockout studies in which a mutant of *Methanococcus maripaludis* was engineered with the *mm10* gene knocked out, and its MCR was compared to that of the wild type strain (personal communication). In the *mm10* knockout strain, MCR did not possess the 5-(*S*)-methylarginine modification. Thus, we anticipated that the coexpression of Mm10 with McrA would produce peptide fragments with the 5-(*S*)-methylarginine modification. The absence of the 5-(*S*)-methylarginine modification in our experiments brings to light a number of possible deficiencies in using an *E. coli* expression host that might prevent the formation of fully mature MCR. Radical SAM enzymes that catalyze the methylation of unactivated sp^3 hybridized centers require an additional methyl donor, such as methylcobalamin or 5-methyltetrahydrofolate. Methanogens also utilize a unique methyl donor, 5-methyl tetrahydromethanopterin. *E. coli* produced 5-methyltetrahydrofolate, but does not have a complete methylcobalamin biosynthetic pathway (nor does BL21 derived strains contain a fully functional B_{12} transport system) and does not synthesize 5-methyl tetrahydromethanopterin. It is possible that this may be the root cause for the lack of the 5-(*S*)-methylarginine modification. Also, bioinformatic analysis of

Mm10 suggests that it contains several FeS clusters that may require supplementation with enzymes dedicated to the manufacturing FeS clusters.

The peptide fragment expected to contain the thioglycine residue was observed in the repertoire of peptide fragments produced in the trypsin digest of McrA coexpressed with PrmA, but not with the McrA coexpressed with ThiI. In examining the former, the thioglycine modification did not appear to be present, which suggests that more than just the Mm1 or TfuA enzymes are required to complete this modification. Since McrA coexpressed with the ThiI enzyme did not yield the expectant peptide fragment, we cannot conclusively deduce whether the lack of the thioglycine modification is due to the absence of the ThiI enzyme, or if other contributing factors are required for the reaction. However, the thioglycine residue is prone to hydrolysis, and it is possible that the McrA subunit purified in isolation could have lost this modification prior to MS analysis.

With the MALDI-TOF MS data indicating that both the 1-*N*-methylhistidine and *S*-methylcysteine residues of McrA are introduced by the PrmA enzyme and the reporting of Mm10 being responsible for the 5-(*S*)-methylarginine modification, the origin of only one modification, the thioglycine, remains to be discovered in *M. acetivorans*. The work of this study brings us closer to the ultimate goal of heterologously expressing active *holo* MCR.

Conclusion

Achieving the ultimate goal of constructing active *holo* MCR brings with it the potential for its use in methane bioconversion strategies. By and large, the C—H bond of methane is stable and not easily activated for oxidation [105, 106]. And while methane itself is a fuel, at atmospheric temperatures and pressures, its gaseous state make transportation both cumbersome and hazardous [107]. Thus, activation of the C—H bond for oxidation of methane to methanol or another more easily transported commodity chemical is desirable. In principle, this reaction can be carried out by employing reverse methanogenesis. While reverse methanogenesis is possible using MCR, it can only be accomplished by coupling methane oxidation to the reduction of a thermodynamically favorable electron acceptor (e.g., sulfate) due to the large barrier that must be overcome to carry out this reaction [108]. Indeed, anaerobic methanotrophic archea (ANME) that use MCR to catalyze the anaerobic oxidation of methane (AOM) are often found as part of syntrophic consortia with sulfate reducing bacteria that accept the electrons via direct interspecies electron transfer (DIET) [109]. Another approach to producing a viable biofuel focuses on the production of acetate by strains of *Methanosarcina acetivorans* engineered to express ANME MCR. Acetate can be used as a precursor in producing ethanol, fatty alcohols, alkanes and acyl esters through acetyl-CoA or isopentanol utilizing the isoprenoid pathway [105]. To this end, the work described in the preceding chapters of this dissertation represents a significant advancement in the knowledge required to ultimately engineer a strain with fully active MCR that is capable of biosynthesizing biofuels or biofuel precursors.

Appendix 1:

Chapter 1

The Biosynthetic Pathway of Dinoflagellate Luciferin

Supplemental Information

Identification and Characterization of Novel Chlorophyll Catabolites from the Bioluminescent Dinoflagellate *Pyrocystis fusiformis*

Supplemental Information

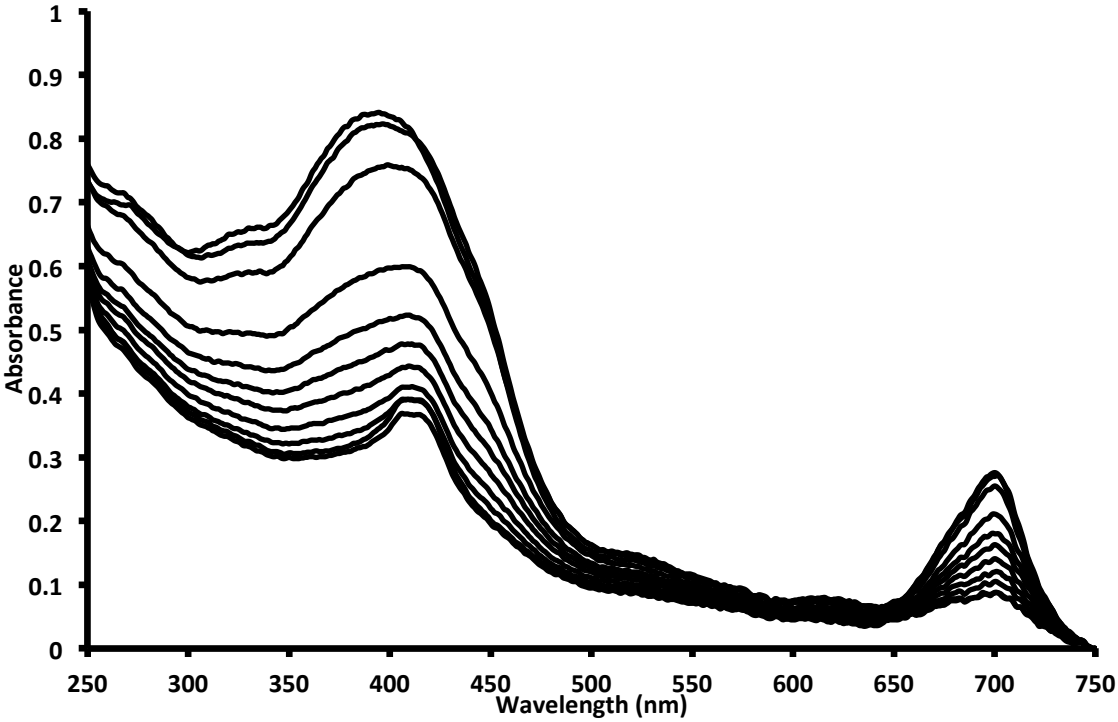


Figure A1S1. Trifluoroacetic acid titration of P710.

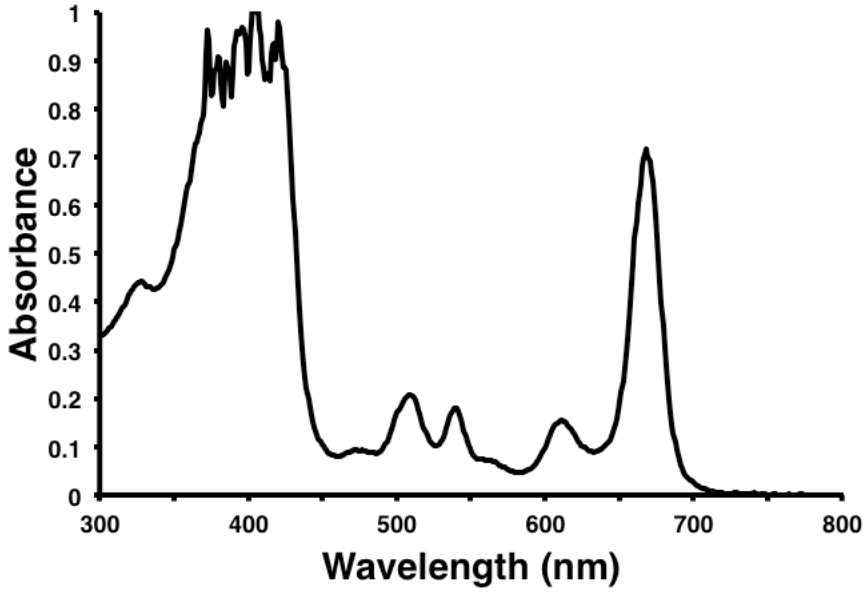


Figure A1S2. UV-visible spectrum of pheophorbide *a* in DMSO.

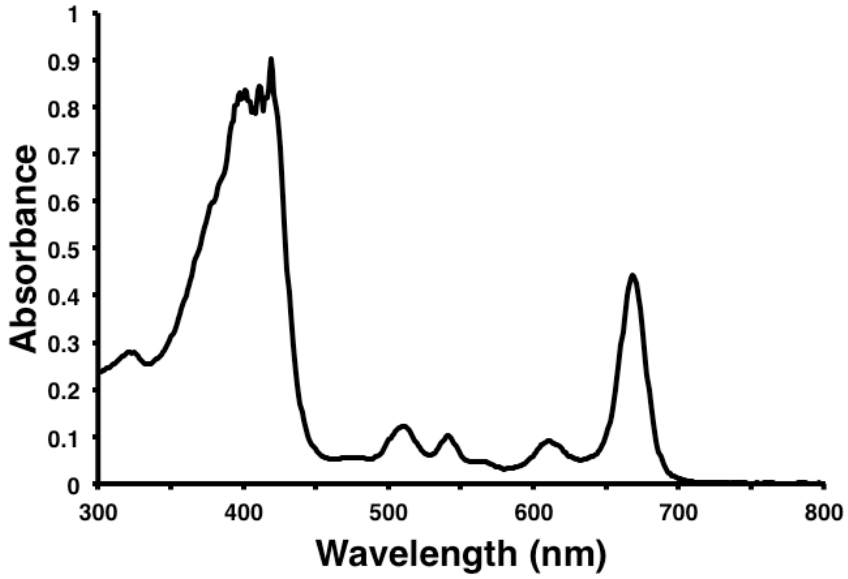


Figure A1S3. UV-visible spectrum of pyropheophorbide *a* in DMSO.

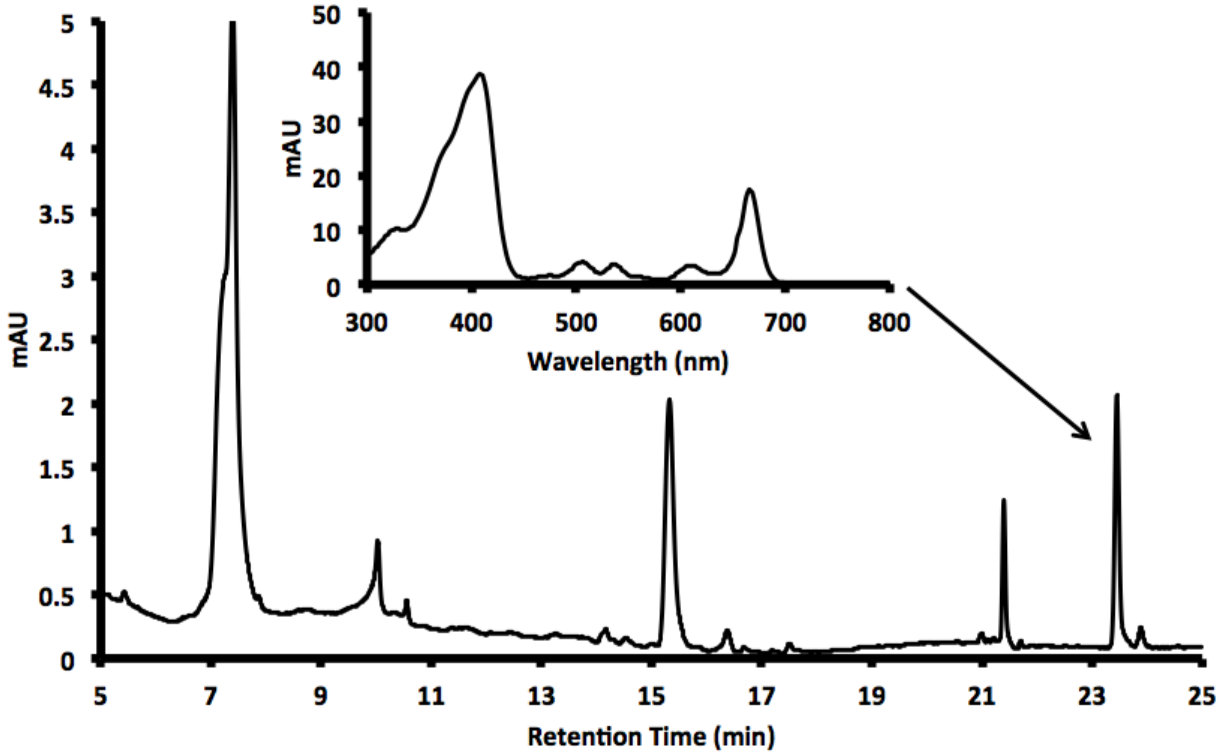


Figure A1S4. HPLC chromatogram of a crude extract of dinoflagellate tetrapyrroles featuring a peak at retention time 23.4 minutes displaying the UV-visible spectrum of either pheophorbide *a* or pyropheophorbide *a*.

Appendix 2:

Chapter 2

The pH Regulation of Dinoflagellate Luciferse

Supplemental Information

Probing the pH Regulation Mechanism of Dinoflagellate Luciferase Using Constant pH Accelerated Molecular Dynamics

Supplemental Information

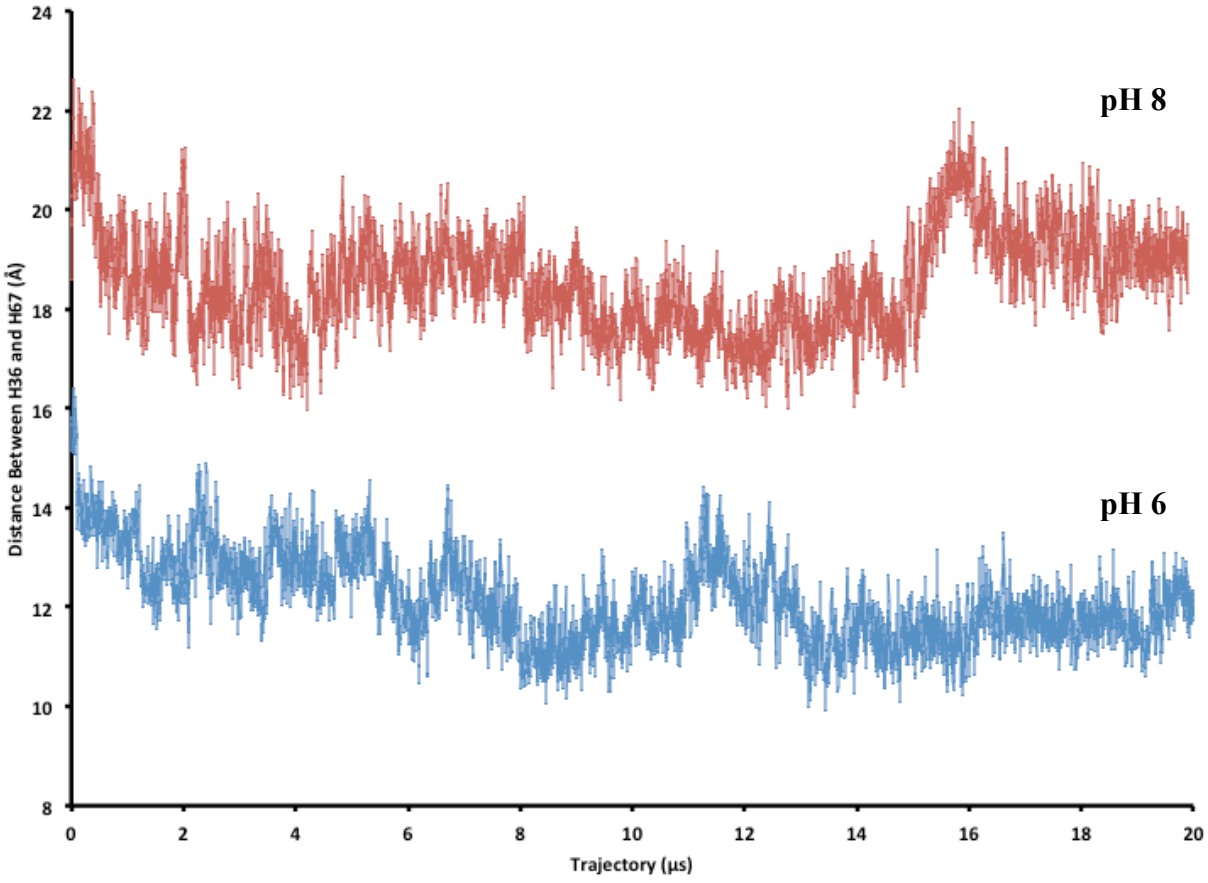


Figure A2S1. Residue-to-residue distance over the course of the trajectory of His36 and His67 in *Pyrocystis fusiformis* luciferase domain II.

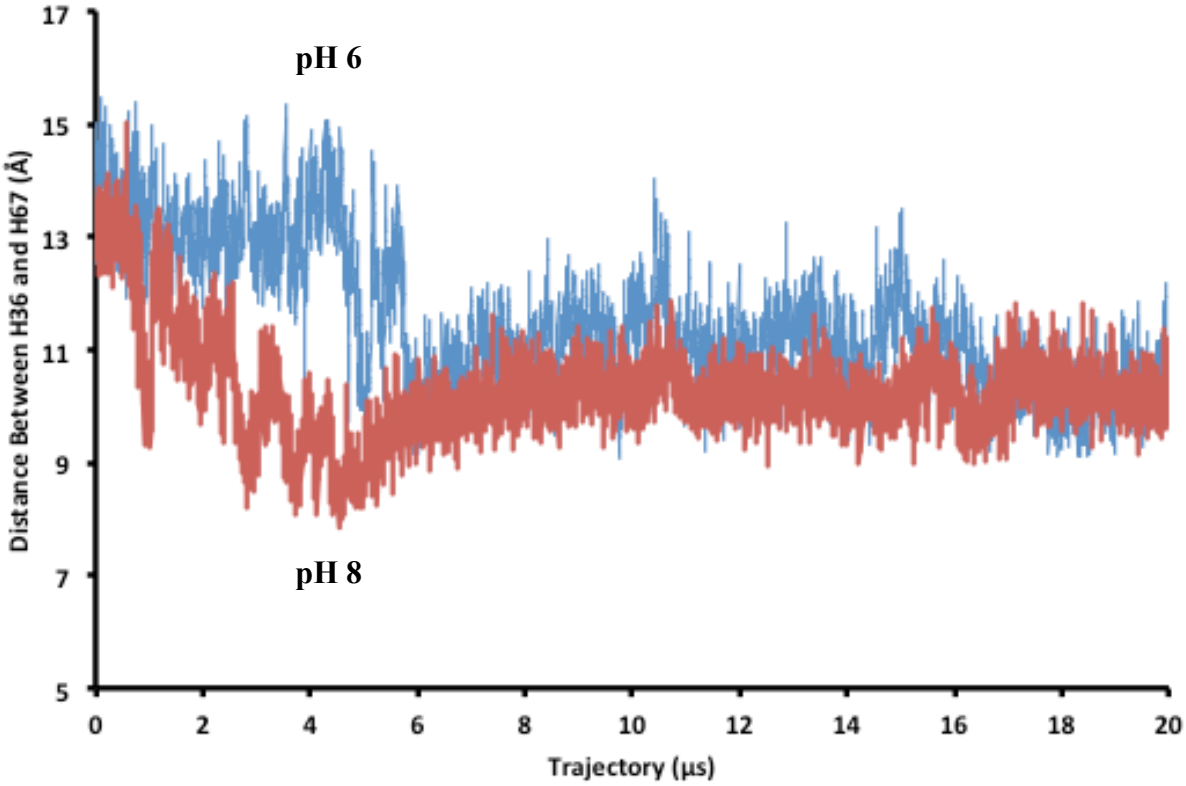


Figure A2S2. Residue-to-residue distance over the course of the trajectory of His36 and His67 in *Pyrocystis fusiformis* luciferase domain I.

