DIRECT CORRELATION OF PROTEIN STRUCTURE AND FUNCTION USING HIGH-PRESSURE X-RAY CRYSTALLOGRAPHY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Buz Michael Barstow

May 2009

DIRECT CORRELATION OF PROTEIN STRUCTURE AND FUNCTION USING HIGH-PRESSURE X-RAY CRYSTALLOGRAPHY

Buz Michael Barstow, Ph. D.

Cornell University 2009

A protein molecule is an intricate system whose function is highly sensitive to small external perturbations. However, few examples that directly correlate protein function with continuous, progressive sub-angstrom structural perturbations have thus far been presented. In order to elucidate this relationship, we have investigated the *Aequorea* yellow fluorescent protein Citrine as a model system under high-pressure perturbation. Citrine has been compressed by high pressure to produce deformations of its β-barrel scaffold and light absorbing and emitting center, the *chromophore*, by applying a novel high pressure cryo-cooling technique.

A closely spaced series of high-pressure X-ray crystallographic structures of Citrine from 0.1 to 500 MPa reveal that the chromophore undergoes a progressive deformation of up to 0.8 Å at an applied pressure of 500 MPa. It is experimentally demonstrated that deformation of the chromophore is directly correlated with a progressive shift of the fluorescence peak of Citrine from yellow to green under these conditions.

The re-orientation of the Citrine chromophore is actuated by the differential motion of two clusters of atoms that compose the β-barrel scaffold of the molecule, resulting in a bending or buckling of the β-barrel. The high-pressure structures also reveal a perturbation of the hydrogen bonding network stabilizing the excited state of the

Citrine chromophore that is implicated in the reduction of fluorescence intensity of the molecule under high pressure.

The blue-shift of the Citrine fluorescence spectrum resulting from the bending of the β-barrel provides structural insight into the transient blue-shifting of isolated yellow fluorescent protein molecules under ambient conditions and suggests mechanisms to alter the time-dependent behavior of Citrine under ambient conditions.

The results presented in this thesis demonstrate that the fluorescence spectrum of Citrine is highly sensitive to sub-angstrom deformations and its fluorescent function must be understood at the sub-angstrom level. These results provide important general lessons for the structure-function relationship of enzymes, and may have significant implications for protein function prediction and biomolecule design and engineering as they suggest methods to tune protein function by modification of the protein scaffold.

BIOGRAPHICAL SKETCH

Buz Barstow arrived at Cornell University in 2001. His first graduate project was to assist in the development of the superconducting radio-frequency cavities for the energy recovery linac (ERL) X-ray light source. During this time he developed his interests in biological experiments that could be performed with the ERL, and decided to switch the focus of his research from accelerator physics to biological physics: to develop a greater understanding of the function of biological macromolecules, and to gain insights that would assist in the engineering of macromolecules. The goal of his career is to develop synthetic proteins for the environmentally sustainable production of materials, fuels, chemicals and medicines and to help remediate environmental degradation.

ACKNOWLEDGEMENTS

At many points in this thesis, I will use the pronoun “we”. This pronoun is used to refer to my co-authors: Nozomi Ando, Chae Un Kim, and my advisor, Sol Gruner, to whom I owe an enormous debt of gratitude. Nozomi and Chae Un were both enormous contributors to my thinking on Citrine, to the design and conduct of experiments, and a great source of encouragement. I can’t thank Sol enough for his encouragement and support, both financial and intellectual. Sol provided me with the freedom that I needed to make this work happen, and allowed me to develop at my own pace. I wrote this thesis with one eye on an earlier version of myself, and have included information and advice that I would have liked to have known 7 years ago. The first piece of advice that I would give to a younger me is to pick Sol as an advisor. Sol might not be everyone, but I’m convinced that he was the right choice for me.

In addition to my co-authors, there is a long list of people whose assistance I relied upon heavily. During the course of this thesis project, I was given the freedom to perform almost each step at my own speed, from molecular biology, to protein purification and crystallization, design and construction of apparatus, machining to numerical analysis and simulation. Learning so much required help from many experts. I would like to thank Dr Cynthia Kinsland and the staff of the Cornell Life Sciences Core Lab Centers Protein Facility for assistance with the modification and expression of Citrine vectors and the purification of Citrine. I enormously enjoyed working with Cynthia. The MacCHESS staff at the Cornell High Energy Synchrotron Source for assistance in data collection and reduction. Martin Novak in the Gruner group provided invaluable assistance with the construction of experiments.