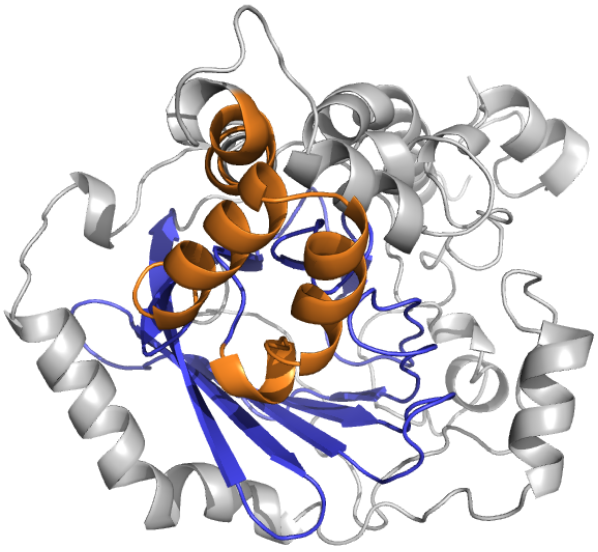


Figure A2S3. Structure of *Pyrocystis fusiformis* luciferase domain I at the end of a 1 µs CpHaMD simulation.

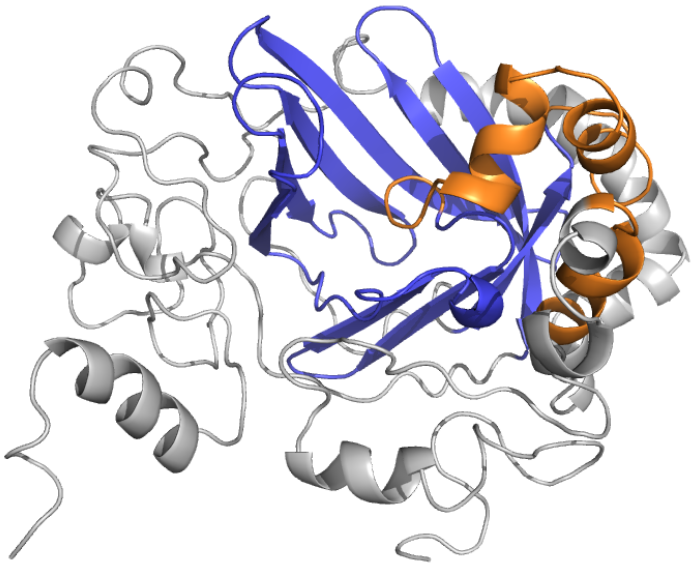


Figure A2S4. Structure of *Pyrocystis fusiformis* luciferase domain II at the end of a 1 μs CpHaMD simulation.

Appendix 3:

Chapter 3

The Mechanism of Dinoflagellate Bioluminescence

Supplemental Information

Computational Investigation of the Mechanism of Dinoflagellate Bioluminescence: Chemically Initiated Electron Exchange Luminescence or Twisted Intramolecular Charge Transfer?

Supporting Information

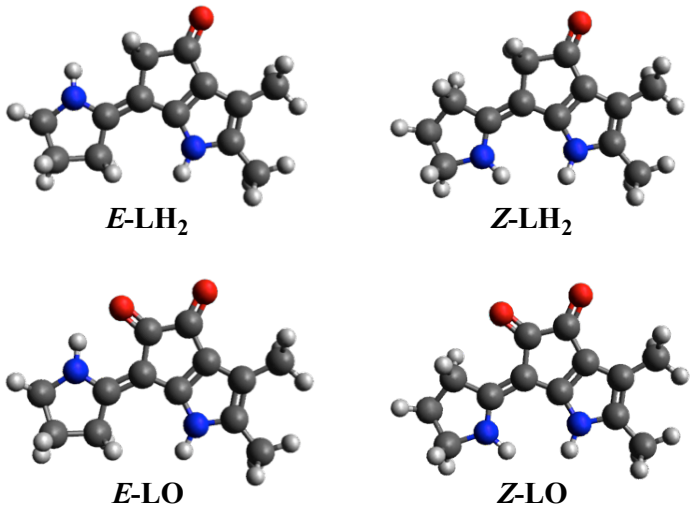


Figure A3S1. Optimized structures of the *E*- and *Z*-isomers of dinoflagellate luciferin (LH_2) and oxyluciferin (LO) in aqueous medium.

Table A3S1. Calculated substrate and product energies of the LCF-catalyzed reaction in aqueous and proteinaceous environments. Substrate energies are assigned to 0.0 kcal/mol and product energies are given with respect to their cognate substrate. Values in parentheses are absolute energies relative to *E*-LH_2 + O_2 + $CH_3CO_2^-$ (aq).

Environment	**Aqueous**		**Proteinaceous**	
Intermediate	**$^1\Delta_g$ (kcal/mol)**	**$^3\Sigma_g^-$ (kcal/mol)**	**$^1\Delta_g$ (kcal/mol)**	**$^3\Sigma_g^-$ (kcal/mol)**
E-LH_2 + O_2 + $CH_3CO_2^-$	0.0 (22.5)	0.0 (0.0)	0.0 (25.5)	0.0 (3.0)
Z-LH_2 + O_2 + $CH_3CO_2^-$	0.0 (25.9)	0.0 (3.4)	0.0 (29.0)	0.0 (6.5)
E-LO + H_2O + $CH_3CO_2^-$	-113.4 (-90.9)	-90.9 (-90.9)	-113.2 (-87.7)	-90.7 (-87.7)
Z-LO + H_2O + $CH_3CO_2^-$	-114.4 (-88.6)	-91.9 (-88.6)	-114.3 (-85.3)	-91.8 (-85.3)

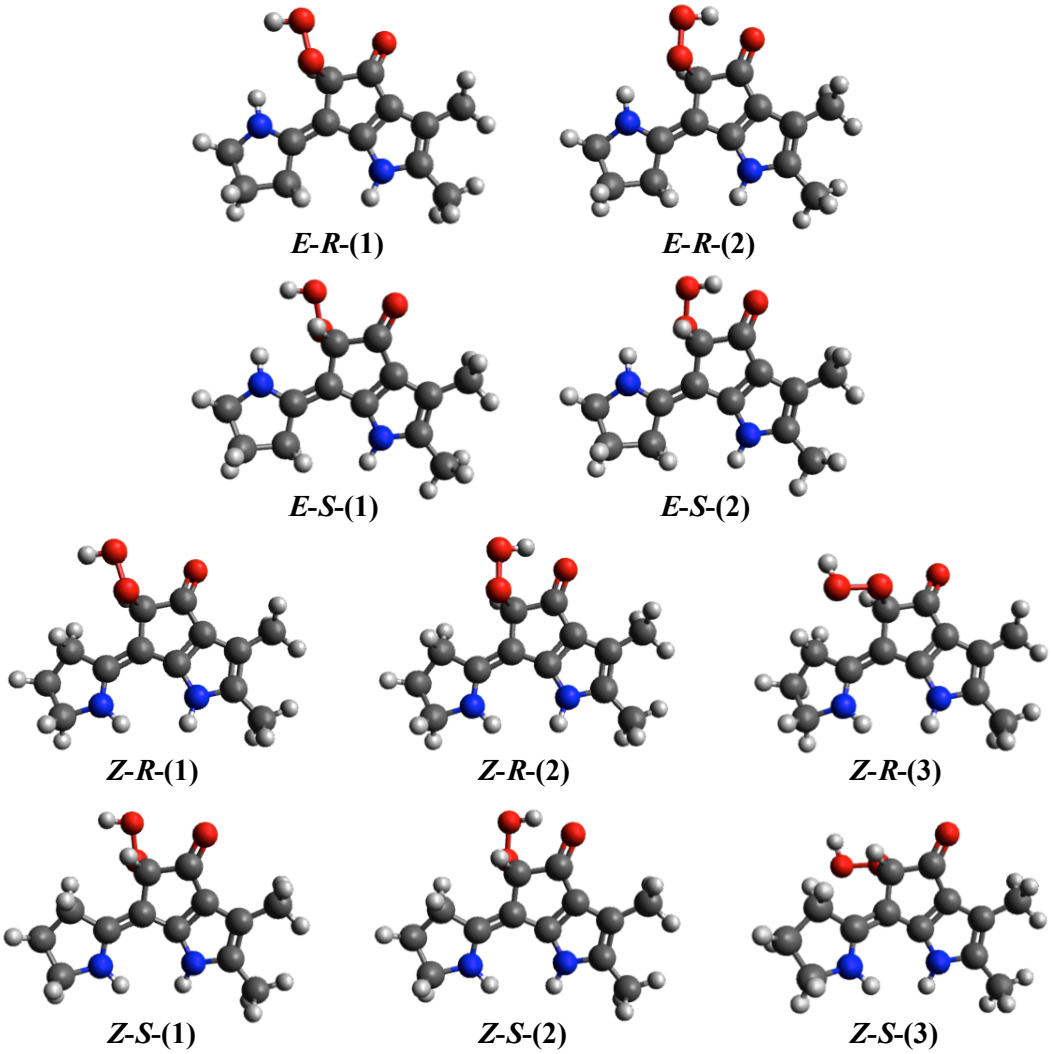


Figure A3S2. Optimized structures of the *E*- and *Z*-isomers of the hydroperoxide (LHOOH) intermediate in aqueous medium.

Table A3S2. Calculated ground and excited state energies of the hydroperoxide (LHOOH) intermediate in an aqueous (proteinaceous) environment. Intermediate energies are given with respect to their cognate substrate energy using triplet ($^3\Sigma_g^-$) oxygen. The corresponding excitation wavelengths (λ) and oscillator strengths (f) are also shown.

LHOOH Intermediate	**Ground State (kcal/mol)**	**Excited State (kcal/mol)**	**λ (nm)**	***f***
E-*R*-(1) + $CH_3CO_2^-$	-28.2 (-28.1)	34.5 (35.1)	456 (452)	0.21 (0.20)
E-*R*-(2) + $CH_3CO_2^-$	-33.7 (-34.4)	30.2 (30.5)	448 (440)	0.29 (0.29)
E-*S*-(1) + $CH_3CO_2^-$	-28.3 (-28.3)	34.1 (34.8)	458 (453)	0.20 (0.19)
E-*S*-(2) + $CH_3CO_2^-$	-33.8 (-35.2)	29.8 (30.1)	450 (438)	0.29 (0.30)
Z-*R*-(1) + $CH_3CO_2^-$	-30.4 (-30.2)	31.8 (32.4)	459 (457)	0.27 (0.27)
Z-*R*-(2) + $CH_3CO_2^-$	-35.0 (-35.2)	28.3 (28.5)	452 (449)	0.32 (0.31)
Z-*R*-(3) + $CH_3CO_2^-$	-29.7 (-29.8)	30.1 (30.4)	479 (475)	0.22 (0.22)
Z-*S*-(1) + $CH_3CO_2^-$	-29.6 (-29.0)	32.1 (32.6)	464 (464)	0.26 (0.25)
Z-*S*-(2) + $CH_3CO_2^-$	-34.6 (-34.9)	28.4 (28.6)	454 (451)	0.31 (0.30)
Z-*S*-(3) + $CH_3CO_2^-$	-28.9 (-28.8)	30.4 (30.7)	482 (481)	0.21 (0.20)

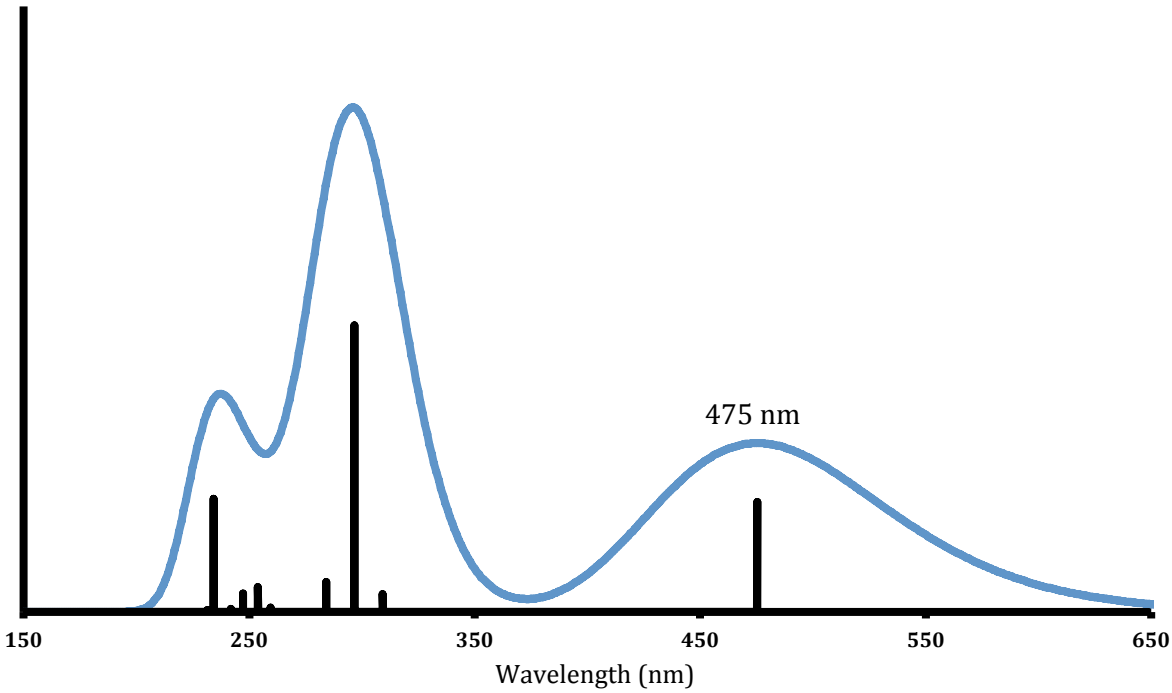


Figure A3S3. Calculated absorption spectrum of the *Z*-*R*-(3)-hydroperoxide intermediate.

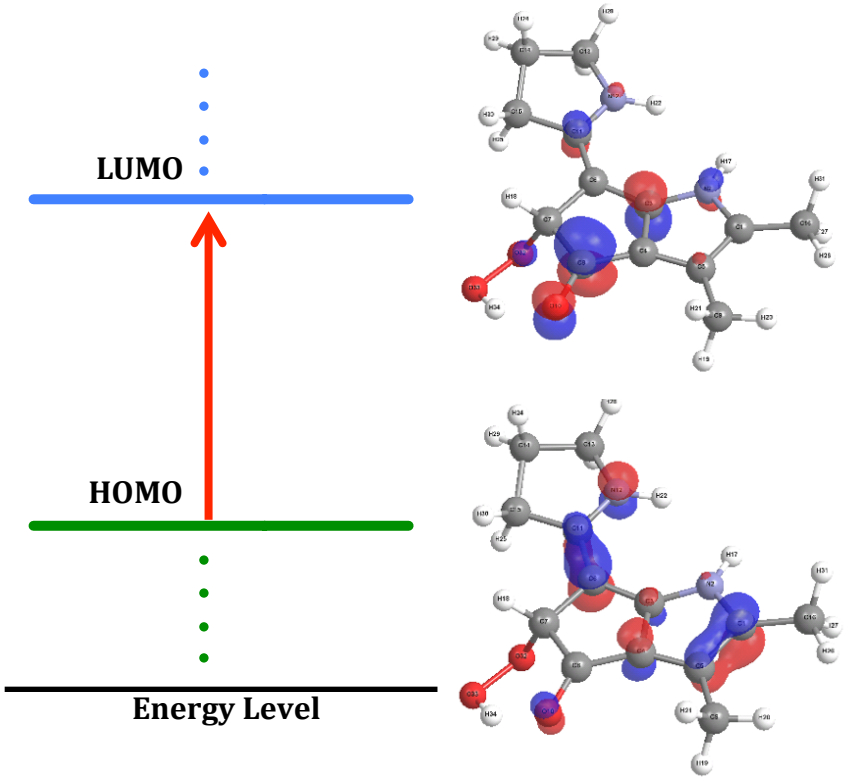


Figure A3S4. Calculated molecular orbitals of the *Z*-*R*-(3)-hydroperoxide intermediate.

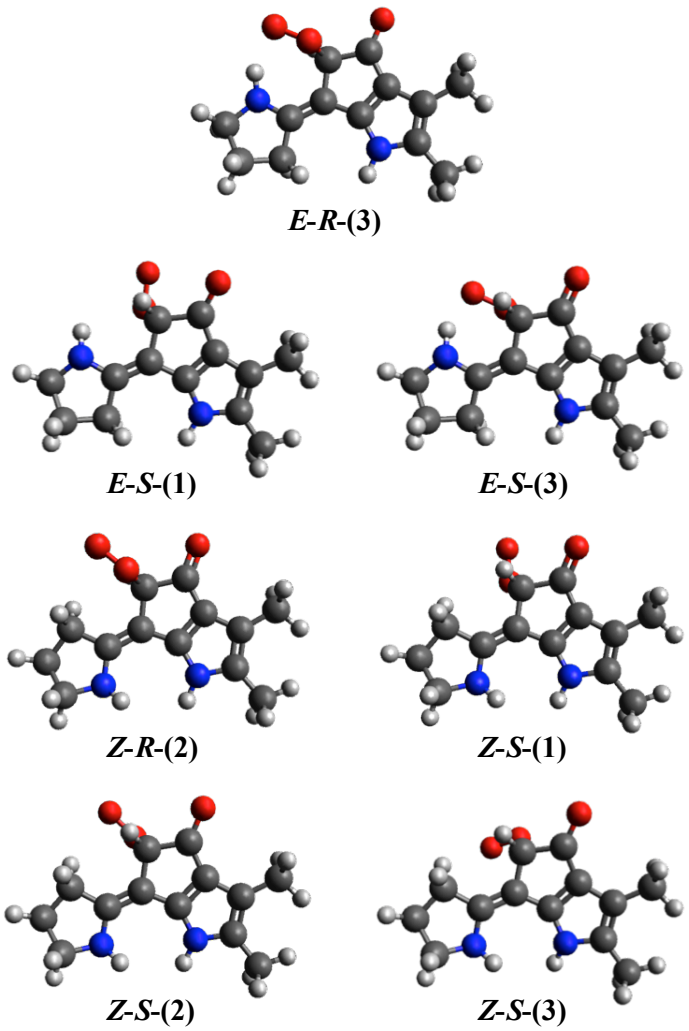


Figure A3S5. Optimized structures of the *E*- and *Z*-isomers of the peroxy anion ($LHOO^-$) intermediate in aqueous medium (except for *E*-*R*-(3), which was optimized in proteinaceous medium).

Table A3S3. Calculated ground and excited state energies of the peroxy anion ($LHOO^-$) intermediate in an aqueous (proteinaceous) environment. Intermediate energies are given with respect to their cognate substrate energy using triplet ($^3\Sigma_g^-$) oxygen. The corresponding excitation wavelengths (λ) and oscillator strengths (f) are also shown.

$LHOO^-$ Intermediate	**Ground State (kcal/mol)**	**Excited State (kcal/mol)**	**λ (nm)**	***f***
E-*R*-(3) + CH_3CO_2H	N/A (-20.6)	N/A (33.8)	N/A (525)	N/A (0.05)
E-*S*-(1) + CH_3CO_2H	-14.9 (-13.9)	36.1 (35.5)	561 (580)	0.01 (0.01)
E-*S*-(3) + CH_3CO_2H	-24.0 (-21.5)	33.1 (32.6)	501 (528)	0.09 (0.04)
Z-*R*-(2) + CH_3CO_2H	-16.2 (-15.3)	32.4 (39.1)	588 (609)	0.02 (0.01)
Z-*S*-(1) + CH_3CO_2H	-18.8 (-17.6)	33.4 (32.7)	548 (568)	0.02 (0.01)
Z-*S*-(2) + CH_3CO_2H	-17.5 (-16.3)	32.6 (31.9)	571 (594)	0.02 (0.01)
Z-*S*-(3) + CH_3CO_2H	-12.7 (-12.3)	35.5 (34.5)	593 (610)	0.01 (0.01)

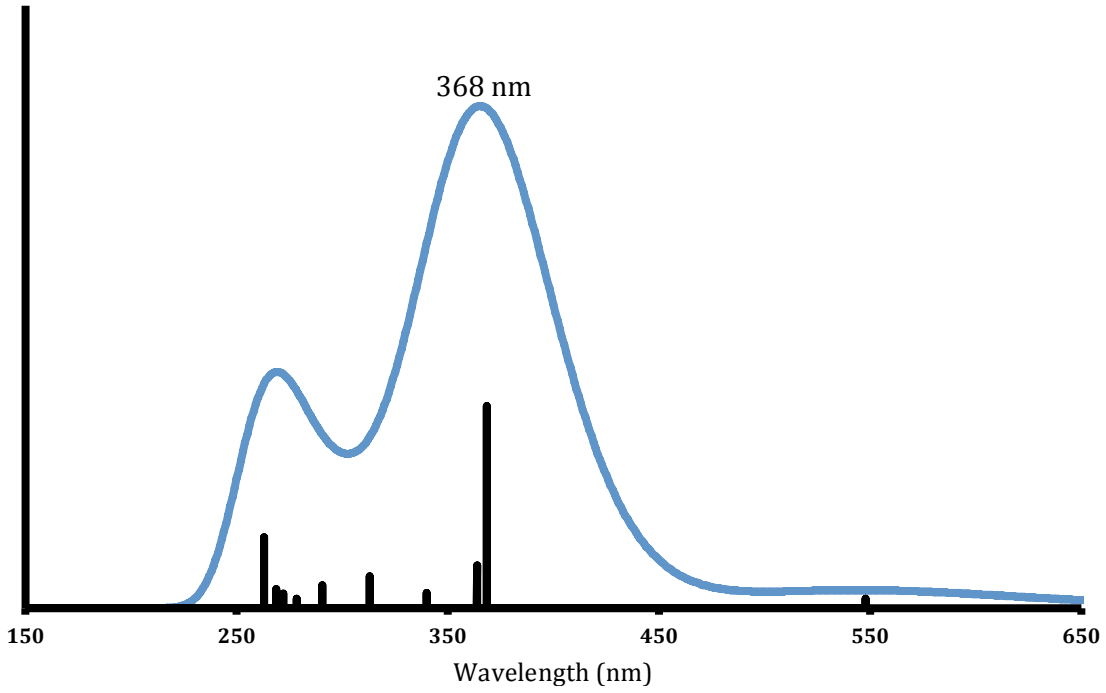


Figure A3S6. Calculated absorption spectrum of the *Z*-*S*-(2)-peroxy anion intermediate.

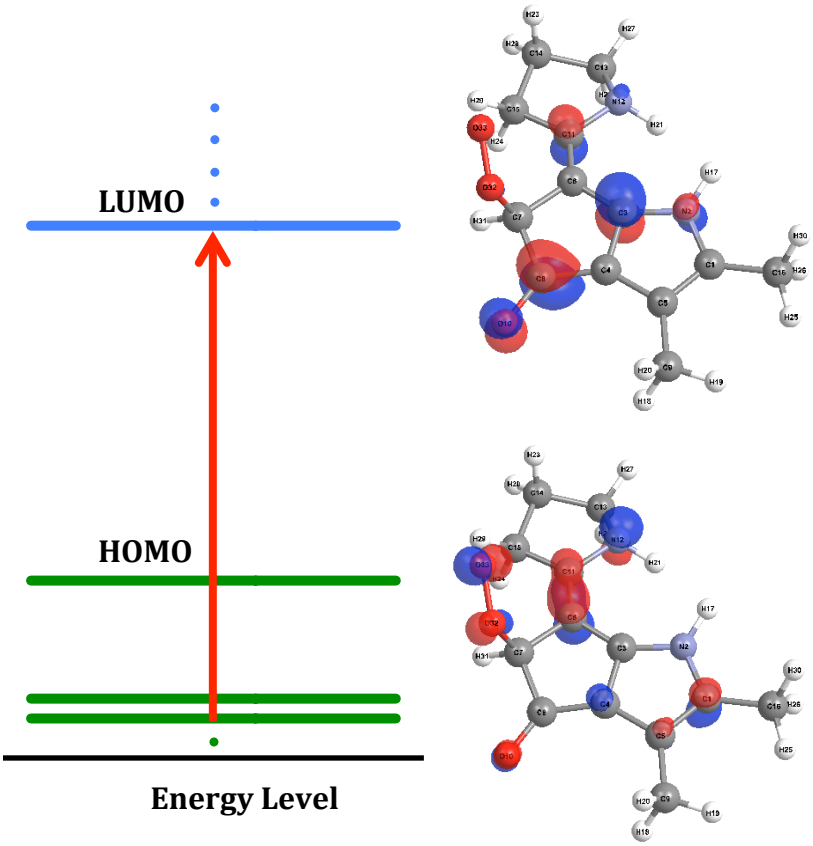


Figure A3S7. Calculated molecular orbitals of the *Z*-*S*-(2)-peroxy anion intermediate.

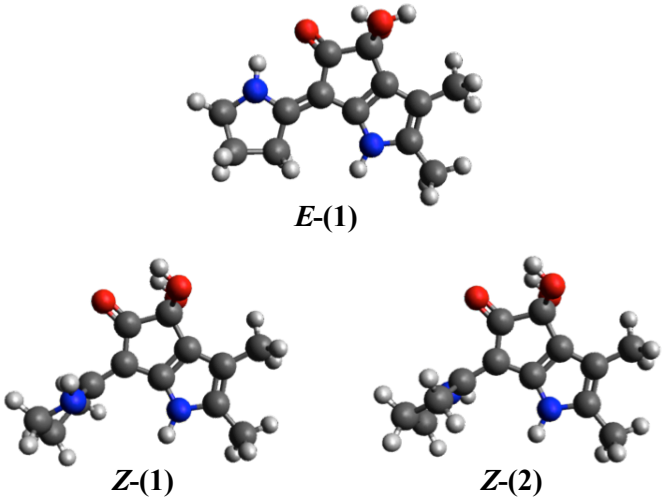


Figure A3S8. Optimized structures of the *E*- and *Z*-isomers of the *gem*-diol (LOHOH) intermediate in aqueous medium.

Table A3S4. Calculated ground and excited state energies of the *gem*-diol (LOHOH) intermediate in an aqueous (proteinaceous) environment. Intermediate energies are given with respect to their cognate substrate energy using triplet ($^3\Sigma_g^-$) oxygen. The corresponding excitation wavelengths (λ) and oscillator strengths (f) are also shown.

LOHOH Intermediate	**Ground State (kcal/mol)**	**Excited State (kcal/mol)**	**λ (nm)**	***f***
E-(1) + $CH_3CO_2^-$	-93.8 (-93.4)	0.1 (0.6)	507 (512)	0.10 (0.09)
Z-(1) + $CH_3CO_2^-$	-63.0 (-65.2)	17.4 (15.3)	356 (355)	0.12 (0.16)
Z-(2) + $CH_3CO_2^-$	-64.6 (-64.2)	16.0 (15.6)	355 (359)	0.15 (0.16)

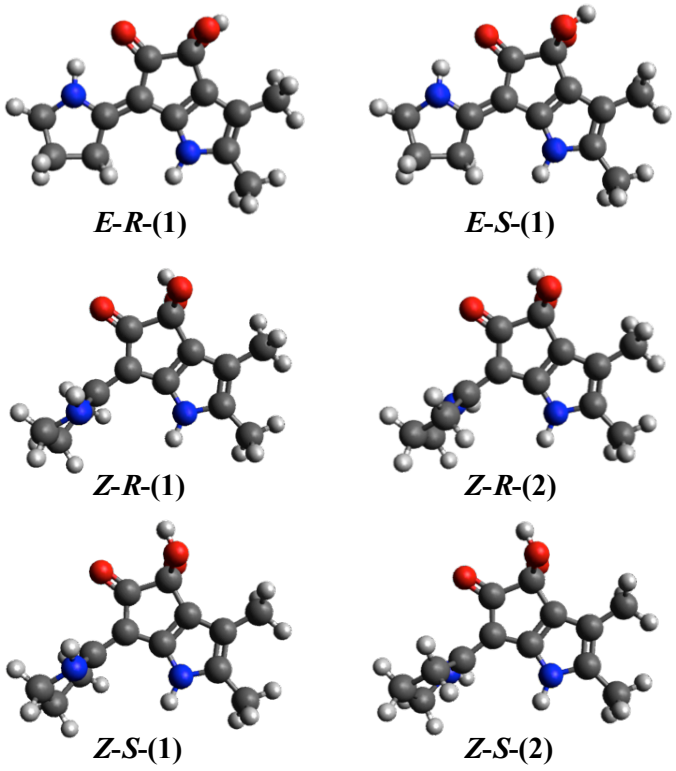


Figure A3S9. Optimized structures of the *E*- and *Z*-isomers of the *gem*-diolate ($LOHO^-$) intermediate in aqueous medium.

Table A3S5. Calculated ground and excited state energies of the *gem*-diolate ($LOHO^-$) intermediate in an aqueous (proteinaceous) environment. Intermediate energies are given with respect to their cognate substrate energy using triplet ($^3\Sigma_g^-$) oxygen. The corresponding excitation wavelengths (λ) and oscillator strengths (*f*) are also shown.

$LOHO^-$ Intermediate	**Ground State (kcal/mol)**	**Excited State (kcal/mol)**	**λ (nm)**	***f***
E-*R*-(1) + CH_3CO_2H	-91.1 (-90.8)	-25.7 (-24.9)	437 (433)	0.22 (0.23)
E-*S*-(1) + CH_3CO_2H	-90.4 (N/A)	-26.3 (N/A)	446 (N/A)	0.23 (N/A)
Z-*R*-(1) + CH_3CO_2H	-50.3 (N/A)	8.6 (N/A)	486 (N/A)	0.08 (N/A)
Z-*R*-(2) + CH_3CO_2H	-51.8 (-51.3)	8.9 (9.6)	471 (469)	0.10 (0.14)
Z-*S*-(1) + CH_3CO_2H	-50.4 (N/A)	8.5 (N/A)	486 (N/A)	0.09 (N/A)
Z-*S*-(2) + CH_3CO_2H	-52.1 (-51.0)	10.5 (10.0)	457 (469)	0.14 (0.14)

Appendix 4

Chapter 4

Coexpression of Coenzyme F430 Biosynthetic Genes

Supplemental Information

The *in vivo* Biosynthesis of Coenzyme F430 through Heterologous Expression of Multiple Plasmids in *Escherichia coli*

Supplemental Information

Table A4S2. Forward and reverse primers utilized in PCR reactions where restriction sites are incorporated into the PCR product which included NdeI, XhoI, FseI, BspHI, PciI, BamHI, and HindIII.

Gene	Primer
sirC	Forward: 5′-atataccatggctgaaacaaataattttc-3′ Reverse: 5′-atataaagcttttattcgagcttatccg-3′
sirA	Forward: 5′-cggcgcatatgtcagaaaattacgg-3′ Reverse: 5′-atgagctcgagtcagaaatcctttcctgc-3′
cfbA	Forward: 5′-gcgactcatgactgagaaactcgg-3′ Reverse: 5′-attacggatccttacagggcttcctg-3′
cfbB	Forward: 5′-acacaacatgtcccacagcaaacaatc-3′ Reverse: 5′-attaaggatccctaccggggagccc-3′
cfbC	Forward: 5′-cgctgcatatgaaaaaacaaaagatcgttgc-3′ Reverse: 5′-attatggccggccttattttgtcatttccc-3′
cfbD	Forward: 5′-cgccgtcatgactcaaaaagagatctc-3′ Reverse: 5′-atcacaagctttcaggcttcttttgcaac-3′
cfbE	Forward: 5′-gacaccatatggacctgttccgg-3′ Reverse: 5′-cgcacctcgagttaacggaaacatttc-3′

Appendix 5

Chapter 5

The Post-translational Modifications of Methyl-coenzyme M Reductase

Supplemental Information

The Origin and Purpose of Novel Post-translational Modifications Present in the Active Site of Methyl-coenzyme M Reductase

Supporting Information

Table A5S1. Forward and reverse primers utilized in PCR reactions where restriction sites are incorporated into the PCR product which included NdeI, XhoI, FseI, BspHI, PciI, KpnI, BamHI, and HindIII.

Gene	Primer
mcrA	Forward: 5′- gatactcatatgatggcagcagacattttc-3′ Reverse: 5′- atgtgctcgagctagtgatggtgatgatggtgttttgccgggatgacgag-3′
mcrB	Forward: 5′- cgcggcatatgtctgacacagtagacatctacgacg-3′ Reverse: 5′- ttataggccggccttagagcgctcctg-3′
mcrG	Forward: 5′- ttactccatggcatacgaagcacagtattatcc-3′ Reverse: 5′- ttattggatcctcatttcggctggaatcc-3′
mcrC	Forward: 5′- aatctcatatgtcagactctgcttcaaacacg-3′ Reverse: 5′- gctctctcgagtcactcatctttatcagtgtc-3′
mcrD	Forward: 5′- agaattcatgatgatcgaccgggaaacac-3′ Reverse: 5′- cacttggatcctcatgcaactcctttatgatc-3′
mm1	Forward: 5′- atataccatggccgagataaaaattg-3′ Reverse: 5′- agataaagcttcatcttcttttcca-3′
mm10	Forward: 5′- gctcgcatatggaagtagttgtcgacg-3′ Reverse: 5′- atataactcgagttactctagaggcaatcccag-3′
tfuA	Forward: 5′- acagaccatggaaaagaagatgaaag-3′ Reverse: 5′- atataggatcctcaggcctcttcaataag-3′
thiI	Forward: 5′- atataattaatgacaggcaattcc-3′ Reverse: 5′- atataggtacctcagagcttgaggatttttac-3′
prmA	Forward: 5′- ggcgccatatggaaataagatgtaggtg-3′ Reverse: 5′- agctagctcgagtcaaatcacaacaac-3′

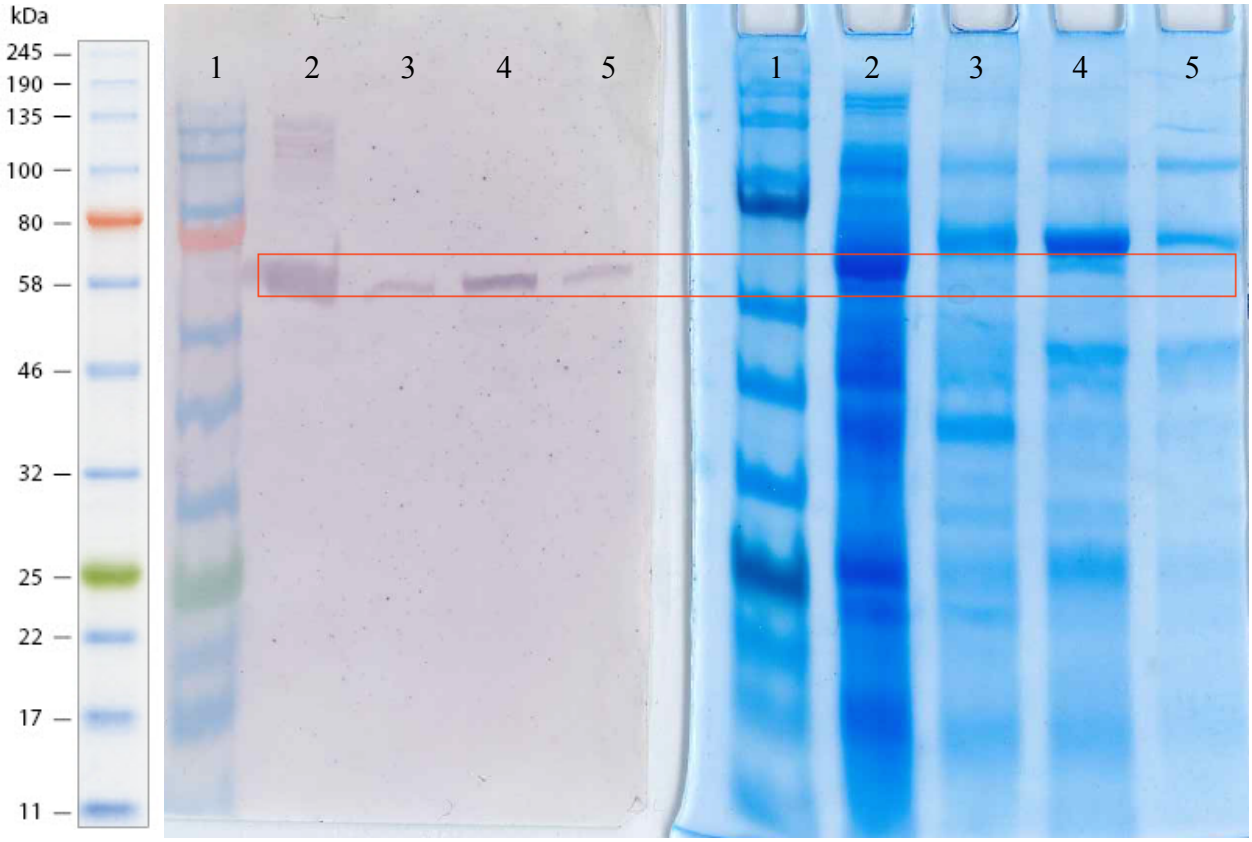


Figure A5S1. Western Blot (Left) and Coomassie Blue stained SDS-PAGE of full-length Mcr α-subunit. Lanes 1-5 on both are as follows: Molecular weight standard, flow through, fraction 1, fraction 2, and fraction 3.

A

```
  1 MAADIFAKFK KSMEVKFTQE YGSNKQAGGD ITGKTEKFLR LGPEQDARKQ
 51 EMIKAGKEIA EKRGIAFYNP MMHMGAPLGQ RAITPYTISG TDIVAEPDDL
101 HYVNNAAMQQ MWDDIRRTCI VGLDMAHETL EKRLGKEVTP ETINHYLETL
151 NHAMPGAAVV QEMMVETHPA LVDDCYVKIF TGDDELADEI DKQYVINVNK
201 MFSEEQAAQI KASIGKTTWQ AIHIPTIVSR TTDGAQTSRW AAMQIGMSFI
251 SAYAMCAGEA AVADLSFAAK HAALVSMGEM LPARRARGPN EPGGLSFGHL
301 SDIVQTSRVS KDPAKIALEV VGAGCMLYDQ IWLGSYMSGG VGFTQYATAA
351 YTDDILDNNT YYDVDYINDK YNGAANLGTD NKVKATLDVV KDIATESTLY
401 GIETYEKFPT ALEDHFGGSQ RATVLAAASG VACALATGNA NAGLSGWYLS
451 MYVHKEAWGR LGFFGFDLQD QCGATNVLSY QGDEGLPDEL RGPNYPNYAM
501 NVGHQGGYAG IAQAAHSGRG DAFTVNPLLK VCFADELMPF NFAEPRREFG
551 RGAIREFMPA GERSLVIPAK
```

B

```
  1 MAADIFAKFK KSMEVKFTQE YGSNKQAGGD ITGKTEKFLR LGPEQDARKQ
 51 EMIKAGKEIA EKRGIAFYNP MMHMGAPLGQ RAITPYTISG TDIVAEPDDL
101 HYVNNAAMQQ MWDDIRRTCI VGLDMAHETL EKRLGKEVTP ETINHYLETL
151 NHAMPGAAVV QEMMVETHPA LVDDCYVKIF TGDDELADEI DKQYVINVNK
201 MFSEEQAAQI KASIGKTTWQ AIHIPTIVSR TTDGAQTSRW AAMQIGMSFI
251 SAYAMCAGEA AVADLSFAAK HAALVSMGEM LPARRARGPN EPGGLSFGHL
301 SDIVQTSRVS KDPAKIALEV VGAGCMLYDQ IWLGSYMSGG VGFTQYATAA
351 YTDDILDNNT YYDVDYINDK YNGAANLGTD NKVKATLDVV KDIATESTLY
401 GIETYEKFPT ALEDHFGGSQ RATVLAAASG VACALATGNA NAGLSGWYLS
451 MYVHKEAWGR LGFFGFDLQD QCGATNVLSY QGDEGLPDEL RGPNYPNYAM
501 NVGHQGGYAG IAQAAHSGRG DAFTVNPLLK VCFADELMPF NFAEPRREFG
551 RGAIREFMPA GERSLVIPAK
```

Figure A5S2. Protein coverage (red) of resultant peptides fragment from trypin digest of MCR α-subunit coexpressed with PrmA enzyme. (A) Peptide fragment containing the 1-*N*-methylhistidine modification highlighted. (B) Peptide fragment containing the *S*-methylcysteine modification highlighted.

References

1. Hastings, J.W. and C.H. Johnson, *Bioluminescence and chemiluminescence.* Methods Enzymol, 2003. **360**: p. 75-104.
2. Martin, D.F. and B.B. Martin, *Red tide, red terror.* J Chem Educ, 1976. **53**(10): p. 614-7.
3. Wang, D.Z., *Neurotoxins from marine dinoflagellates: a brief review.* Mar Drugs, 2008. **6**(2): p. 349-71.
4. Pinto da Silva, L., A.J. Santos, and J.C. Esteves da Silva, *Efficient firefly chemi/bioluminescence: evidence for chemiexcitation resulting from the decomposition of a neutral firefly dioxetanone molecule.* J Phys Chem A, 2013. **117**(1): p. 94-100.
5. Branchini, B.R., et al., *Thermostable red and green light-producing firefly luciferase mutants for bioluminescent reporter applications.* Anal Biochem, 2007. **361**(2): p. 253-62.
6. Abrahams, M. and L. Townsend, *Bioluminescence in dinoflagellate: a test of the burglar alarm hypothesis* Ecology, 1993. **74**: p. 258-260.
7. Fogel, M. and J.W. Hastings, *Bioluminescence: mechanism and mode of control of scintillon activity.* Proc Natl Acad Sci U S A, 1972. **69**(3): p. 690-3.
8. Arrio, B. and B. Lecuyer, *The existence of scintillons in dinoflagellates: an in vitro simulation of the in vivo flashes.* FEBS Lett, 1975. **53**(1): p. 49-52.
9. Schmitter, R., et al., *Dinoflagellate bioluminescence: a comparative study of invitro components.* J Cell Physiol, 1976. **87**(1): p. 123-34.