Yi-Fan Chen and Dr Mark Tate, also in the Gruner group constructed of the internals of the high pressure cryo-cooling apparatus. Odin Wojcik and AccuFab Inc. (Ithaca, NY, USA) constructed the high pressure cryo-cooling safety enclosure. Professor Warren Zipfel (Cornell University) provided a sample of monomeric EGFP. Professor Roger Tsien (University of California San Diego) provided the Citrine plasmid. Dr Gerhard Hummer (National Institute of Diabetes and Digestive and Kidney Diseases) helped me with molecular dynamics simulations of Citrine. Dr Ismail Hafez (University of British Columbia, Vancouver) was the initial source of inspiration for this thesis. I have a feeling that his suggestion to study Citrine was one of the most fortuitous influences on my life. Professors Lois Pollack and Brian Crane (Cornell University) both served on my thesis committee, and went out of their way to help me out. Professor Robert Campbell (University of Alberta) provided invaluable assistance with the crystallization of Citrine. Elizabeth Landrum and Darren Southworth assisted in data collection. Joan Lenz and Dr Raymond Molloy (Cornell University) both assisted with molecular biology.

Finally, thanks must go to the sponsors of this work: the United States National Institutes of Health, US National Science Foundation and US Department of Energy. This work was supported under the Department of Energy Office of Biological Environmental Research (DOE-BER) grant number FG02-97ER62443, and the National Institutes of Health Protein Structure Initiative (NIH-PSI) grant number GM074899. CHESS is supported by the NSF and NIH via NSF award DMR-0225180. MacCHESS is supported via NIH grant RR001646.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF EQUATIONS

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF TABLES

CHAPTER 3

CHAPTER 4

CHAPTER 6

LIST OF ABBREVIATIONS

A. victoria	*Aequorea victoria*
BFP	Blue Fluorescent Protein
CCP4	Collaborative Computational Project Number 4. A suite of programs for manipulating X-ray diffraction data and macromolecular atomic structures.
CHESS	Cornell High Energy Synchrotron Source
Da	1 Dalton, a unit of mass commonly used in biochemistry. 1 Da = 1.67×10^{-27} Kg = $1/12^{th}$ of the mass of 1 ^{12}C atom ≈ 1 ^{1}H atom.
DPS	Data Processing Suite. A software package used to index, integrate, scale and truncate X-ray diffraction images.
ECFP	Enhanced Cyan Fluorescent Protein
EGFP	Enhanced Green Fluorescent Protein
EHT	Extended Hückel Theory
EYFP	Enhanced Yellow Fluorescent Protein
FMN	Flavin mononucleotide
GFP	Green Fluorescent Protein
Hb	Hemoglobin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hi-P	High-pressure
HOMO	Highest Occupied Molecular Orbital
IPTG	Isopropyl β-D-1-thiogalactopyranoside
NADH	Nicotinamide Adenine Dinucleotide Hydride
$NADH_4$	1,4,5,6-tetrahydro-NADH. A non-oxidizable analog of NADH.
KIE	Kinetic Isotope Effect = k_H/k_D (Protium transfer rate / Deuterium transfer rate)
LB	Luria-Bertani broth
LUMO	Lowest Unoccupied Molecular Orbital

Mb	Myoglobin
MR	Morphinone Reductase
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank. An online repository of protein atomic structures. http://www.pdb.org
P. putida	*Pseudomonas putida*
TEV	Tobacco Etch Virus. Typically used to refer to the Tobacco Etch Virus protease enzyme.
TGF-β2	Transforming Growth Factor β2
T. fusca	*Thermobifida fusca*
Tris	Tris(hydroxymethyl)aminomethane
vdW	van der Waals. Refers to a gas displaying simple, but non-ideal behavior, including long-range attraction of particles, and short-range repulsion.
v/v	Volume/volume. A measure of concentration. For example 1% v/v PEG 400 means that 1% of the total volume of a solution should be PEG 400.
w/v	Weight/volume. A measure of concentration. For example 5% w/v PEG 3350 means that 5 grams of PEG 3350 powder should be dissolved in a solution, and the final volume brought up to 100 mL.
wt	Wild-type
wtGFP	Wild-type Green Fluorescent Protein