10. Morse, D., A.J. Pappenheimer, and J. Hastings, *Role of a luciferin-binding protein in the circadian bioluminescent reaction of Gonyaulax polyedra.* J Biol Chem, 1989. **264**(20): p. 11822-6.

11. Hastings, J.W. and J.C. Dunlap, *Cell-free components in dinoflagellate bioluminescence. The particulate activity: Scintillons; the soluble components: Luciferase, luciferin, and luciferin-binding protein.* Methods Enzymol, 1986. **133**: p. 307-327.

12. Morse, D. and M. Mittag, *Dinoflagellate luciferin-binding protein.* Methods Enzymol, 2000. **305**: p. 258-76.

13. Desjardins, M. and D. Morse, *The polypeptide components of scintillons, the bioluminescence organelles of the dinoflagellate Gonyaulax polyedra.* Biochem Cell Biol, 1993. **71**(3-4): p. 176-82.

14. Smith, S.M., et al., *Voltage-gated proton channel in a dinoflagellate.* Proc Natl Acad Sci U S A, 2011. **108**(44): p. 18162-7.

15. Morse, D., et al., *Properties and cellular localization of a luciferin binding protein in the bioluminescence reaction of Gonyaulax polyedra.* J Biolumin Chemilumin, 1989. **3**(2): p. 79-83.

16. Colepicolo, P., et al., *Circadian Regulation of Bioluminescence in the Dinoflagellate Pyrocystis-Lunula.* Journal of Phycology, 1993. **29**(2): p. 173-179.

17. Morse, D., et al., *Circadian regulation of bioluminescence in Gonyaulax involves translational control.* Proc Natl Acad Sci U S A, 1989. **86**(1): p. 172-6.

18. Johnson, C., J. Roeber, and J. Hastings, *Circadian changes in enzyme concentration account for rhythm of enzyme activity in gonyaulax.* Science, 1984. **223**(4643): p. 1428-30.

19. Hastings, J.W., *Chemistry, clones, and circadian control of the dinoflagellate bioluminescent system. The Marlene DeLuca memorial lecture.* J Biolumin Chemilumin, 1989. **4**(1): p. 12-9.

20. Seo, K.S. and L. Fritz, *Cell Ultrastructual Changes Correlate with Circadian Rhythms in Pyrocystis lunula (Pyrrophyta).* J. Phycol., 2000. **36**: p. 351-358.

21. Lee, D.H., et al., *Molecular cloning and genomic organization of a gene for luciferin binding protein from dinoflagellate Gonyaulax polyedra.* J Biol Chem, 1993. **268**: p. 8842-8850.

22. Mittag, M., L. Li, and J. Hastings, *The mRNA level of the circadian regulated Gonyaulax luciferase remains constant over the cycle.* Chronobiol Int, 1998. **15**(1): p. 93-8.

23. Rosemarie, K., et al., *The circadian rhythm of bioluminescence in Pyrocystis is not due to differences in the amount of luciferase: a comparative study of three bioluminescence marine dinoflagellates.* J. Phycol., 1998. **34**: p. 167-172.

24. Mittag, M., D. Lee, and J. Hastings, *Circadian expression of the luciferin-binding protein correlates with the binding of a protein to the 3' untranslated region of its mRNA.* Proc Natl Acad Sci U S A, 1994. **91**(12): p. 5257-61.

25. Hastings, J.W., *Chemistries and colors of bioluminescent reactions: a review.* Gene, 1996. **173**(1 Spec No): p. 5-11.

26. Dunlap, J.C., J.W. Hastings, and O. Shimomura, *Dinoflagellate Luciferin Is Structurally Related to Chlorophyll.* Febs Letters, 1981. **135**(2): p. 273-276.

27. Dunlap, J.C. and J.W. Hastings, *Biochemistry of dinoflagellate bioluminescence: purification and characterization of dinoflagellate luciferin from Pyrocystis lunula.* Biochemistry, 1981. **20**(4): p. 983-9.

28. Schultz, L.W., et al., *Crystal structure of a pH-regulated luciferase catalyzing the bioluminescent oxidation of an open tetrapyrrole.* Proc Natl Acad Sci U S A, 2005. **102**(5): p. 1378-83.

29. Nakamura, H., et al., *Structure of Dinoflagellate Luciferin and Its Enzymatic and Nonenzymatic Air-Oxidation Products.* J Am Chem Soc, 1989. **111**(19): p. 7607-7611.

30. Stojanovic, M.N. and Y. Kishi, *Dinoflagellate biolumeniscence: the chemical behavior of the chromophore towards oxidants.* Tetrahedron Letters, 1994. **35**: p. 9347-9350.

31. Wilson, T. and J.W. Hastings, *Bioluminescence.* Annu Rev Cell Dev Biol, 1998. **14**: p. 197-230.

32. Wu, C., H. Akimoto, and Y. Ohmiya, *Tracer Studies on dinoflagellate luciferin with [^{15}N]-glycine and [^{15}N]-L-glutamic acid in the dinoflagellate Pyrocystis lunula.* Tetrahedron Letters, 2003. **44**: p. 1263-1266.

33. Takamiya, K., T. Tsuchiya, and H. Ohta, *Degradation pathway(s) of chlorophyll: what has gene cloning revealed?* Trends in Plant Science, 2000. **5**(10): p. 426-431.

34. Hortensteiner, S., *Chlorophyll degradation during senecence.* Annu. Rev. Plant. Biol., 2006. **57**: p. 55-77.

35. Topalov, G. and Y. Kishi, *Chlorophyll catabolism leading to the skeleton of Dinoflagellate and Krill luciferins: hypothesis and model studies.* Angew. Chem. Int. Ed., 2001. **40**(20): p. 3892-3894.

36. Liu, L., T. Wilson, and J.W. Hastings, *Molecular evolution of dinoflagellate luciferases, enzymes with three catalytic domains in a single polypeptide.* Proc Natl Acad Sci U S A, 2004. **101**(47): p. 16555-60.

37. Li, L., R. Hong, and J.W. Hastings, *Three functional luciferase domains in a single polypeptide chain.* Proc Natl Acad Sci U S A, 1997. **94**(17): p. 8954-8.

38. Li, L. and J.W. Hastings, *The structure and organization of the luciferase gene in the photosynthetic dinoflagellate Gonyaulax polyedra.* Plant Mol Biol, 1998. **36**(2): p. 275-84.

39. Okamoto, O.K., et al., *Members of a dinoflagellate luciferase gene family differ in synonymous substitution rates.* Biochemistry, 2001. **40**(51): p. 15862-8.

40. Li, L., et al., *N-terminal intramolecularly conserved histidines of three domains in Gonyaulax luciferase are responsible for loss of activity in the alkaline region.* Biochemistry, 2001. **40**(6): p. 1844-9.

41. Suzuki-Ogoh, C., C. Wu, and Y. Ohmiya, *C-terminal region of the active domain enhances enzymatic activity in dinoflagellate luciferase.* Photochem Photobiol Sci, 2008. **7**(2): p. 208-11.

42. Krieger, N. and J. Hastings, *Bioluminescence: pH activity profiles of related luciferase fractions.* Science, 1968. **161**(841): p. 586-9.

43. Fresneau, C. and B. Arrio, *Pyrocystis lunula bioluminescence: physicochemical characterization of the luciferin precursor.* Arch Biochem Biophys, 1988. **265**(1): p. 22-7.

44. Fresneau, C., et al., *Dinoflagellate luminescence: purification of a NAD(P)H-dependent reductase and of its substrate.* Arch Biochem Biophys, 1986. **251**(2): p. 495-503.

45. Fresneau, C., et al., *Demonstration of the Peptidic Nature of Luciferin from Pyrocystis-Lunula.* International Journal of Peptide and Protein Research, 1988. **31**(2): p. 126-129.

46. Krautler, B., *Chlorophyll breakdown and chlorophyll catabolites in leaves and fruit.* Photochem Photobiol Sci, 2008. **7**(10): p. 1114-20.

47. Rodoni, S., et al., *Chlorophyll Breakdown in Senescent Chloroplasts (Cleavage of Pheophorbide a in Two Enzymic Steps).* Plant Physiol, 1997. **115**(2): p. 669-676.

48. Suzuki, Y., M. Doi, and Y. Shioi, *Two enzymatic reaction pathways in the formation of pyropheophorbide a.* Photosynth Res, 2002. **74**(2): p. 225-33.

49. Li, Y., Z.L. Cai, and M. Chen, *Spectroscopic properties of chlorophyll f.* J Phys Chem B, 2013. **117**(38): p. 11309-17.

50. Cussatlegras, A.-S. and P.L. Gal, *Variability in the bioluminescence response of the dinoflagellate Pyrocystis lunula.* Journal of Experimental Marine Biology and Ecology, 2007. **343**: p. 74-81.

51. Morishita, H., et al., *Cloning and characterization of an active fragment of luciferase from a luminescent marine alga, Pyrocystis lunula.* Photochem Photobiol, 2002. **75**(3): p. 311-5.

52. Liu, L. and J. Hastings, *Two different domains of the luciferase gene in the heterotrophic dinoflagellate Noctiluca scintillans occur as two separate genes in photosynthetic species.* Proc Natl Acad Sci U S A, 2007. **104**(3): p. 696-701.

53. Lecuyer, B., et al., *Dinoflagellate luciferases: purification of luciferases from Gonyaulax polyedra, Pyrocystis lunula, and Pyrocystis fusiformis.* Arch Biochem Biophys, 1979. **196**(2): p. 371-84.

54. Williams, S.L., C.A. de Oliveira, and J.A. McCammon, *Coupling Constant pH Molecular Dynamics with Accelerated Molecular Dynamics.* J Chem Theory Comput, 2010. **6**(2): p. 560-568.

55. Baptista, A.M., V.H. Teixeira, and C.M. Soares, *Constant-pH molecular dynamics using stochastic titration.* Journal of Chemical Physics, 2002. **117**(9): p. 4184-4200.

56. Gotz, A.W., et al., *Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born.* J Chem Theory Comput, 2012. **8**(5): p. 1542-1555.

57. Goh, G.B., et al., *Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism.* Proteins, 2014. **82**(7): p. 1319-31.

58. Mongan, J., D.A. Case, and J.A. McCammon, *Constant pH molecular dynamics in generalized Born implicit solvent.* J Comput Chem, 2004. **25**(16): p. 2038-48.

59. Pierce, L.C., et al., *Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics.* J Chem Theory Comput, 2012. **8**(9): p. 2997-3002.

60. Hamelberg, D., J. Mongan, and J.A. McCammon, *Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules.* Journal of Chemical Physics, 2004. **120**(24): p. 11919-11929.

61. DeVane, R., et al., *A molecular dynamics method for calculating molecular volume changes appropriate for biomolecular simulation.* Biophys J, 2003. **85**(5): p. 2801-7.

62. Braslavsky, S.E. and G.E. Heibel, *Time-Resolved Photothermal and Photoacoustic Methods Applied to Photoinduced Processes in Solution.* Chemical Reviews, 1992. **92**(6): p. 1381-1410.

63. Miksovska, J. and R.W. Larsen, *Photothermal studies of pH induced unfolding of apomyoglobin.* J Protein Chem, 2003. **22**(4): p. 387-94.

64. Yip, R.W., et al., *Photochemistry of the Ortho-Nitrobenzyl System in Solution - Identification of the Biradical Intermediate in the Intramolecular Rearrangement.* Journal of Physical Chemistry, 1991. **95**(16): p. 6078-6081.

65. Taylor, F.J.R.H., M.; Saldarriaga, J. F. , *Dinoflagellate diversity and distribution.* Biodivers. Conserv., 2008 **17**(2): p. 407–418.

66. Biggley, W., et al., *Stimulable and spontaneous bioluminescence in the marine dinoflagellates, Pyrodinium bahamense, Gonyaulax polyedra, and Pyrocystis lunula.* J Gen Physiol, 1969. **54**(1): p. 96-122.

67. Morse DS, F.L., Hastings JW., *What is the clock? Translational regulation of circadian bioluminescence.* Trends Biochem Sci., 1990. **15**(7): p. 262-265.

68. Chai, J.D. and M. Head-Gordon, *Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections.* Phys Chem Chem Phys, 2008. **10**(44): p. 6615-20.

69. Frisch, M.J.e.a., Gaussian 09, 2013. Revision D.01; Gaussian, Inc.: Wallingford, CT.

70. Alipour, M., *How Does LCDFT Compare to SAC-CI for the Treatment of Valence and Rydberg Excited States of Organic Compounds?* Journal of Physical Chemistry A, 2014. **118**(9): p. 1741-1747.

71. Stojanovic, M.N. and Y. Kishi, *Dinoflagellate bioluminescence: the chromophore of dinoflagellate luciferin.* Tetrahedron Letters, 1994. **35**: p. 9343-9346.

72. Tomasi, J., B. Mennucci, and R. Cammi, *Quantum mechanical continuum solvation models.* Chem Rev, 2005. **105**(8): p. 2999-3093.

73. Koo, J.A., S.P. Schmidt, and G.B. Schuster, *Bioluminescence of the firefly: key steps in the formation of the electronically excited state for model systems.* Proc Natl Acad Sci U S A, 1978. **75**(1): p. 30-3.

74. Koo, J. and G. Schuster, *Chemiluminescence of diphenoyl peroxide. Chemically initiated electron exchange luminescence. A new general mechanism for chemical production of electronically excited states.* J. Am. Chem. Soc., 1978. **100**(14): p. 4496-4503.

75. Grabowski, Z.R., K. Rotkiewicz, and W. Rettig, *Structural changes accompanying intramolecular electron transfer: focus on twisted intramolecular charge-transfer states and structures.* Chem Rev, 2003. **103**(10): p. 3899-4032.

76. Sanders, J.K. and S.E. Jackson, *The discovery and development of the green fluorescent protein, GFP.* Chem Soc Rev, 2009. **38**(10): p. 2821-2.

77. Kajiyama, N. and E. Nakano, *Isolation and characterization of mutants of firefly luciferase which produce different colors of light.* Protein Eng, 1991. **4**(6): p. 691-3.

78. Vu, L.T., et al., *Changing blue fluorescent protein to green fluorescent protein using chemical RNA editing as a novel strategy in genetic restoration.* Chem Biol Drug Des, 2015. **86**(5): p. 1242-52.

79. Nakatani, N., J.Y. Hasegawa, and H. Nakatsuji, *Red light in chemiluminescence and yellow-green light in bioluminescence: color-tuning mechanism of firefly, Photinus pyralis, studied by the symmetry-adapted cluster-configuration interaction method.* J Am Chem Soc, 2007. **129**(28): p. 8756-65.

80. Thauer, R.K., et al., *Methanogenic archaea: ecologically relevant differences in energy conservation.* Nat Rev Microbiol, 2008. **6**(8): p. 579-91.

81. Won, H.S., MF; Olson, KD; Wolfe, RS, *Two-dimensional NMR studies of native coenzyme F430.* J. Am. Chem. Soc., 1990. **112**(6): p. 2178-84.

82. Ermler, U., et al., *Crystal structure of methyl-coenzyme M reductase: the key enzyme of biological methane formation.* Science, 1997. **278**(5342): p. 1457-62.

83. Thauer, R.K., *Biochemistry of methanogenesis: a tribute to Marjory Stephenson. 1998 Marjory Stephenson Prize Lecture.* Microbiology, 1998. **144 (Pt 9)**: p. 2377-406.

84. Wongnate, T., et al., *The radical mechanism of biological methane synthesis by methyl-coenzyme M reductase.* Science, 2016. **352**(6288): p. 953-8.

85. Grabarse, W., et al., *Comparison of three methyl-coenzyme M reductases from phylogenetically distant organisms: unusual amino acid modification, conservation and adaptation.* J Mol Biol, 2000. **303**(2): p. 329-44.

86. Goubeaud, M., G. Schreiner, and R.K. Thauer, *Purified methyl-coenzyme-M reductase is activated when the enzyme-bound coenzyme F430 is reduced to the nickel(I) oxidation state by titanium(III) citrate.* Eur J Biochem, 1997. **243**(1-2): p. 110-4.

87. Kahnt, J., et al., *Post-translational modifications in the active site region of methyl-coenzyme M reductase from methanogenic and methanotrophic archaea.* Febs Journal, 2007. **274**(18): p. 4913-4921.

88. Prakash, D., et al., *Elucidating the process of activation of methyl-coenzyme M reductase.* J Bacteriol, 2014. **196**(13): p. 2491-8.

89. Shima, S. and R.K. Thauer, *Methyl-coenzyme M reductase and the anaerobic oxidation of methane in methanotrophic Archaea.* Current Opinion in Microbiology, 2005. **8**(6): p. 643-648.

90. Zheng, K., Ngo, P.D., Owens, V.L., Yang, X., Mansoorabadi, S.O., *The biosynthetic pathway of coenzyme F430 in methanogenic and methanotrophic archaea.* Science, 2016. **354**(6310): p. 339-342.

91. Ferry, J.G., *Methane: small molecule, big impact.* Science, 1997. **278**(5342): p. 1413-4.

92. Ferry, J.G., *Enzymology of one-carbon metabolism in methanogenic pathways.* FEMS Microbiol Rev, 1999. **23**(1): p. 13-38.

93. Energy, U.S.D.o., *Modern Shale Gas Development in the United States.* Office of Fossil Energy, National Energy Technology Laboratory, 2009.

94. Caldwell, S.L., et al., *Anaerobic oxidation of methane: mechanisms, bioenergetics, and the ecology of associated microorganisms.* Environ Sci Technol, 2008. **42**(18): p. 6791-9.

95. Hallam, S.J., et al., *Reverse methanogenesis: testing the hypothesis with environmental genomics.* Science, 2004. **305**(5689): p. 1457-62.

96. Knittel, K. and A. Boetius, *Anaerobic oxidation of methane: progress with an unknown process.* Annu Rev Microbiol, 2009. **63**: p. 311-34.

97. Mayr, S., et al., *Structure of an F430 variant from archaea associated with anaerobic oxidation of methane.* J Am Chem Soc, 2008. **130**(32): p. 10758-67.

98. Scheller, S., et al., *Methyl-coenzyme M reductase from methanogenic archaea: isotope effects on label exchange and ethane formation with the homologous substrate ethyl-coenzyme M.* J Am Chem Soc, 2013. **135**(40): p. 14985-95.

99. Scheller, S., et al., *Methyl-coenzyme M reductase from methanogenic archaea: isotope effects on the formation and anaerobic oxidation of methane.* J Am Chem Soc, 2013. **135**(40): p. 14975-84.

100. Ellermann, J., et al., *The final step in methane formation. Investigations with highly purified methyl-CoM reductase (component C) from Methanobacterium thermoautotrophicum (strain Marburg).* Eur J Biochem, 1988. **172**(3): p. 669-77.

101. Dunbar, K.L., J.O. Melby, and D.A. Mitchell, *YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations.* Nature Chemical Biology, 2012. **8**(6): p. 569-575.

102. Dunbar, K.L., et al., *Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis.* Nat Chem Biol, 2014. **10**(10): p. 823-9.

103. Breil, B., J. Borneman, and E.W. Triplett, *A newly discovered gene, tfuA, involved in the production of the ribosomally synthesized peptide antibiotic trifolitoxin.* J Bacteriol, 1996. **178**(14): p. 4150-6.

104. Palenchar, P.M., et al., *Evidence that ThiI, an enzyme shared between thiamin and 4-thiouridine biosynthesis, may be a sulfurtransferase that proceeds through a persulfide intermediate.* J Biol Chem, 2000. **275**(12): p. 8283-6.

105. Soo, V.W.C., et al., *Reversing methanogenesis to capture methane for liquid biofuel precursors.* Microbial Cell Factories, 2016. **15**.

106. Shima, S., et al., *Structure of a methyl-coenzyme M reductase from Black Sea mats that oxidize methane anaerobically.* Nature, 2012. **481**(7379): p. 98-101.

107. Haynes, C.A. and R. Gonzalez, *Rethinking biological activation of methane and conversion to liquid fuels.* Nat Chem Biol, 2014. **10**(5): p. 331-9.

108. Thauer, R.K., *Anaerobic oxidation of methane with sulfate: on the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO2.* Current Opinion in Microbiology, 2011. **14**(3): p. 292-299.

109. Mueller, T.J., et al., *Methane oxidation by anaerobic archaea for conversion to liquid fuels.* J Ind Microbiol Biotechnol, 2015. **42**(3): p. 391-401.

Additional Primary Efforts

A version of this manuscript was published in the Journal of Physical Chemistry

Ngo, P.D., S.O. Mansoorabadi, and P.A. Frey, *Serine Protease Catalysis: A Computational Study of Tetrahedral Intermediates and Inhibitory Adducts.* J Phys Chem B, 2016. 120(30): p. 7353-9.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Serine Protease Catalysis: A Computational Study of Tetrahedral Intermediates and Inhibitory Adducts

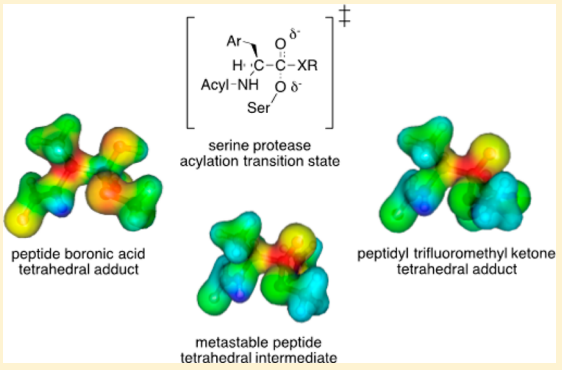
Phong D. Ngo,[†] Steven O. Mansoorabadi,*[,†] and Perry A. Frey*[,‡]

[†]Department of Chemistry and Biochemistry, Auburn University, 179 Chemistry Building, Auburn, Alabama 36849, United States

[‡]Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706, United States

Supporting Information

ABSTRACT: Peptide boronic acids and peptidyl trifluoromethyl ketones (TFKs) inhibit serine proteases by forming monoanionic, tetrahedral adducts to serine in the active sites. Investigators regard these adducts as analogs of monoanionic, tetrahedral intermediates. Density functional theory (DFT) calculations and fractional charge analysis show that tetrahedral adducts of model peptidyl TFKs are structurally and electrostatically very similar to corresponding tetrahedral intermediates. In contrast, the DFT calculations show the structures and electrostatic properties of analogous peptide boronate adducts to be significantly different. The peptide boronates display highly electrostatically positive boron, with correspondingly negative ligands in the tetrahedra. In addition, the computed boron−oxygen and boron−carbon bond lengths in peptide boronates (which are identical or very similar to the corresponding bonds in a peptide boronate adduct of α-lytic protease determined by X-ray crystallography at subangstrom resolution) are significantly longer than the corresponding bond lengths in model tetrahedral intermediates. Since protease-peptidyl TFKs incorporate low-barrier hydrogen bonds (LBHBs) between an active site histidine and aspartate, while the protease-peptide boronates do not, these data complement the spectroscopic and chemical evidence for the participation of LBHBs in catalysis by serine proteases. Moreover, while the potency of these classes of inhibitors can be correlated to the structures of the peptide moieties, the present results indicate that the strength of their bonds to serine contribute significantly to their inhibitory properties.



INTRODUCTION

Serine and threonine proteases such as chymotrypsin (Cht) catalyze hydrolytic cleavage of peptide bonds in proteins. Many other proteases including thrombin, elastase, subtilisin, α-lytic protease (αLP), and the proteasome fall into this class. The active sites include the side chains of serine or threonine and histidine (Ser195 and His57 in chymotrypsin), as well as a site for a carbonyl or anionic oxygen. The β-hydroxyl group of serine or threonine and the imidazole ring of histidine catalyze peptide hydrolysis by multistep mechanisms. The substrate acylates the β-hydroxyl group of serine or threonine in two distinct steps as shown in Scheme 1. The transient tetrahedral intermediate is, based on the Hammond postulate,[1] thought to be structurally similar to the transition state. Completion of the catalytic cycle proceeds by hydrolysis of the acyl-enzyme intermediate by a water molecule.

In 1970−71 investigators discovered inhibition of the serine protease chymotrypsin by alkane- and 2-phenylethaneboronic acids. They attributed this to the formation of anionic tetrahedral boronate adducts to Ser195 in the active site and suggested that the adducts resembled the transient tetrahedral intermediate and transition state (Scheme 2).[2,3]

Simple alkaneboronic acids inhibited serine proteases with millimolar dissociation (inhibition) constants (K_i). Because of the biomedical importance of serine and threonine proteases, much research centered around development of oligopeptide boronic acids with high inhibitory potency. In an early example, MeO-Succ-Ala-Ala-Pro-boroPhe-OH displayed a K_i value of 0.16 nM against chymotrypsin.[4] Extensive research on thrombin has led to clinical trials of several inhibitors.[5] The threonine-specific Bortezomib (Chart 1), a proteasome inhibitor, has been approved by the FDA for treatment of multiple myeloma.[5]

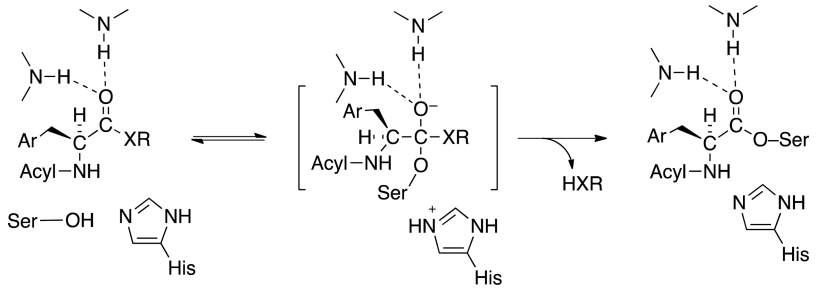
In another class of serine protease inhibitors, the leaving group in a peptide substrate is replaced by the trifluoromethyl group to form a peptidyl trifluoromethyl ketone (TFK).[6−14] The keto group in a peptidyl TFK reacts with the β-hydroxyl group of serine to form an adduct analogous to the tetrahedral intermediate of a substrate (Scheme 3). Because $-CF_3$ cannot leave, the reaction is arrested at the adduct, which is stable and inhibitory.

Boronate and TFK adducts to active site serine in serine proteases share similarities with tetrahedral intermediates in that they include tetrahedral, monoanionic centers. Yet boron

Received: April 22, 2016
Revised: June 21, 2016
Published: July 7, 2016

DOI: 10.1021/acs.jpcb.6b04089
J. Phys. Chem. B 2016, 120, 7353−7359

Scheme 1. Mechanism of Acylation of Serine in Serine Proteases



Scheme 2. Reaction of Boronic Acids with Serine in Serine Proteases

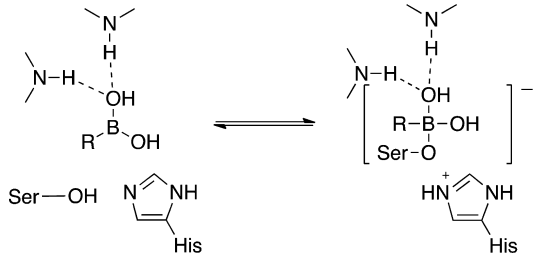
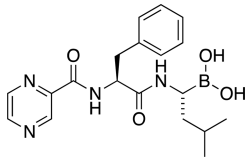
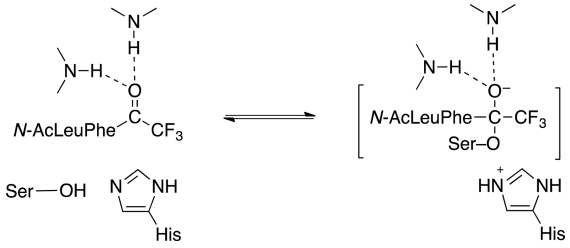


Chart 1. Structure of Bortezomib, an FDA-Approved Drug for Multiple Myeloma



Scheme 3. Reaction of a Peptidyl Trifluoromethyl Ketone with Serine in a Serine Protease



and carbon are chemically different, and the adducts incorporate different electrostatic and molecular properties. The TFK adducts retain the oxyanionic group of the tetrahedral intermediate. In contrast, the boronate complexes incorporate a hydroxyl group in place of the oxyanion, as proven by an ultra-high-resolution X-ray crystal structure of the MeO-Succ-Ala-Ala-Pro-boroVal-OH adduct in α-lytic protease.[15] It thus seems unlikely that peptide boronate and peptidyl TFK adducts could both similarly mimic transition states. The present work examines the similarities and differences among models of tetrahedral intermediates and analogous boronate and TFK adducts. The models are methoxide adducts of *N*-formyl-L-alanine (fAla) methylamide, fAla methyl ester, fAla boronic acid, fAla boronic acid monomethyl ester, and fAla TFK.

COMPUTATIONAL METHODOLOGY

In an effort to elucidate the covalent interactions of serine-β–OH with model peptide and ester substrates, and peptidyl TFK and peptide boronic acid inhibitors, the bond dissociation enthalpies and Gibbs free energies of the heterolytic and homolytic cleavage of the C–O and B–O bonds of the respective model tetrahedral species were calculated using Gaussian09.[16] Geometry optimizations were executed with density functional theory (DFT) using the Becke-style three-parameter Lee–Yang–Parr correlation functional (B3LYP) and Pople's diffuse polarized triple-ζ 6-311+G(d,p) basis set.[17,18] Opt = Tight and Int = UltraFine convergence criteria were utilized for all calculations. Gas-phase bond dissociation enthalpies and Gibbs free energies were obtained by performing vibrational frequency calculations on the optimized geometries using the same B3LYP/6-311+G(d,p) scheme at 298.15 K and 1.0000 atm, with a scale factor of 0.9877.[19] Aqueous-phase bond dissociation enthalpy and Gibbs free energy calculations were also performed on the optimized geometries with an identical B3LYP/6-311+G(d,p) method and basis set using polarizable continuum model (PCM) solvation with water as the solvent.[20] This computational methodology was chosen because it has been previously applied to the analysis of enzymatic and nonenzymatic acyltransfer reactions and was shown to provide reasonably accurate geometries and energies.[21−23] Estimates for atomic partial charges were obtained from (1) Mulliken population analysis (MPA), (2) the electrostatic potential (ESP) using the Merz–Singh–Kollman scheme, and (3) natural bond orbital (NBO) analysis.[24−26]

RESULTS

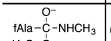
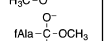
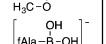
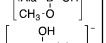
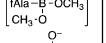
1. Anionic Tetrahedral Adducts of *N*-Formyl-L-Alanine Derivatives. Tetrahedral intermediates such as in Scheme 1 cannot be isolated for experimental comparison with peptide boronates or peptidyl TFKs in Schemes 2 and 3. They can, however, be compared through computational methods such as density functional theory (DFT).[27] For this purpose, the model tetrahedral intermediates adopted here are methoxide adducts of *N*-formyl-L-alanine (fAla) methyl ester and of fAla methylamide. The same computational methods applied to the corresponding methoxide adducts of fAla boronic acids and fAla TFK allow comparisons to be made.

Complete results of DFT calculations on methoxide adducts of fAla methyl ester and fAla methylamide appear in the Supporting Information, together with comparable results on the corresponding fAla TFK and fAla boronic acids. Relevant covalent bond lengths in the tetrahedral centers appear in Table 1. The calculated bond lengths refer to aqueous solutions and

DOI: 10.1021/acs.jpcb.6b04089
J. Phys. Chem. B 2016, 120, 7353−7359

are similar to those in the gas phase (see Supporting Information).

Table 1. Calculated Bond Lengths of Peripheral Groups to C and B in Tetrahedral Adducts of Peptide Substrates, Peptidyl TFK, and Peptide Boronates

Adduct	Bond Length (Å)			
	C-Cα 1.58	C-O⁻ 1.30	C-NHCH$_3$ 1.50	C-OCH$_3$ 1.52
	C-Cα 1.57	C-O⁻ 1.28	C-OCH$_3$ 1.51	C-OCH$_3$ 1.48
	B-Cα 1.65	B-OH 1.49	B-OH 1.50	B-OCH$_3$ 1.48
	B-Cα 1.65	B-OH 1.49	B-OCH$_3$ 1.49	B-OCH$_3$ 1.49
	C-Cα 1.58	C-O⁻ 1.30	C-CF$_3$ 1.58	C-OCH$_3$ 1.51

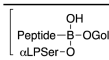
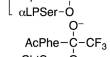
[a]The quoted bond lengths refer to the aqueous phase. Computed bond lengths in the gas phase are similar. A complete listing can be found in the Supporting Information.

2. Experimental and Computed Structures. To be reliable, computed structures must be similar to experimental structures. As noted, transient tetrahedral intermediates cannot be isolated and structurally characterized for comparison with computed structures. However, computed structures of methoxide adducts to fAla boronic acids and fAla TFK can be compared with experimental structures of inhibitory tetrahedral adducts of serine proteases determined by X-ray crystallography. Bond lengths at the tetrahedral centers are of particular interest.

Table 2 lists the bond lengths about the tetrahedral centers in the adducts of MeO-Succ-Ala-Ala-Pro-boroVal in the active site of α-lytic protease and of *N*-AcPhe TFK in the active site of chymotrypsin.[14,15] The experimental bond lengths in Table 2 are identical or very similar to the corresponding computed bond lengths for the models in Table 1, well within experimental uncertainty. Moreover, the computed bond lengths in methoxide adducts to fAla boronic acids are also nearly the same as comparable bonds in experimentally determined crystal structures of small molecule alkyl boronates.[28,29] These relationships support the computational methods and inspire confidence in the computed structures of methoxide adducts of fAla methyl ester, fAla methylamide, and fAla TFK.

The calculated B−Cα and B−OH lengths in the boronates are significantly longer than the calculated C−Cα and C−O⁻ lengths in the model tetrahedral intermediates. The corresponding calculated bond lengths in the model fAla TFK adduct are similar to those in the model tetrahedral intermediates.

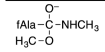
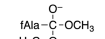
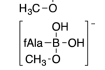
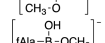
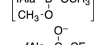
Table 2. Experimental Bond Lengths of Peripheral Groups to C and B in Enzyme-Inhibited Tetrahedral Adducts[a]

Adduct	Bond Length (Å)			
	B-Cα 1.61	B-OH 1.48	B-OSer 1.47	B-OGol 1.52
	C-Cα 1.55	C-O⁻ 1.40	C-CF$_3$ 1.57	C-OSer 1.42

[a]In the first line, the peptide inhibitor is MeO-Succ-Ala-Ala-Pro-boroVal, the enzyme is α-lytic protease (αLP) and the published structure is at a resolution of 0.89 Å.[15] In the second line, the inhibitor is *N*-AcPhe-TFK, the enzyme is chymotrypsin (Cht), and the published crystal structure is at a resolution of 1.4 Å.[14].

3. Strength of Bonds in Tetrahedral Adducts. The bond linking Ser195 to boron in a serine protease-boronate complex can be expected to be strong because the electronegativity difference between boron and oxygen leads to high polarity similar to that in the C−F bond, the strongest single bond in organic chemistry. In the model tetrahedral adducts, methoxyl groups serve as surrogates for the β-hydroxyl group of Ser195. Therefore, the relative strengths of the C−OCH$_3$ and B−OCH$_3$ bonds in the tetrahedral adducts are of interest. The calculated enthalpies and free energies for homolytic and heterolytic cleavage of these bonds appear in Table 3.

Table 3. Calculated Homolytic and Heterolytic Bond Dissociation Enthalpies and Free Energies of Tetrahedral Adducts of Peptide Substrates, Peptidyl TFK, and Peptide Boronates in Gas and Aqueous Phases

Cleavage	Homolytic (kcal/mol)				Heterolytic (kcal/mol)			
Phase	Gas		Aqueous		Gas		Aqueous	
Adduct	Δ*H*	Δ*G*	Δ*H*	Δ*G*	Δ*H*	Δ*G*	Δ*H*	Δ*G*
	69.4	55.2	68.0	54.7	22.0	8.5	-6.5	-19.1
	74.6	61.3	70.2	56.7	32.6	19.9	5.7	-7.2
	111	98.3	118	104	54.7	43.4	26.5	14.7
	104	90.9	105	91.6	57.7	45.3	28.1	15.9
	61.7	47.2	58.0	44.3	56.5	43.6	26.7	13.8

Homolysis of the bond linking the β-oxygen of Ser195 to tetrahedral carbon does not occur in the reactions of serine proteases. However, for purposes of comparing the basic chemistry of tetrahedral adducts, the energies of homolytic cleavage of the methoxyl group in the models are of interest. The calculations show that the homolytic B−OCH$_3$ cleavages display very positive enthalpies and free energies (Δ*H* = +104−118 kcal mol^{-1} and Δ*G* = +91−104 kcal mol^{-1}) in the gas and aqueous phases, respectively. In contrast, homolysis energies of the C−OCH$_3$ bonds in the model tetrahedral intermediates are much less positive (Δ*H* = +68−75 kcal mol^{-1} and Δ*G* = +55−61 kcal mol^{-1}). In the case of the tetrahedral adduct of fAla TFK, C−OCH$_3$ homolysis energies are Δ*H* = +58−62 kcal mol^{-1} and Δ*G* = +44−47 kcal mol^{-1}. From the standpoint of homolysis (i.e., bond dissociation energy), the methoxide adduct of fAla TFK is more similar to the tetrahedral intermediates than the fAla boronates.

In heterolytic scission, calculated energies for departure of the methoxide ion from the model tetrahedral intermediates and the model boronate or TFK adducts are strikingly different. The calculations do not refer to an enzymatic process, which is acid catalyzed with proton donation, so that the leaving group would be methanol. However, because the objective is to compare intrinsic chemical relations among aminoacyl tetrahedral intermediates and peptide boronate and peptidyl TFK adducts, simple dissociation of methoxide serves the purpose.

Focusing on the aqueous enthalpies and free energies in Table 3, methoxide departure is very highly exothermic and spontaneous for the tetrahedral fAla methylamide intermediate, with Δ*H* = −6.5 kcal mol^{-1} and Δ*G* = −19 kcal mol^{-1}. This is

DOI: 10.1021/acs.jpcb.6b04089
J. Phys. Chem. B 2016, 120, 7353−7359

perfectly consistent with the tetrahedral intermediate being metastable for enzymatic and nonenzymatic reactions. Reaction of the model tetrahedral intermediate for fAla methyl ester is less spontaneous (ΔH = 5.7 kcal mol^{-1} and ΔG = −7.2 kcal mol^{-1}).

The heterolytic B−OCH_3 cleavages in the fAla boronates are highly endothermic (ΔH = +27−28 kcal mol^{-1} and ΔG = +15−16 kcal mol^{-1}). The free energy for heterolytic cleavage of C−OCH_3 in the fAla TFK adduct is also highly endothermic (ΔH = +27 kcal mol^{-1} and ΔG = +14 kcal mol^{-1}). Methoxide binds slightly more strongly to boron in the boronate adducts than to carbon in the methoxyl adduct of fAla TFK. However, both display very strong bonding to the methoxyl group, unlike the methoxide adducts to fAla methyl ester and fAla methylamide. Therefore, bonding of methoxide in the model boronate and TFK adducts do not share energetic similarities with tetrahedral intermediates or transition states of acyltransfer in either enzymatic or nonenzymatic processes.

4. Localized and Dispersed Anionic Charge in Tetrahedral Adducts. As Lewis acids, boronic acids react to form adducts with Lewis bases such as alcohols or water. These reactions can be formulated as in Scheme 4, where a reactive electron pair in R′-OH forms a covalent bond to boron in the boronic acid.

Scheme 4. Reaction of a Boronic Acid with an Alcohol

Many investigators show the formal charge from the Lewis base centered on boron in the anionic adduct.[5,15] In the ultrahigh resolution structure of the complex of MeO-Succ-Ala-Ala-Pro-boroVal-OH with α-lytic protease, Fuhrmann et al. interpret the electron density to indicate excess negative charge on boron.

All of the model tetrahedral adducts investigated here carry one negative charge. However, the charges do not localize to the same locations in the adducts. The electrostatic landscapes can be described as in the following paragraph and illustrated in Figure 1, which graphically displays the distribution of negative charge in the methoxide adducts as electrostatic potential surface maps.[30] Gradations of charge range from blue for the most electropositive regions of the molecule to red for the most negative charge. The images show that the electrostatic landscape of the fAla TFK adduct is more similar to both model tetrahedral intermediates than are either of the fAla boronates.ised

One can also compare the distribution of charge in molecules by calculating partial charges on atoms and groups. These calculations give ranges of values for individual atoms or groups depending upon the computational methods employed. The results of three methods applied to the tetrahedral monoanions studied here appear in the Supporting Information. The methods are the Mulliken population analysis (MPA),[24] the Merz−Singh−Kollman scheme for deriving partial charges from the electrostatic potential (ESP),[25] and the natural bond orbital analysis (NBO).[26] The quantitative results vary among these methods, but all three methods reveal similar overall charge localization and dispersal.