LIST OF SYMBOLS

F	Structure factor $F = A\exp(i\phi)$
A	Amplitude of structure factor
ϕ	Phase factor of structure factor
I	Intensity, either of Bragg spot; $I(h,k,l)$, or intensity in general; $I(x)$. $I = FF^* = A^2$.
h, k, l	Reciprocal lattice and diffraction spot indices.
d	Separation of the NADH (or $NADH_4$) C4 and FMN N5 atoms Morphinone Reductase
r	Separation of proton donor (attached to NADH C4 atom) and acceptor (attached to FMN N5 atom) sites in Morphinone Reductase.
R	Crystallographic *R*-factor. An indicator of the quality of a model of a macromolecule calculated during refinement procedure. Root mean square difference between observed diffraction amplitudes and those calculated from model. Usually calculated through L_2 norm.
R_{Free}	Crystallographic R factor based upon the match of the model to a small subset of an X-ray dataset (usually $\approx$ 5% of reflection) that is sequestered from refinement. Usually calculated through L_2 norm using sequestered set of reflections.
k_{H}	Rate constant for protium (^{1}H) transfer.
k_{D}	Rate constant for deuterium (^{2}H) transfer.
$\nu_{\text{Fe-His}}$	Heme iron-proximal histidine stretching mode Raman peak
λ	Wavelength
k_{i}	Incident wave vector
k_{s}	Scattered wave vector
Q	Scattering vector. $\vec{Q} = \vec{k}_{\text{i}} - \vec{k}_{\text{s}}$.
G	Inverse lattice vector
Δk	Spread in wave number of incident X-ray beam
n	Number of observations used in the refinement of a macromolecular structure.

p	Number of parameters of a macromolecular structure. Typically 4-13 × the number of atoms in the structure.
$\mathbf{x}$	Set of model parameters for a macromolecular structure.
$\mathbf{M}(\mathbf{x})$	Diffraction amplitudes and phases calculated from model parameters, $\mathbf{x}$.
$\mathbf{y}$	Set of diffraction data
L_2	Residual between calculated, $\mathbf{M}(\mathbf{x})$ and observed, $\mathbf{y}$, diffraction data. Defined in equation 2.4.
$\mathbf{W}$	Weight matrix in macromolecular refinement. Defined in equation 2.9.
$\mathbf{e}$	Set of errors on model parameters $\mathbf{x}$.
$\sigma(x_i)$	The estimated standard uncertainty on the model coordinate x_i. See equations 2.27 to 2.32.
$d_{\min}$	The maximum resolution of an X-ray diffraction dataset. Expressed in Å.
Ψ	Complete wave function of a molecule.
ϕ_i	Slater type atomic orbital. See equation 3.2.
l	Angular momentum quantum number.
n	Principal quantum number.
r, θ, φ	Spherical polar coordinates.
ζ	Slater coefficient. An empirical radial scaling coefficient Slater type atomic orbitals.
Y_l^m	Spherical harmonic function.
$\mathbf{H}$	Hamiltonian operator
E	Total energy of a wavefunction.
H_{ij}	Resonance integral matrix element. See equation 3.6.
S_{ij}	Overlap matrix element. See equation 3.7.
a_i	Probability amplitude
ψ_i	Molecular orbital i

E_i	Energy of molecular orbital i
$C_{\mu\nu}$	Hückel coefficient. Usually set to 1.75 for all μ and ν.
$H_{\mu\nu}$	Ionization potential for atomic orbital μ
W	Work. Energy liberated upon the expansion of a gas.
P	Pressure
V	Volume
γ	Adiabatic expansion coefficient. For an ideal gas = 5/3.
K	Petry coefficient for penetration of projectiles into a barrier. See equation 4.8.
d	Penetration depth of projectile. See equation 4.8.
A	Area of projectile. See equation 4.8.
m	Mass of projectile. See equation 4.8.
v	Velocity of projectile. See equation 4.8.
U	Strain energy deposited into plate. See equation 4.9.
F	Force
δ	Central deflection of a plate. See equation 4.10.
μ	Poisson ratio of material
E	Young's modulus of material
h	Thickness of plate
D	Flexural rigidity of plate. See equation 4.11.
s	Stress
ε	Strain
E_{Elastic}	Energy absorbed by elastic deformation
E_{Plastic}	Energy absorbed by plastic deformation
Y	Maximum stress. Stress above which deformations are no longer elastic.
N	Number of bolts