All three methods show the central boron in both fAla boronates to be more electrostatically positive (with average partial charges of +1.0) than carbon in either model tetrahedral intermediate (+0.4 and +0.5 for fAla methylamide and fAla methyl ester, respectively) or in the methoxide adduct of fAla TFK (+0.1). As a consequence, the peripheral ligands to boron in the fAla boronates are more electrostatically negative than the corresponding ligands in the tetrahedral intermediates or the methoxide adduct of fAla TFK, and the negative charge is dispersed throughout the outer ligands of the tetrahedral center (Figure 1).

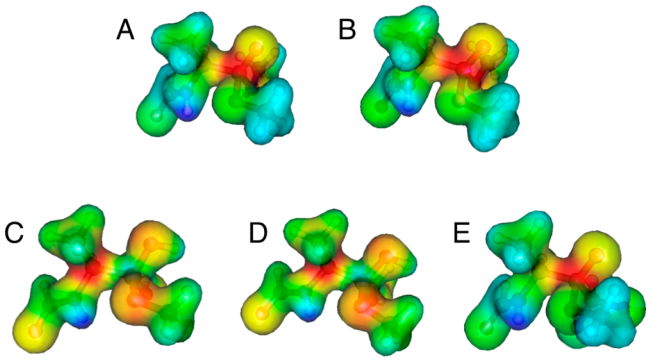


Figure 1. Electrostatic potential surface map of the methoxide adducts of (A) *N*-formyl-L-alanine methylamide, (B) *N*-formyl-L-alanine methyl ester, (C) *N*-formyl-L-alanineboronic acid, (D) *N*-formyl-L-alanineboronic acid monomethyl ester, and (E) *N*-formyl-L-alanine-TFK. Maps were generated using Molekel 5.4.0.8.[30]

DISCUSSION

The computational results describe significant differences in covalent structure and distribution of negative charge in methoxide adducts of fAla boronic acids compared with methoxide adducts of fAla methyl ester and fAla methylamide, which serve as models of tetrahedral intermediates. In contrast, the methoxide adduct of fAla TFK bears more similarities to the model tetrahedral intermediates. On this basis, the complexes of active site serine in serine proteases to peptidyl TFKs are better analogs of tetrahedral intermediates in protease catalysis than are the corresponding complexes with peptide boronic acids.

Certain proteases, such as HIV protease, catalyze peptide hydrolysis by way of tetrahedral intermediates without the intervention of serine or any enzymatic nucleophile. The tetrahedral intermediates arise through the addition of hydroxide to the peptide bond. In a theoretical study, a model tetrahedral adduct of hydroxide to *N*-methylacetamide was computationally compared with anionic *N*-methylmethanephosphonamidate and neutral *N*-methylmethanesulfonamide.[27] All were found to be tetrahedral. However, only anionic phosphonamidates were highly inhibitory and analogous to tetrahedral intermediates, highlighting the importance of electrostatic potential. The present results with a serine protease complement the earlier work on a nonserine protease.

Experimental work delineates other differences between boronate adducts to serine proteases and catalytic intermediates. Many X-ray crystallographic structures of protease−boronate complexes are available.[31−34] In several structures, the boron does not bind the active site serine-β-oxygen but instead is ligated by the $N^{\varepsilon 2}$ in the imidazole ring of histidine.[32,33,35] In the ultrahigh resolution structure of α-lytic protease inhibited with MeO-Succ-Ala-Ala-Pro-boroVal, in which hydrogen atoms are resolved, the boron-bound oxygen in the oxyanion binding

DOI: 10.1021/acs.jpcb.6b04089
J. Phys. Chem. B 2016, 120, 7353−7359

site is protonated, as B−OH, and not as B−O$^-$.[15] Therefore, the resultant hydroxyl group in B−OH can both accept and donate a hydrogen bond, unlike the oxyanionic centers in tetrahedral intermediates, which exclusively accept hydrogen bonds. Specifically, the B−OH group donates a hydrogen bond to another ligand to boron (a glycerol molecule), which cannot occur in a tetrahedral intermediate.

Boron ligation in proteases perturbs interactions between other active site-amino acids as well. The third amino acid in the catalytic triad of serine proteases, the β-carboxylate of aspartate (Asp102 in chymotrypsin) forms a hydrogen bond to N$^{\delta1}$ of histidine in the active site. Robillard and Shulman described a downfield ^{1}H NMR resonance at 18 ppm in chymotrypsin at low pH.[36−38] They assigned this signal to the hydrogen bond bridging His57 and Asp102. Hibbert and Emsley defined three classes of hydrogen bonds according to strength in small molecules: weak, 2−12 kcal mol^{-1}; intermediate strength, 12−24 kcal mol^{-1}; and very strong, >24 kcal mol^{-1}.[39] The intermediate and strong hydrogen bonds display downfield chemical shift values and curvilinear correlations with length and chemical shift.[40,41] The 18 ppm signal in chymotrypsin corresponds to an intermediate strength or low-barrier hydrogen bond (LBHB).[42] Downfield ^{1}H NMR signals can also arise from the effects of internal magnetic fields due to ring currents or nearby paramagnetic ions, which are absent in serine proteases. The 18 ppm signal would therefore have to be attributed to a property of the proton bridging His57 and Asp102, the most obvious being deshielding through elongation of the bond N$^{\delta1}$−H in His57.

Reactions of chymotrypsin with boric acid, phenylboronic acid, and 2-phenylethaneboronic acid gave complexes displaying ^{1}H NMR signals at 15.9, 16.3, and 17.2 ppm, well upfield from the signals in free chymotrypsin.[37,38] Deeply penetrating ^{11}B-, ^{15}N-, and ^{1}H NMR analysis of peptide boronate complexes of chymotrypsin and subtilisin as functions of pH and peptide structure revealed information about hydrogen bonding and ligation of boron.[43−46] In general, the complexes were found to be tetrahedral about boron, all complexes involved bonding of serine to boron, the most downfield histidinium proton was N$^{\delta1}$−H bonded to aspartate-β-COO$^-$, and the basicity of the active site histidine was elevated. Ligation of histidine to boron could be observed with a less specific boronic acid.[32,33,35,44] The most downfield proton appeared at 17.2−17.4 ppm, depending on the enzyme and inhibitor, upfield from the signal for chymotrypsin at low pH. The downfield proton displayed deuterium fractionation factors of 0.53 or 0.65 for complexes of MeO-Succ-Ala-Ala-Pro-boroPhe with subtilisin or chymotrypsin, respectively, indicative of enhanced hydrogen bonding.[46]

Unlike peptide boronic acids, peptidyl TFKs form tetrahedral complexes with serine proteases in which protons bridging histidine and aspartate in the active sites resonate between 18.6 and 18.9 ppm, well downfield from native chymotrypsin.[47−52] Deuterium fractionation factors for the downfield protons in these complexes are 0.3 to 0.4.[48] These properties indicate very strong, low-barrier His-Asp hydrogen bonding upon formation of the tetrahedral TFK complexes. An ultrahigh-resolution structure of the serine protease subtilisin revealed an elongated N$^{\delta1}$−H bond in the active site histidine, consistent with a LBHB.[53] A recent subangstrom-resolution structure of a β-lactamase revealed two authentic LBHBs, one in the active site serine and a structural LBHB outside the active site.[54]

Fuhrmann et al. assert that peptide boronate adducts in serine proteases are more closely related to tetrahedral peptidyl transition states than are peptidyl TFK adducts.[15] The present results do not support this hypothesis; rather, they suggest that the methoxide adduct of fAla TFK more closely resembles the model tetrahedral intermediates in terms of covalent and electrostatic structure. One could imagine that interactions of peptide boronate complexes with active sites might distort the complexes toward the transition state and away from the most stable structures in solution. However, the similarities between the computed covalent bond lengths of the model fAla boronates in the present work with those found experimentally in the complex of α-lytic protease with MeO-Succ-Ala-Ala-boroVal-OH (Table 2) do not support this interpretation. Given the differences in bond lengths and negative charge distribution in the tetrahedral centers, a tetrahedral intermediate and related boronate likely interact differently with an active site. In contrast, the structural and electrostatic similarities of tetrahedral intermediates to peptidyl TFK adducts suggests that the latter is a better mimic of the transition state. The results complement the spectroscopic and chemical case for the participation of LBHBs in catalysis by serine proteases.[47−52]

CONCLUSIONS

In the present results, computations show that the B−OCH$_3$ bonds of methoxide adducts to fAla boronic acid are much stronger than the C−OCH$_3$ bond in the methoxide adducts of fAla methyl ester and fAla methylamide, the model tetrahedral intermediates. Similarly, the C−OCH$_3$ bond in the methoxide adduct of fAla TFK is comparable in strength to the peptide boronates, that is, much stronger than in the model tetrahedral intermediates. The methoxyl in the model tetrahedral intermediates can leave spontaneously as methoxide, as expected for a transient intermediate, but not in the model inhibitory complexes, in which it is strongly bound. Similar relationships can be expected for the adducts of serine-β-oxyl in enzymatic tetrahedral intermediates and inhibited complexes. From a chemical standpoint, these relationships are rational and might have been predicted, at least in the cases of the peptide boronates. The B−OCH$_3$ bond can be expected to be stronger than the C−OCH$_3$ bond because of its greater polarity. In this respect, the peptide boronate adducts to serine proteases are unlike the catalytic tetrahedral intermediates, and the strength of this bond can contribute to the tight binding of peptide boronic acids to serine proteases. The peptidyl TFK adducts display nearly the same property, strong covalent bonding to the methoxide group. This is also at variance with catalytic tetrahedral intermediates. Evidently, electron withdrawal by the trifluoromethyl group in a peptidyl TFK enhances polarity in the C−O-alkyl bonds in tetrahedral adducts and likely contributes to the inhibitory potency of these compounds as well. In 1975, W. P. Jencks anticipated such behavior in certain classes of apparent transition state analogs binding tightly to enzymes through interactions not related to transition states.[55] Despite this anomaly, the alkoxide adducts of peptidyl TFKs appear structurally and electrostatically more similar to tetrahedral intermediates than do the same adducts to peptide boronic acids.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.6b04089.

DOI: 10.1021/acs.jpcb.6b04089
J. Phys. Chem. B 2016, 120, 7353−7359

Calculated bond lengths and partial charges (Tables S1−10) (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: som@auburn.edu. Tel. (334)844-6954.
*E-mail: frey@biochem.wisc.edu. Tel. (608)262-0055.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Hammond, G. S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77*, 334−338.
(2) Antonov, V. K.; Ivanina, T. V.; Berezin, I. V.; Martinek, K. *n*-Alkylboronic Acids as Bifunctional Reversible Inhibitors of Alpha-Chymotrypsin. *FEBS Lett.* **1970**, *7*, 23−25.
(3) Koehler, K. A.; Lienhard, G. E. Phenylethaneboronic acid, a Possible Transition State Analog of Chymotrypsin. *Biochemistry* **1971**, *10*, 2477−2483.
(4) Kettner, C.; Shenvi, A. Inhibition of the Serine Proteases Leukocyte Elastase, Pancreatic Elastase, Cathepsin G, and Chymotrypsin by Peptide Boronic Acids. *J. Biol. Chem.* **1984**, *259*, 15106−15114.
(5) Smoum, R.; Rubinstein, A.; Dembitsky, V. M.; Srebnik, M. Boron Containing Compounds as Protease Inhibitors. *Chem. Rev.* **2012**, *112*, 4156−4220.
(6) Imperiali, B.; Abeles, R. H. Inhibition of Serine Proteases by Peptidyl Fluoromethyl Ketones. *Biochemistry* **1986**, *25*, 3760−3767.
(7) Brady, K.; Abeles, R. H. Inhibition of Chymotrypsin by Peptidyl Trifluoromethyl Ketones: Determinants of Slow-Binding Kinetics. *Biochemistry* **1990**, *29*, 7608−7617.
(8) Liang, T. C.; Abeles, R. H. Complex of α-Chymotrypsin and *N*-Acetyl-L-Leucyl-L-Phenylalanyl Trifluoromethyl Ketone: Structural Studies with NMR Spectroscopy. *Biochemistry* **1987**, *26*, 7603−7608.
(9) Brady, K.; Liang, T. C.; Abeles, R. H. pH Dependence of the Inhibition of Chymotrypsin by a Peptidyl Trifluoromethyl Ketone. *Biochemistry* **1989**, *28*, 9066−9070.
(10) Allen, K. N.; Abeles, R. H. Inhibition Kinetics of Acetylcholinesterase with Fluoromethyl Ketones. *Biochemistry* **1989**, *28*, 8466−8473.
(11) Takahashi, L. H.; Radhakrishnan, R.; Rosenfield, R. E., Jr.; Meyer, E. F., Jr.; Trainor, D. A.; Stein, M. X-ray Diffraction Analysis of the Inhibition of Porcine Pancreatic Elastase by a Peptidyl Trifluoromethylketone. *J. Mol. Biol.* **1988**, *201*, 423−428.
(12) Brady, K.; Wei, A. Z.; Ringe, D.; Abeles, R. H. Structure of Chymotrypsin-Trifluoromethyl Ketone Inhibitor Complexes: Comparison of Slowly and Rapidly Equilibrating Inhibitors. *Biochemistry* **1990**, *29*, 7600−7607.
(13) Jackson, S. E.; Fersht, A. R. Contribution of Long-Range Electrostatic Interactions to the Stabilization of the Catalytic Transition State of the Serine Protease Subtilisin BPN′. *Biochemistry* **1993**, *32*, 13909−13916.
(14) Neidhart, D. I.; Wei, Y.; Cassidy, C.; Lin, J.; Cleland, W. W.; Frey, P. A. Correlation of Low-Barrier Hydrogen Bonding and Oxyanion Binding in Transition State Analogue Complexes of Chymotrypsin. *Biochemistry* **2001**, *40*, 2439−2447.
(15) Fuhrmann, C. N.; Daugherty, M. D.; Agard, D. A. Subangstrom Crystallography Reveals that Short Ionic Hydrogen Bonds, and not a His-Asp Low-Barrier Hydrogen Bond, Stabilize the Transition State in Serine Protease Catalysis. *J. Am. Chem. Soc.* **2006**, *128*, 9086−9102.
(16) Frisch, M. J. et al. *Gaussian 09*, Revision D.01; Gaussian, Inc., Wallingford, CT, 2013.
(17) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(18) Raghavachari, K.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. 20. Basis Set for Correlated Wave-Functions. *J. Chem. Phys.* **1980**, *72*, 650−654.
(19) Andersson, M. P.; Uvdal, P. New Scale Factors for Harmonic Vibrational Frequencies Using the B3LYP Density Functional Method with the Triple-Zeta Basis Set 6-311+G(d,p). *J. Phys. Chem. A* **2005**, *109*, 2937−2941.
(20) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999−3093.
(21) Aranda, J.; Cerqueira, N. M.; Fernandes, P. A.; Roca, M.; Tuñon, I.; Ramos, M. J. The Catalytic Mechanism of Carboxylesterases: a Computational Study. *Biochemistry* **2014**, *53*, 5820−5829.
(22) Takano, Y.; Houk, K. N. Conformational Preferences in the Transition States and Tetrahedral Intermediates of Transacylations. Relationships to Enzyme-Bound Conformations of Phosphonate Inhibitors of Lipases and Esterases. *J. Phys. Chem. A* **2004**, *108*, 11740−11751.
(23) Lee, I.; Kim, C. K.; Li, H. G.; Sohn, C. K.; Kim, C. K.; Lee, H. W.; Lee, B.-S. Acyl-Transfer Mechanisms Involving Various Acyl Functional Groups: > X=Y with X = C, S, P and Y = O, S. *J. Am. Chem. Soc.* **2000**, *122*, 11162−11172.
(24) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(25) Besler, B. H.; Merz, K. M., Jr; Kollman, P. A. Atomic Charges Derived from Semiempirical Methods. *J. Comput. Chem.* **1990**, *11*, 431−439.
(26) Foster, J. P.; Weinhold, F. Natural Hybrid Orbitals. *J. Am. Chem. Soc.* **1980**, *102*, 7211−7218.
(27) Radkiewicz, J. L.; McAllister, M. A.; Goldstein, E.; Houk, K. N. A Theoretical Investigation of Phosphonamidates and Sulfonamides as Protease Transition State Isosteres. *J. Org. Chem.* **1998**, *63*, 1419−1428.
(28) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388 CSD accession numbers JIQHUM, JIQJEY, VUSTUY, HEMBOQ..
(29) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. New Software for Searching the Cambridge Structural Database and Visualizing Crystal Structures. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 389−397.
(30) Varetto, U. *Molekel* v 5.4.0.8; Swiss National Supercomputing Centre, Manno, Switzerland, 2009.
(31) Matthews, D. A.; Alden, R. A.; Birktoft, J. J.; Freer, S. T.; Kraut, J. X-ray Crystallographic Study of Boronic Acid Adducts with Subtilisin BPN′ (Novo). A Model for the Catalytic Transition State. *J. Biol. Chem.* **1975**, *250*, 7120−7126.
(32) Stoll, V. S.; Eger, B. T.; Hynes, R. C.; Martichonok, V.; Jones, J. B.; Pai, E. F. Differences in Binding Modes of Enantiomers of 1-Acetamido Boronic Acid Based Protease Inhibitors: Crystal Structures of Gamma-Chymotrypsin and Subtilisin Carlsberg Complexes. *Biochemistry* **1998**, *37*, 451−462.
(33) Transue, T. R.; Gabel, S. A.; London, R. F. NMR and Crystallographic Characterization of Adventitious Borate Binding by Trypsin. *Bioconjugate Chem.* **2006**, *17*, 300−308.
(34) Tulinsky, A.; Blevins, R. A. Structure of a Tetrahedral Transition State Complex of Alpha-Chymotrypsin Dimer at 1.8 Å Resolution. *J. Biol. Chem.* **1987**, *262*, 7737−7743.
(35) Tsilikounas, E.; Kettner, C. A.; Bachovchin, W. W. ^{11}B NMR Spectroscopy of Peptide Boronic Acid Inhibitor Complexes of α-Lytic Protease. Direct Evidence for Tetrahedral Boron in both Boron-Histidine and Boron-Serine Adduct Complexes. *Biochemistry* **1993**, *32*, 12651−12655.
(36) Robillard, G.; Shulman, R. G. High Resolution Nuclear Magnetic Resonance Study of the Histidine-Aspartate Hydrogen Bond in Chymotrypsin and Chymotrypsinogen. *J. Mol. Biol.* **1972**, *71*, 507−511.
(37) Robillard, G.; Shulman, R. G. High Resolution Nuclear Magnetic Resonance Studies of the Active Site of Chymotrypsin: The Hydrogen Bonded Protons of the "Charge Relay" System. *J. Mol. Biol.* **1974**, *86*, 519−540.

DOI: 10.1021/acs.jpcb.6b04089
J. Phys. Chem. B 2016, 120, 7353−7359

(38) Robillard, G.; Shulman, R. G. High Resolution Nuclear Magnetic Resonance Studies of the Active Site of Chymotrypsin: II. *J. Mol. Biol.* **1974**, *86*, 541−558.
(39) Hibbert, F.; Emsley, J. Hydrogen bonding and chemical reactivity. *Adv. Phys. Org. Chem.* **1991**, *26*, 255−379.
(40) McDermott, A.; Rydenour, C. F. *Proton Chemical Shift Measurements in Biological Solids, in Encyclopaedia of Nuclear Magnetic Resonance*, Grant, D. M.; Harris, R. K., Eds.; John Wiley and Sons: New York, 1996; p 1320.
(41) Harris, T. K.; Mildvan, A. S. High Precision Measurement of Hydrogen Bond Lengths by Nuclear Magnetic Resonance measurements. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 275−282.
(42) Frey, P. A.; Whitt, S. A.; Tobin, J. B. A Low-Barrier Hydrogen Bond in the Catalytic Triad of Serine Proteases. *Science* **1994**, *264*, 1927−1930.
(43) Adebodun, F.; Jordan, F. ^{11}B-Nuclear Magnetic Resonance Studies of the structure of the Transition State Analogue Phenylboronic Acid Bound to Chymotrypsin. *J. Am. Chem. Soc.* **1988**, *110*, 309−310.
(44) Zhong, S.; Haghjoo, K.; Kettner, C.; Jordan, F. Proton Magnetic Resonance Studies of the Active Center Histidine of Chymotrypsin Complexed with Peptideboronic Acids. Solvent Accessibility to the $N^{\delta 1}$ and N^{ϵ} Sites Can Differentiate Slow-Binding and Rapidly Reversible Inhibitors. *J. Am. Chem. Soc.* **1995**, *117*, 7048−7055.
(45) Bao, D.; Cheng, J.-T.; Kettner, C.; Jordan, F. Assignment of the $N^{\epsilon 2}$ H and $N^{\delta 1}$ H Resonances at the Active Center Histidine in Chymotrypsin and Subtilisin Complexed to Peptideboronic Acids without Specific ^{15}N Labeling. *J. Am. Chem. Soc.* **1998**, *120*, 3485−3489.
(46) Bao, D.; Huskey, W. P.; Kettner, C. A.; Jordan, F. Hydrogen Bonding to Active Site Histidine in Peptidyl Boronic Acid Inhibitor Complexes of Chymotrypsin and Subtilisin. Proton Magnetic Resonance Assignments and H/D Fractionation. *J. Am. Chem. Soc.* **1999**, *121*, 4684−4689.
(47) Cassidy, C. S.; Lin, J.; Frey, P. A. A New Concept for the Mechanism of Action of Chymotrypsin: The Role of the Low-Barrier Hydrogen Bond. *Biochemistry* **1997**, *36*, 4576−4584.
(48) Lin, J.; Westler, W. M.; Cleland, W. W.; Markley, J. L.; Frey, P. A. Fractionation Factors and Activation Energies for Exchange of the Low Barrier Hydrogen Bonding Proton in Peptidyl Trifluoromethyl Ketone Complexes of Chymotrypsin. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 14664−14668.
(49) Cassidy, C. S.; Lin, J.; Frey, P. A. The Deuterium Isotope Effect on the NMR Signal of the Low-Barrier Hydrogen Bond in a Transition-State Analog Complex of Chymotrypsin. *Biochem. Biophys. Res. Commun.* **2000**, *273*, 789−792.
(50) Westler, W. M.; Frey, P. A.; Lin, J.; Wemmer, D. E.; Morimoto, H.; Williams, P. G.; Markley, J. L. Evidence for a Strong Hydrogen Bond in the Catalytic Dyad of Transition-State Analogue Inhibitor Complexes of Chymotrypsin from Proton-Triton NMR Isotope Shifts. *J. Am. Chem. Soc.* **2002**, *124*, 4196−4197.
(51) Lin, J.; Cassidy, C. S.; Frey, P. A. Correlations of the Basicity of His 57 with Transition State Analogue Binding, Substrate Reactivity, and the Strength of the Low-Barrier Hydrogen Bond in Chymotrypsin. *Biochemistry* **1998**, *37*, 11940−11948.
(52) Halkides, C. J.; Wu, Y. Q.; Murray, C. J. A Low-Barrier Hydrogen Bond in Subtilisin: ^{1}H and ^{15}NMR Studies with Peptidyl Tifluoromethyl Ketones. *Biochemistry* **1996**, *35*, 15941−15948.
(53) Kuhn, P.; Knapp, M.; Soltis, S. M.; Ganshaw, G.; Thoene, M.; Bott, R. The 0.78 Å Structure of a Serine Protease: *Bacillus lentus* Subtilisin. *Biochemistry* **1998**, *37*, 13446−13452.
(54) Nichols, D. A.; Hargis, J. C.; Sanishvili, R.; Jaishankar, P.; DeFrees, K.; Smith, E.; Wang, K.; Prati, F.; Renslo, A. R.; Woodcock, H. L.; Chen, Y. Ligand-Induced Proton Transfer and Low-Barrier Hydrogen Bond Revealed by X-ray Crystallography. *J. Am. Chem. Soc.* **2015**, *137*, 8086−8095.
(55) Jencks, W. P. Binding Energy, Specificity, and Enzyme Catalysis: The Circe Effect. *Adv. Enzymol. RAMB* **2006**, *43*, 219−410.

DOI: 10.1021/acs.jpcb.6b04089
J. Phys. Chem. B 2016, 120, 7353−7359

Supporting Information for

Serine protease catalysis: A computational study of tetrahedral intermediates and inhibitory adducts

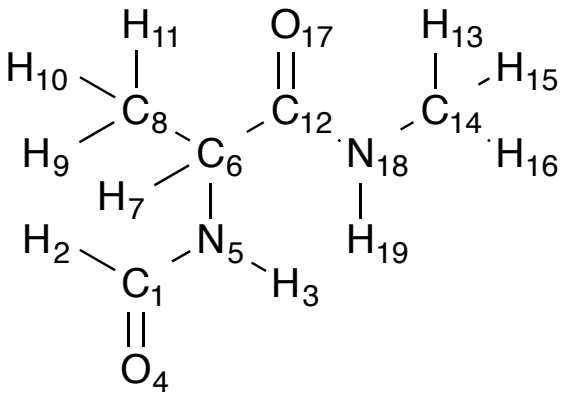
Phong D. Ngo,[†] Steven O. Mansoorabadi,[†*] and Perry A. Frey[‡*]

[†]*Department of Chemistry and Biochemistry, Auburn University, 179 Chemistry Building, Auburn, Alabama 36849, United States*

[‡]*Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706, United States*

Table S1. Calculated bond lengths and partial charges for *N*-formyl-L-alanine methylamide.

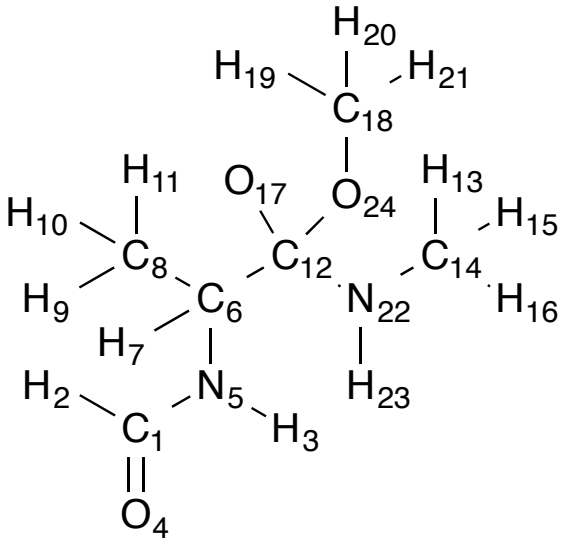
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_2-C_1	1.10610	1.10186
H_7-C_6	1.09510	1.09410
H_{13}-C_{14}	1.09454	1.09278
O_{17}-C_{12}	1.22165	1.23258
C_1-O_4	1.21261	1.22462
C_1-N_5	1.36734	1.35375
C_6-C_{12}	1.54761	1.54426
C_6-N_5	1.46066	1.46141
C_6-C_8	1.53261	1.53312
C_{12}-N_{18}	1.35635	1.34403
H_{10}-C_8	1.08932	1.08935
N_{18}-C_{14}	1.45527	1.45743
N_{18}-H_{19}	1.00832	1.00833
C_{14}-H_{16}	1.08890	1.09165
C_{14}-H_{15}	1.09145	1.08850
N_5-H_3	1.01352	1.01443
H_{11}-C_8	1.09176	1.09123
C_8-H_9	1.09445	1.09353



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.185673	0.542545	0.53857	0.208360	0.564355	0.54971
2	H	0.096989	0.003683	0.10846	0.127617	0.025438	0.12926
3	H	0.272161	0.311858	0.39508	0.298631	0.330930	0.40981
4	O	-0.358594	-0.553204	-0.59945	-0.465956	-0.656694	-0.67500
5	N	-0.228311	-0.498783	-0.66309	-0.203948	-0.450694	-0.64065
6	C	-0.057865	0.101803	-0.09819	-0.074422	0.043758	-0.09607
7	H	0.167496	0.100725	0.22172	0.187773	0.123902	0.23296
8	C	-0.482034	-0.301934	-0.57848	-0.485977	-0.283322	-0.57814
9	H	0.145790	0.106470	0.19924	0.162752	0.109663	0.20738
10	H	0.184973	0.103875	0.23615	0.169913	0.090663	0.22532
11	H	0.153109	0.094668	0.20676	0.163816	0.101719	0.21366
12	C	0.006734	0.430382	0.66737	0.060570	0.511643	0.67838
13	H	0.156402	0.074350	0.19104	0.167593	0.143746	0.19984
14	C	-0.216798	-0.095060	-0.37254	-0.224958	-0.359505	-0.37033
15	H	0.141786	0.074141	0.19441	0.160106	0.169965	0.21046
16	H	0.177458	0.106665	0.22543	0.165936	0.156255	0.20287
17	O	-0.364685	-0.513673	-0.64366	-0.467822	-0.624620	-0.70576
18	N	-0.272664	-0.351111	-0.62991	-0.278465	-0.276839	-0.60987
19	H	0.292379	0.262599	0.40109	0.328481	0.279636	0.41617

Table S2. Calculated bond lengths and partial charges for the methoxide adduct of *N*-formyl-L-alanine methylamide.

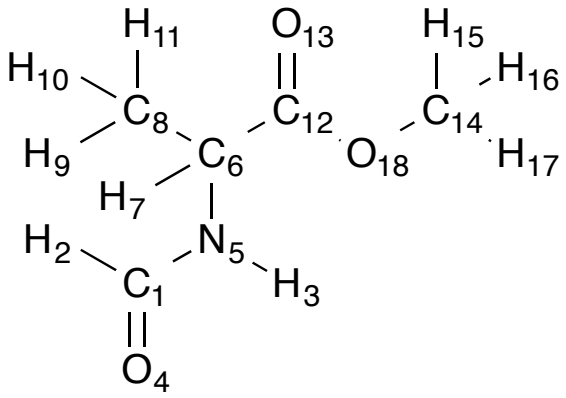
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_{21}-C_{18}	1.10080	1.09769
H_{20}-C_{18}	1.09719	1.09371
C_{18}-O_{24}	1.40500	1.41289
C_{18}-H_{19}	1.09725	1.09576
H_{23}-N_{22}	1.01499	1.01544
H_3-N_5	1.01486	1.01403
O_4-C_1	1.22993	1.23519
O_{24}-C_{12}	1.55662	1.52329
N_{22}-C_{12}	1.49985	1.49562
N_{22}-C_{14}	1.45389	1.46199
H_{15}-C_{14}	1.09699	1.09395
N_5-C_1	1.34031	1.33804
N_5-C_6	1.47041	1.46962
C_1-H_2	1.10858	1.10287
C_{12}-C_6	1.58409	1.58058
C_{12}-O_{17}	1.27668	1.29732
C_{14}-H_{16}	1.09091	1.09057
C_{14}-H_{13}	1.10858	1.10327
C_6-H_7	1.09560	1.09367
C_6-C_8	1.52898	1.52965
H_9-C_8	1.09287	1.09274
C_8-H_{11}	1.09681	1.09426
C_8-H_{10}	1.09087	1.09026



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.180972	0.474888	0.52924	0.219189	0.518180	0.53557
2	H	0.062483	-0.008470	0.08794	0.115751	0.026852	0.12031
3	H	0.288328	0.205575	0.41514	0.290205	0.232916	0.41091
4	O	-0.418159	-0.655481	-0.68052	-0.491873	-0.716999	-0.71893
5	N	-0.112204	-0.222665	-0.62803	-0.131096	-0.253803	-0.61910
6	C	0.037218	-0.119273	-0.07451	0.038860	-0.078946	-0.07387
7	H	0.126326	0.097163	0.18233	0.150169	0.132646	0.19698
8	C	-0.556080	-0.385495	-0.58396	-0.551230	-0.423562	-0.58153
9	H	0.148629	0.091690	0.20270	0.150885	0.107132	0.20078
10	H	0.167076	0.097795	0.22782	0.153381	0.109465	0.21549
11	H	0.101610	0.061820	0.17100	0.138723	0.101939	0.19382
12	C	-0.716473	1.127168	0.73578	-0.638084	1.020757	0.73924
13	H	0.071478	-0.060931	0.13120	0.105167	0.011607	0.15105
14	C	-0.240506	0.018724	-0.36742	-0.229684	-0.113418	-0.36907
15	H	0.096815	0.027967	0.16291	0.132571	0.095925	0.18487
16	H	0.163121	0.039483	0.21152	0.144042	0.072058	0.19171
17	O	-0.461011	-0.853727	-0.87721	-0.625941	-0.958811	-0.95328
18	C	-0.117654	-0.062934	-0.20323	-0.101797	-0.066299	-0.20770
19	H	0.138671	0.052753	0.17698	0.128891	0.054425	0.16504
20	H	0.086087	0.037052	0.14299	0.125880	0.075107	0.16637
21	H	0.108827	-0.007559	0.14582	0.121013	0.015822	0.15263
22	N	-0.070129	-0.695994	-0.73393	-0.165533	-0.752042	-0.74790
23	H	0.191608	0.230963	0.34108	0.220723	0.293285	0.35217
24	O	-0.277033	-0.490511	-0.71564	-0.300212	-0.504238	-0.70555

Table S3. Calculated bond lengths and partial charges for *N*-formyl-L-alanine methyl ester.

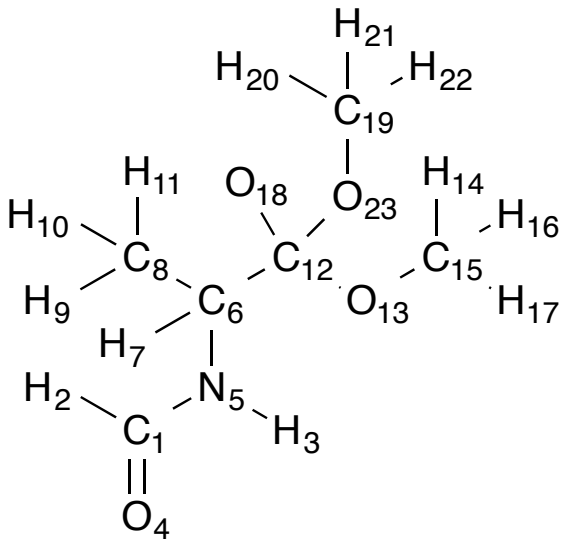
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_9-C_8	1.09280	1.09285
H_{16}-C_{14}	1.08725	1.08646
H_3-N_5	1.01214	1.01308
H_{11}-C_8	1.09153	1.09098
C_8-H_{10}	1.09232	1.09105
C_8-C_6	1.53797	1.53841
O_{18}-C_{14}	1.44345	1.44831
O_{18}-C_{12}	1.34495	1.33418
H_{15}-C_{14}	1.09044	1.08964
C_{14}-H_{17}	1.09021	1.08956
N_5-C_6	1.45300	1.45483
N_5-C_1	1.36309	1.34884
O_4-C_1	1.21321	1.22653
C_6-C_{12}	1.53172	1.53207
C_6-H_7	1.09158	1.09153
C_1-H_2	1.10651	1.10192
C_{12}-O_{13}	1.20522	1.21124



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.206706	0.457155	0.54009	0.210285	0.515808	0.54692
2	H	0.087407	0.040803	0.10521	0.124212	0.058494	0.12766
3	H	0.280261	0.293615	0.40295	0.301015	0.330357	0.41153
4	O	-0.368961	-0.536759	-0.60359	-0.478604	-0.651326	-0.68372
5	N	-0.114576	-0.421360	-0.63946	-0.134401	-0.474283	-0.61634
6	C	-0.241871	0.031926	-0.10600	-0.196597	0.098383	-0.10852
7	H	0.169370	0.079408	0.22316	0.187450	0.081266	0.23709
8	C	-0.392096	-0.263415	-0.57488	-0.409662	-0.245968	-0.57451
9	H	0.160042	0.076950	0.20684	0.169572	0.091984	0.21113
10	H	0.162627	0.093505	0.21577	0.169773	0.081612	0.22051
11	H	0.163922	0.102710	0.21430	0.171463	0.100928	0.21869
12	C	-0.036161	0.681247	0.81158	0.016116	0.706020	0.82686
13	O	-0.253319	-0.542784	-0.59009	-0.333932	-0.609808	-0.63759
14	C	-0.183434	-0.295731	-0.21593	-0.197489	-0.228257	-0.21823
15	H	0.161885	0.162024	0.18857	0.169763	0.155465	0.19263
16	H	0.156814	0.152961	0.19081	0.172007	0.152272	0.19947
17	H	0.165742	0.157710	0.19088	0.172920	0.146719	0.19302
18	O	-0.124359	-0.269964	-0.56017	-0.113890	-0.309667	-0.54661

Table S4. Calculated bond lengths and partial charges for the methoxide adduct of *N*-formyl-L-alanine methyl ester.

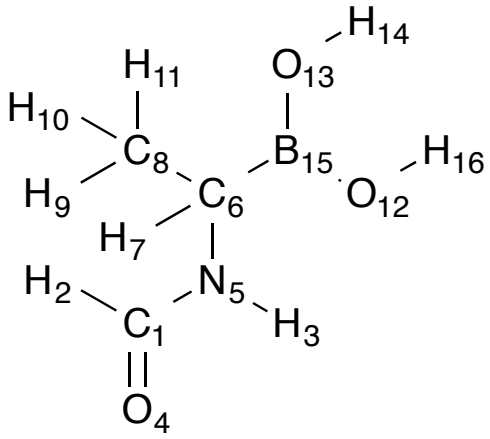
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_{10}-C_8	1.09128	1.09046
H_{11}-C_8	1.09623	1.09365
H_7-C_6	1.09461	1.09324
H_{14}-C_{15}	1.10598	1.10077
C_8-C_6	1.53123	1.53033
C_8-H_9	1.09298	1.09314
H_{17}-C_{15}	1.09565	1.09478
O_{18}-C_{12}	1.26618	1.28180
H_2-C_1	1.10905	1.10275
C_6-C_{12}	1.56884	1.56866
C_6-N_5	1.46643	1.46522
C_{15}-H_{16}	1.09665	1.09344
C_{15}-O_{13}	1.40419	1.41406
C_{12}-O_{13}	1.52047	1.50772
C_{12}-O_{23}	1.49587	1.47649
C_1-N_5	1.33894	1.33641
C_1-O_4	1.22968	1.23579
N_5-H_3	1.01385	1.01347
H_{20}-C_{19}	1.09633	1.09463
O_{23}-C_{19}	1.40812	1.41696
C_{19}-H_{22}	1.09926	1.09669
C_{19}-H_{21}	1.09545	1.09232



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.199444	0.487345	0.53637	0.228260	0.470098	0.53688
2	H	0.056234	0.000162	0.08389	0.115039	0.057086	0.11893
3	H	0.280202	0.232011	0.41478	0.281998	0.257339	0.40881
4	O	-0.424717	-0.653284	-0.68081	-0.500411	-0.705394	-0.72204
5	N	-0.051177	-0.257437	-0.61456	-0.065698	-0.252741	-0.60660
6	C	-0.111463	-0.277382	-0.07684	-0.135575	-0.299194	-0.07558
7	H	0.132387	0.168912	0.18695	0.156279	0.198458	0.20030
8	C	-0.487729	-0.418240	-0.58018	-0.505794	-0.396750	-0.58016
9	H	0.146044	0.106484	0.20141	0.150220	0.113321	0.20056
10	H	0.160513	0.109660	0.22115	0.152776	0.104498	0.21446
11	H	0.105507	0.095765	0.17390	0.142087	0.123541	0.19656
12	C	-0.537023	1.115507	0.84002	-0.431479	1.168253	0.84501
13	O	-0.178411	-0.565985	-0.67880	-0.244787	-0.655003	-0.69013
14	H	0.078921	-0.046121	0.12551	0.114926	-0.030662	0.14609
15	C	-0.182349	0.066778	-0.20570	-0.160807	0.104603	-0.21077
16	H	0.097619	0.028418	0.14933	0.134121	0.054652	0.17040
17	H	0.150950	0.017132	0.18133	0.138747	0.017538	0.16796
18	O	-0.448892	-0.828946	-0.86493	-0.592282	-0.950199	-0.92632
19	C	-0.116215	-0.179866	-0.20500	-0.108541	-0.199288	-0.20989
20	H	0.139317	0.076245	0.17564	0.133953	0.086197	0.16757
21	H	0.094678	0.059547	0.14850	0.132657	0.105740	0.17143
22	H	0.118079	0.062600	0.15183	0.128419	0.083889	0.15796
23	O	-0.221919	-0.399304	-0.68378	-0.264107	-0.455983	-0.68142

Table S5. Calculated bond lengths and partial charges for *N*-formyl-L-alanineboronic acid.