ξ	Extinction coefficient
C	Concentration
l	Length of sample
Σ	Estimated uncertainty on a positional deviation between two structures. See equation 5.1.
β	Isothermal compressibility. See equation 5.3.
κ	Isotropic bulk modulus.
I_z	Moment of inertia along *z*-axis. See equation 5.4.
I_x	Moment of inertia along *x*-axis. See equation 5.5.
I_y	Moment of inertia along *y*-axis. See equation 5.5.
ΔE_c	Calculated energy shift of fluorescence peak
ΔE_m	Measured energy shift of fluorescence peak

CHAPTER 1

INTRODUCTION

1.1 Overview of this Thesis

A protein molecule is an intricate system whose function is highly sensitive to small external perturbations. However, few examples that directly correlate protein function with continuous, progressive sub-angstrom structural perturbations have thus far been presented. In order to elucidate this relationship, we have investigated the *Aequorea* yellow fluorescent protein Citrine [1, 2] as a model system under high-pressure perturbation. Citrine has been compressed by high pressure to produce deformations of its β-barrel scaffold and light absorbing and emitting center, the *chromophore*, by applying a novel high pressure cryo-cooling technique [3, 4].

A closely spaced series of high-pressure X-ray crystallographic structures of Citrine from 0.1 to 500 MPa reveal that the chromophore undergoes a progressive deformation of up to 0.8 Å at an applied pressure of 500 MPa. It is experimentally demonstrated that deformation of the chromophore is directly correlated with a progressive shift of the fluorescence peak of Citrine from yellow to green under these conditions [5].

The re-orientation of the Citrine chromophore is actuated by the differential motion of two clusters of atoms that compose the β-barrel scaffold of the molecule, resulting in a bending or buckling of the β-barrel [6]. The high-pressure structures also reveal a perturbation of the hydrogen bonding network stabilizing the excited state of the Citrine chromophore that is implicated in the reduction of fluorescence intensity of the molecule under high pressure.

The blue-shift of the Citrine fluorescence spectrum resulting from the bending of the β-barrel provides structural insight into the transient blue-shifting of isolated yellow fluorescent protein molecules under ambient conditions [7] and suggests mechanisms to alter the time-dependent behavior of Citrine under ambient conditions.

The results presented in this thesis demonstrate that the fluorescence spectrum of Citrine is highly sensitive to sub-angstrom deformations and its fluorescent function must be understood at the sub-angstrom level. These results provide important general lessons for the structure-function relationship of enzymes, and may have significant implications for protein function prediction and biomolecule design and engineering as they suggest methods to tune protein function by modification of the protein scaffold.

1.2 Introduction

It is well known that the three-dimensional structure of a protein molecule is crucial to understanding its function [8]. As protein molecules have dimensions of tens to hundreds of angstroms, sub-angstrom perturbations to the positions of atoms and functional groups in the molecule may appear insignificant [9]. However, if one considers two objects interacting through a Lennard-Jones potential that are separated by 3.5 Å and moves them together by just 0.1 Å, the potential energy between these objects will change by $(1/3.4^6) / (1/3.5^6) = 1.19$, or 19%. This result suggests that changes as small as 0.1 Å in the relative positions of critical functional groups could make a large difference in the energy of an electronic transition, and thus a notable change in protein function [10]. Fersht and Winter [11] noted that the activity of triose phosphate isomerase is reduced by approximately three to four orders of magnitude by a change in the position of the active site base by approximately 1 Å.

The sensitivity of protein function to sub-angstrom positioning of critical functional groups is further implied by the observation that modest pressures (less than a few hundred MPa) significantly modify protein function [12]. Using high-pressure stopped flow techniques Ueda *et al*. [13] observed that pressures in the range of tens of MPa reduce the flash decay rate of firefly luciferase. Bruner and Hall [14] used a high-pressure patch clamp apparatus to measure the conductance properties of the pore forming membrane protein alamethicin. Bruner and Hall found that pressures up to 100 MPa exponentially increase the average lifetime of the conductance states of alamethicin. Unno *et al*. [15] found that the R (relaxed or oxygenated) to T (tense or de-oxygenated) transition of hemoglobin is biased towards the R-state by the application of pressure. Adachi and Morishima [16] established that the rate of carbon monoxide and oxygen association of horse and sperm whale myoglobin are exponentially modified by pressures up to 200 MPa using time resolved spectroscopy and flash photolysis. Miller *et al.* [17] observed that the methyl-viologen reducing activity (a proxy for H_2 consumption) of a hydrogenase from the thermophilic bacterium *Methanococcus jannaschii* is increased by the application of pressures up to 75 MPa. Hay *et al*. [18] observed that pressures up to 200 MPa exponentially increase the rate of NADH oxidation by the flavoprotein morphinone reductase.

Several studies have highlighted the sensitivity of inter- and intra-molecular biological electron transfer to the application of high pressure. Using high-pressure radiolysis and optical absorption Wishart *et al*. [19] demonstrated that the application of high pressure causes an exponential increase in the rate of intra-molecular electron transfer in metal-labeled cytochrome c, and electron transfer from metal in solution to cytochrome c. Meier *et al*. [20] measured the effects of high pressure on the rates of intra-molecular electron transfer in ruthenium modified cytochrome c. Bänsch *et al*.

[21] studied the effect of high pressure on the rate of inter-molecular electron transfer from reduced horse heart cytochrome c to Pentaammine (isonicotinamide) ruthenium(II) perchlorate ($[Ru(NH_3)_5\text{-isn}](ClO_4)_2$) using a custom high-pressure stopped flow absorption spectrometer. Bänsch *et al*. found that these transfer rates were sensitive to both temperature and pressure. Scott *et al*. [22] studied the sensitivity to pressure of electron transfer between attached metal-labels and the heme group in cytochrome b_5 molecules. Scott *et al.* [22] found that these intra-molecular transfer rates displayed an exponential dependence upon applied pressure.

The structural perturbations needed to actuate these functional changes are suggested by several high-pressure atomic structures of protein molecules. Kundrot and Richards [23, 24] solved the crystal structure of hen egg white lysozyme at a pressure of 100 MPa at room temperature using a high-pressure, X-ray transparent beryllium cell. Urayama *et al.* [4] solved the structure crystal structure of sperm whale myoglobin at a pressure of 200 MPa at room temperature using a high-pressure beryllium cell and at low temperature using a high-pressure cryo-cooling technique. Kim *et al.* [3] solved the structure of glucose isomerase at 130 MPa and of thaumatin at 185 MPa using the high-pressure cryo-cooling technique. Fourme *et al*. [25] solved the crystal structure of hen egg white lysozyme using an X-ray transparent high-pressure diamond anvil cell at pressures from 300 to 690 MPa. Refaee *et al.* [26] solved the structure of hen egg white lysozyme in solution at 200 MPa using a high-pressure NMR cell. Williamson *et al.* [27] solved the structure of bovine pancreatic trypsin inhibitor also using a high-pressure NMR cell. Collins *et al.* [28, 29] solved the crystal structure of the wild type and L99A mutant of T4 lysozyme using the high-pressure beryllium cell at pressures up to 200 MPa. Wilton *et al*. [30] reported the structure of protein G at 200 MPa using a high-pressure NMR cell. All of these high-pressure protein atomic structures indicate

that pressures of a few hundred MPa typically displace atoms by approximately 0.1 to 1 Å from their ambient pressure positions.

Numerical simulations of the structural deformation of cytochrome c mutants at high pressure by Miyashita and Go [31] and by Prabhakaran *et al.* [32] also indicate that deformations to protein structure induced by pressures of several hundreds of MPa are on the order 0.1 to 1.0 Å.

Observations of the small structural perturbations of protein structure by high-pressure concomitant with dramatic modification of protein function suggests that the exact positioning of atoms, especially in the active sites of catalytic proteins is an important feature of protein operation, and that this positioning is subject to environmental perturbation.

The number of reported high-pressure protein structures is extremely small in comparison to the total number of reported protein structures. The number of investigations that attempt to correlate pressure-induced sub-angstrom structural deformations to impacts upon the functional properties of a protein molecule is smaller still. The three most notable investigations of this type are the investigation of the high-pressure modification of the rate constant of morphinone reductase by Hay *et al*. [18], the investigation of the carbon monoxide recombination rate of myoglobin by high-pressure Raman spectroscopy by Galkin *et al*. [33] and the investigation of conformational substates of myoglobin using high-pressure X-ray crystallography by Urayama *et al*. [4].