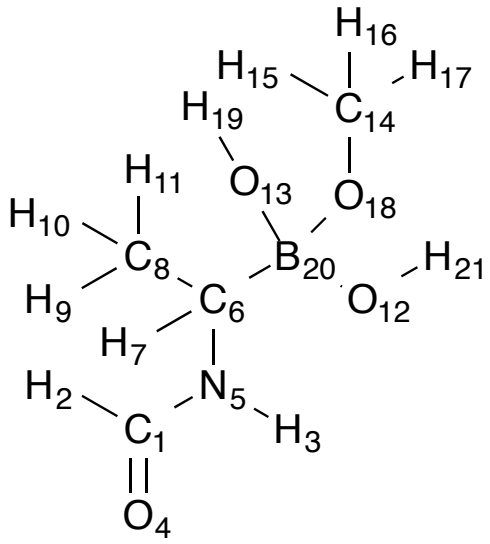
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_9-C_8	1.09432	1.09375
H_3-N_5	1.01408	1.01435
N_5-C_6	1.47389	1.47420
N_5-C_1	1.36632	1.34888
C_8-H_{11}	1.09311	1.09287
C_8-H_{10}	1.09165	1.09165
C_8-C_6	1.53766	1.53740
H_{16}-O_{12}	0.964578	0.964776
O_{12}-B_{15}	1.36540	1.36693
O_4-C_1	1.21351	1.22801
B_{15}-O_{13}	1.36280	1.36516
B_{15}-C_6	1.59484	1.59498
H_{14}-O_{13}	0.963892	0.965197
C_6-H_7	1.09929	1.09814
C_1-H_2	1.10684	1.10243



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.188462	0.543190	0.53763	0.208829	0.540783	0.54302
2	H	0.090764	0.019943	0.10486	0.121643	0.047398	0.12562
3	H	0.272832	0.360968	0.38973	0.300595	0.366728	0.40416
4	O	-0.365160	-0.552146	-0.60438	-0.482410	-0.664857	-0.69041
5	N	-0.230203	-0.672750	-0.65936	-0.212863	-0.579941	-0.62866
6	C	-0.473326	0.305026	-0.36415	-0.456168	0.247215	-0.36441
7	H	0.149596	0.005945	0.22108	0.171933	0.025882	0.23203
8	C	-0.444275	-0.316203	-0.57553	-0.468354	-0.289523	-0.57523
9	H	0.145288	0.083588	0.19749	0.159024	0.080745	0.20275
10	H	0.168337	0.089736	0.22070	0.161761	0.080100	0.21592
11	H	0.145674	0.086113	0.20031	0.150643	0.081522	0.20294
12	O	-0.386121	-0.675967	-0.87939	-0.441370	-0.720380	-0.89346
13	O	-0.411352	-0.727650	-0.87182	-0.457811	-0.777270	-0.88940
14	H	0.295533	0.476347	0.48763	0.321099	0.501223	0.49930
15	B	0.562179	0.577794	1.10127	0.609290	0.640117	1.11262
16	H	0.291773	0.396066	0.49391	0.314159	0.420260	0.50321

Table S6. Calculated bond lengths and partial charges for the methoxide adduct of *N*-formyl-L-alanineboronic acid.

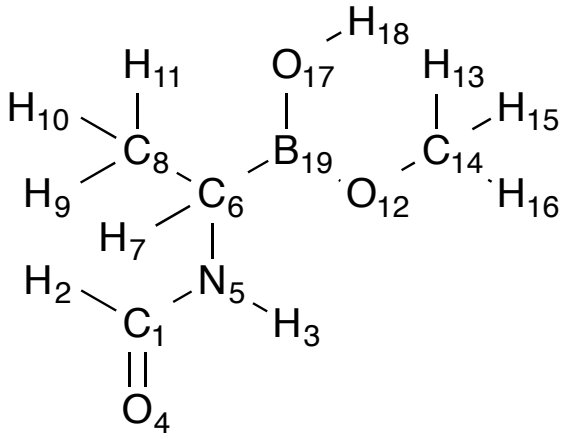
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_9-C_8	1.09426	1.09461
H_{16}-C_{14}	1.09850	1.09537
H_{11}-C_8	1.09919	1.09679
H_{15}-C_{14}	1.10519	1.10087
C_{14}-O_{18}	1.39713	1.40876
C_{14}-H_{17}	1.10021	1.09846
C_8-H_{10}	1.09301	1.09270
C_8-C_6	1.53639	1.53583
O_{18}-B_{20}	1.48580	1.48497
H_3-N_5	1.01453	1.01489
N_5-C_6	1.47483	1.47714
N_5-C_1	1.33602	1.33241
O_4-C_1	1.23300	1.23996
C_6-B_{20}	1.65275	1.65111
C_6-H_7	1.09918	1.09758
C_1-H_2	1.10967	1.10363
B_{20}-O_{13}	1.48779	1.48810
B_{20}-O_{12}	1.49117	1.49570
O_{13}-H_{19}	0.960743	0.961957
O_{12}-H_{21}	0.959822	0.961294



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.174450	0.504135	0.52835	0.214945	0.473194	0.52810
2	H	0.053694	-0.004129	0.08122	0.106055	0.049617	0.11460
3	H	0.285752	0.263047	0.40490	0.291240	0.261634	0.40159
4	O	-0.435780	-0.667721	-0.69460	-0.519030	-0.717979	-0.73891
5	N	-0.100595	-0.369381	-0.60559	-0.152572	-0.335726	-0.59994
6	C	-0.611030	-0.110220	-0.30915	-0.588797	-0.136480	-0.31577
7	H	0.087500	0.040484	0.16456	0.116539	0.060200	0.17921
8	C	-0.430637	-0.262029	-0.57637	-0.447258	-0.231480	-0.57664
9	H	0.142669	0.040198	0.19682	0.142744	0.032081	0.19153
10	H	0.144451	0.081003	0.20656	0.144575	0.074740	0.20346
11	H	0.095140	0.037736	0.16334	0.126511	0.062224	0.18225
12	O	-0.631082	-0.876447	-0.97647	-0.719479	-0.967915	-1.00274
13	O	-0.629838	-0.905673	-0.97824	-0.707802	-0.980496	-0.99870
14	C	-0.317589	-0.072077	-0.18809	-0.309121	-0.110473	-0.19831
15	H	0.093718	0.014346	0.12935	0.118658	0.050581	0.14413
16	H	0.096451	0.020161	0.14300	0.128163	0.062296	0.16008
17	H	0.136969	0.052776	0.15385	0.134220	0.076359	0.15214
18	O	-0.485021	-0.413635	-0.80345	-0.566180	-0.500081	-0.82178
19	H	0.237626	0.378303	0.44270	0.273819	0.438377	0.45895
20	B	0.877458	0.912576	1.07428	0.956107	0.940699	1.07649
21	H	0.215693	0.336546	0.44305	0.256664	0.398629	0.46027

Table S7. Calculated bond lengths and partial charges for *N*-formyl-L-alanineboronic acid monomethyl ester.

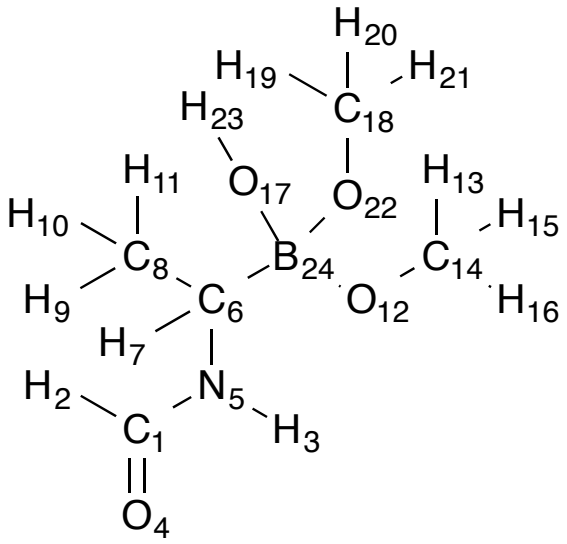
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_3-N_5	1.01283	1.01360
H_{18}-O_{17}	0.964406	0.965629
O_{17}-B_{19}	1.36632	1.36882
O_4-C_1	1.21521	1.22932
N_5-C_1	1.36066	1.34605
N_5-C_6	1.47471	1.47708
H_9-C_8	1.09230	1.09268
B_{19}-O_{12}	1.36353	1.36283
B_{19}-C_6	1.59909	1.60086
O_{12}-C_{14}	1.42740	1.43242
H_{13}-C_{14}	1.09214	1.09152
C_1-H_2	1.10745	1.10258
H_{15}- C_{14}	1.08969	1.08935
C_6-C_8	1.53544	1.53491
C_6-H_7	1.09740	1.09626
C_{14}-H_{16}	1.09384	1.09226
C_8-H_{11}	1.09348	1.09331
C_8-H_{10}	1.09346	1.09292



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.179118	0.489969	0.53424	0.197371	0.488995	0.53985
2	H	0.086414	0.009806	0.10168	0.122569	0.041638	0.12544
3	H	0.261365	0.303968	0.39318	0.287654	0.313395	0.40382
4	O	-0.359917	-0.557758	-0.61067	-0.477115	-0.666325	-0.69411
5	N	-0.122348	-0.473753	-0.64030	-0.126672	-0.416784	-0.61608
6	C	-0.437928	0.325949	-0.36137	-0.398132	0.288550	-0.36340
7	H	0.136418	0.005611	0.20761	0.167587	0.040397	0.22292
8	C	-0.554360	-0.551895	-0.57516	-0.600613	-0.574425	-0.57550
9	H	0.168097	0.151067	0.20865	0.165918	0.149189	0.20506
10	H	0.147212	0.135037	0.20715	0.154687	0.145277	0.21040
11	H	0.144667	0.129315	0.19864	0.150712	0.138222	0.20151
12	O	-0.380289	-0.466243	-0.73838	-0.408425	-0.505616	-0.74614
13	H	0.169677	0.052057	0.18013	0.169963	0.059998	0.17999
14	C	-0.304122	-0.002094	-0.19789	-0.313737	-0.000231	-0.20094
15	H	0.158988	0.096297	0.18466	0.165610	0.100088	0.18709
16	H	0.140319	0.043724	0.16522	0.160313	0.051341	0.17539
17	O	-0.407249	-0.696903	-0.87054	-0.463387	-0.749625	-0.88765
18	H	0.295548	0.426819	0.48664	0.319948	0.451810	0.49788
19	B	0.678389	0.579026	1.12650	0.725750	0.644107	1.13449

Table S8. Calculated bond lengths and partial charges for the methoxide adduct of *N*-formyl-L-alanineboronic acid monomethyl ester.

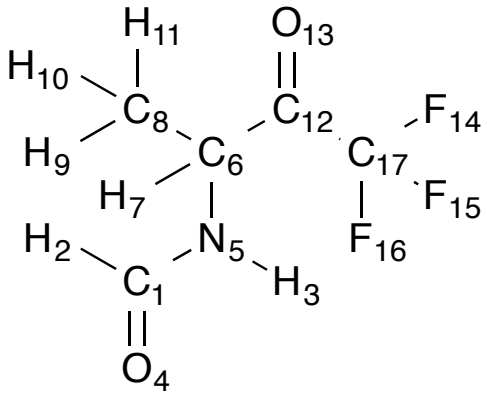
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_{20}-C_{18}	1.09843	1.09551
H_{21}-C_{18}	1.10026	1.09853
C_{18}-H_{19}	1.10545	1.10107
C_{18}-O_{22}	1.39681	1.40836
H_3-N_5	1.01446	1.01479
O_4-C_1	1.23273	1.24005
O_{22}-B_{24}	1.48571	1.48703
O_{12}-B_{24}	1.49402	1.49353
O_{12}-C_{14}	1.39507	1.40689
H_{15}-C_{14}	1.09921	1.09580
N_5-C_1	1.33650	1.33236
N_5-C_6	1.47517	1.47734
C_1-H_2	1.10953	1.10362
B_{24}-C_6	1.65052	1.65112
B_{24}-O_{17}	1.48203	1.48530
C_{14}-H_{13}	1.10444	1.10127
C_{14}-H_{16}	1.10254	1.09964
C_6-H_7	1.09758	1.09630
C_6-C_8	1.53650	1.53589
H_{23}-O_{17}	0.960789	0.961954
H_9-C_8	1.09408	1.09452
C_8-H_{11}	1.09908	1.09685
C_8-H_{10}	1.09307	1.09273



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.179485	0.617249	0.52831	0.220628	0.605498	0.52833
2	H	0.055573	-0.046477	0.08245	0.106093	0.006606	0.11501
3	H	0.283097	0.277540	0.40424	0.288554	0.281063	0.40175
4	O	-0.426786	-0.692784	-0.69326	-0.510552	-0.749002	-0.73886
5	N	-0.105732	-0.510873	-0.60618	-0.146016	-0.481606	-0.60075
6	C	-0.581271	0.185491	-0.32190	-0.574871	0.060183	-0.32693
7	H	0.099977	0.019666	-0.32190	0.120936	0.075172	0.18265
8	C	-0.494452	-0.411141	-0.57393	-0.502017	-0.357928	-0.57459
9	H	0.144454	0.080965	0.19728	0.144136	0.076048	0.19168
10	H	0.145865	0.099565	0.20654	0.145794	0.093400	0.20356
11	H	0.096683	0.046091	0.16430	0.126289	0.068444	0.18226
12	O	-0.502495	-0.558703	-0.80251	-0.585346	-0.658037	-0.82378
13	H	0.096700	-0.011350	0.13328	0.115209	-0.010362	0.14287
14	C	-0.276633	-0.011637	-0.18825	-0.268952	0.037054	-0.19823
15	H	0.097623	0.042764	0.14261	0.126240	0.059985	0.15959
16	H	0.111739	-0.001768	0.14256	0.121057	0.017982	0.14787
17	O	-0.620285	-0.926127	-0.97462	-0.718284	-0.992556	-0.99767
18	C	-0.337242	-0.037032	-0.18739	-0.328118	-0.056407	-0.19805
19	H	0.092719	0.002450	0.12803	0.119276	0.035099	0.14340
20	H	0.099372	0.037938	0.14406	0.129146	0.063960	0.15990
21	H	0.135444	0.014395	0.15296	0.134766	0.039423	0.15168
22	O	-0.465699	-0.504458	-0.80268	-0.558752	-0.556158	-0.82207
23	H	0.225633	0.419605	0.44347	0.265636	0.475527	0.45944
24	B	0.946231	0.868630	1.10892	1.029150	0.866611	1.11093

Table S9. Calculated bond lengths and partial charges for *N*-formyl-L-alanine trifluoromethyl ketone.

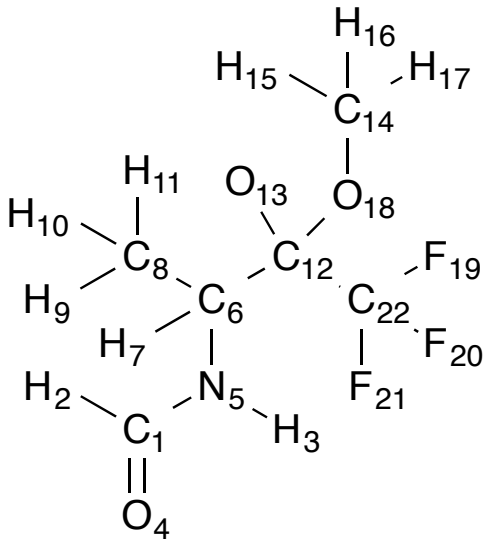
Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_3-N_5	1.01314	1.01408
F_{16}-C_{17}	1.34599	1.34889
H_9-C_8	1.09211	1.09210
O_{13}-C_{12}	1.19868	1.19985
N_5-C_1	1.37042	1.35644
N_5-C_6	1.45578	1.45821
O_4-C_1	1.20913	1.22176
C_{12}-C_{17}	1.56501	1.56704
C_{12}-C_6	1.53607	1.53295
C_{17}-F_{15}	1.33457	1.33880
C_{17}-F_{14}	1.35129	1.34739
C_8-C_6	1.52810	1.52766
C_8-H_{11}	1.09116	1.09054
C_8-H_{10}	1.09114	1.09043
C_1-H_2	1.10584	1.10166
C_6-H_7	1.09654	1.09656



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.201909	0.462208	0.54598	0.218813	0.502660	0.55273
2	H	0.099990	0.044241	0.10994	0.135243	0.056997	0.13297
3	H	0.278439	0.336409	0.40044	0.307758	0.345459	0.41425
4	O	-0.341036	-0.508083	-0.58151	-0.446103	-0.619996	-0.66029
5	N	-0.156328	-0.561174	-0.64990	-0.157921	-0.539982	-0.63201
6	C	-0.168203	0.492040	-0.11623	-0.180726	0.519152	-0.11884
7	H	0.188078	-0.007745	0.23357	0.226186	0.010099	0.25324
8	C	-0.498194	-0.555519	-0.58480	-0.534651	-0.547597	-0.58472
9	H	0.164622	0.157413	0.21400	0.173980	0.155603	0.21605
10	H	0.161051	0.154444	0.22181	0.167303	0.150874	0.22473
11	H	0.161906	0.143494	0.21544	0.172326	0.147789	0.22219
12	C	0.119108	0.223623	0.50196	0.221186	0.290945	0.52520
13	O	-0.191623	-0.374779	-0.47703	-0.238166	-0.434553	-0.50244
14	F	-0.095729	-0.253529	-0.35479	-0.087773	-0.263259	-0.35048
15	F	-0.099708	-0.221522	-0.33475	-0.122277	-0.239519	-0.34432
16	F	-0.073405	-0.224528	-0.34015	-0.084275	-0.244940	-0.34620
17	C	0.249121	0.693007	0.99602	0.229097	0.710265	0.99793

Table S10. Calculated bond lengths and partial charges for the methoxide adduct of *N*-formyl-L-alanine trifluoromethyl ketone.

Calculated Bond Lengths (Å)		
Bond	Gas Phase	Aqueous Phase
H_{10}-C_8	1.09027	1.08953
H_7-C_6	1.09463	1.09304
H_{11}-C_8	1.09522	1.09302
F_{19}-C_{22}	1.36810	1.36575
C_8-C_6	1.53053	1.52999
C_8-H_9	1.09318	1.09277
O_{13}-C_{12}	1.27609	1.29554
H_2-C_1	1.10859	1.10272
C_6-C_{12}	1.57877	1.57679
C_6-N_5	1.46561	1.46536
C_{12}-C_{22}	1.57921	1.58194
C_{12}-O_{18}	1.55415	1.50584
C_{22}-F_{20}	1.35440	1.35806
C_{22}-F_{21}	1.36390	1.36193
C_1-N_5	1.34150	1.33853
C_1-O_4	1.22782	1.23400
N_5-H_3	1.01388	1.01350
H_{15}-C_{14}	1.09579	1.09284
O_{18}-C_{14}	1.40931	1.42030
C_{14}-H_{17}	1.10086	1.09752
C_{14}-H_{16}	1.09744	1.09262



Calculated Partial Charges							
Atom		Gas Phase			Aqueous Phase		
No.	Element	MPA	ESP	NBO	MPA	ESP	NBO
1	C	0.189974	0.504087	0.53618	0.218148	0.538921	0.53903
2	H	0.062459	-0.014339	0.08735	0.115999	0.024491	0.12074
3	H	0.289370	0.235643	0.41697	0.289955	0.256398	0.41190
4	O	-0.412474	-0.636249	-0.67209	-0.490206	-0.699076	-0.71418
5	N	-0.074570	-0.385680	-0.62310	-0.103717	-0.413513	-0.61339
6	C	-0.038053	0.292605	-0.06802	-0.032837	0.278604	-0.06501
7	H	0.148148	0.016628	0.19357	0.171374	0.038398	0.20654
8	C	-0.513479	-0.491575	-0.58165	-0.529116	-0.370574	-0.58103
9	H	0.149177	0.129325	0.20005	0.155922	0.107816	0.20227
10	H	0.168202	0.141807	0.22739	0.157546	0.109920	0.21851
11	H	0.112328	0.069333	0.17841	0.147995	0.071043	0.20050
12	C	-0.840618	0.419619	0.51182	-0.743961	0.380475	0.51517
13	O	-0.397967	-0.749034	-0.84655	-0.555643	-0.855429	-0.91901
14	C	-0.137923	-0.176573	-0.20994	-0.116870	-0.213254	-0.21175
15	H	0.160828	0.123626	0.19569	0.149323	0.131480	0.18045
16	H	0.094478	0.052158	0.14383	0.135840	0.098603	0.17140
17	H	0.102873	0.045619	0.13938	0.127987	0.081582	0.15375
18	O	-0.211801	-0.347846	-0.70781	-0.216844	-0.340013	-0.69222
19	F	-0.149871	-0.264276	-0.38899	-0.145094	-0.236124	-0.38717
20	F	-0.144795	-0.195220	-0.37411	-0.153125	-0.182467	-0.37902
21	F	-0.119771	-0.251731	-0.38274	-0.117390	-0.226467	-0.38181
22	C	0.563486	0.482074	1.02435	0.534714	0.419184	1.02432