1.2.1 Numerical Simulations Link Sub-Angstrom Structural Perturbations in the Active Site of Morphinone Reductase to an Accelerated Rate Constant

The flavoprotein morphinone reductase is expressed by the gram-negative bacterium *Pseudomonas putida* M10. *P. putida* M10 was isolated from opiate factory effluent[1] [34, 35] and utilizes morphine alkaloids as its sole energy and carbon source[2] [35, 36].

The first two steps of the degradation of morphine alkaloids by *P. putida* M10 are catalyzed by the enzymes morphine dehydrogenase and morphinone reductase [36]. Morphine dehydrogenase catalyzes the oxidation of morphine to morphinone and codeine to codeinone in the presence of $NADP^+$ [35]. Morphinone reductase catalyzes the saturation of the carbon-carbon double bond of morphinone and codeinone yielding hydromorphone and hydrocodone in the presence of NADH [37]. A cartoon stereo diagram of morphinone reductase is shown in figure 1.1.

To perform the reduction of morphinone and codeinone, morphinone reductase acquires a hydride (H^-) ion from NADH that is transiently stored on the flavin mononucleotide (FMN) prosthetic group of the enzyme. This stored charge is used to reduce morphinone to hydromorphone or codeinone to hydrocodone. The NADH oxidation half-reaction proceeds in three distinct steps: docking of the NADH cofactor and morphinone reductase, formation of a charge transfer (CT) complex [37] and

---

[1] It is unclear if *Pseudomonas putida* M10 occurs naturally in the waste streams of heroin factories. French *et al*. commented that *P. putida* M10 was isolated from opiate factory waste liquor, implying that the bacterium does occur naturally in this environment. However, the citation used to support this statement by Bruce *et al*. refers to the source of the organism only as industrial waste liquor enriched with morphine.

[2] Bruce *et al*. noted that although *P. putida* M10 can utilize morphine as its sole energy and carbon source, it is cultured with supplemental glucose.

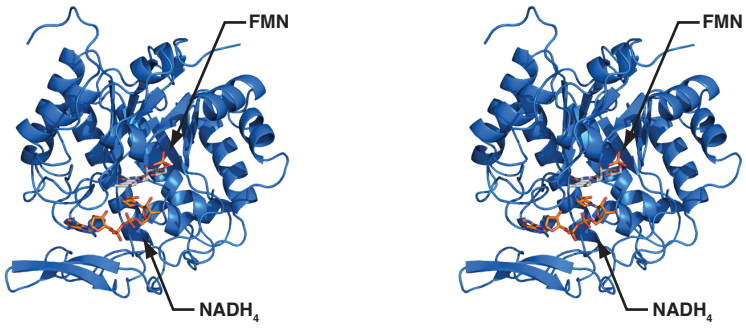


Figure 1.1: Stereo image of the structure of Morphinone Reductase from *Pseudomonas putida* M10 by Pudney *et al.* [38]. The cofactor analog $NADH_4$ [38] and flavin prosthetic group (FMN) are highlighted. PDB accession code 2R14.

finally by the transfer of a hydride ion from the C4 atom of the NADH to the N5 atom of the FMN group [18]. Hay *et al.* [18] used high-pressure perturbation of morphinone reductase to probe the mechanism of hydride transfer from the NADH to the FMN prosthetic group. The FMN group and $NADH_4$ cofactor analog [38] are highlighted in figure 1.1. A diagram of the FMN group and $NADH_4$ cofactor analog is shown in figure 1.2. The direction of hydride transfer and the separation of the $NADH_4$ C4 and FMN N5 atoms, *d*, are shown in figure 1.2.

The rate of hydride transfer from NADH to the FMN prosthetic group in morphinone reductase is highly dependent upon pressure, temperature and the hydrogen isotope composition of the NADH cofactor [18]. Hay *et al.* [18] measured the hydride transfer rate from NADH and NAD^2H (deuterated NADH) by time resolved observation of the absorbance of the FMN prosthetic group at 464 nm in a high-pressure stopped flow apparatus at pressures ranging from atmospheric to 200 MPa, and temperatures ranging from 5 to 40 ˚C. The ratio of the protium (k_H) to the deuterium ion transfer

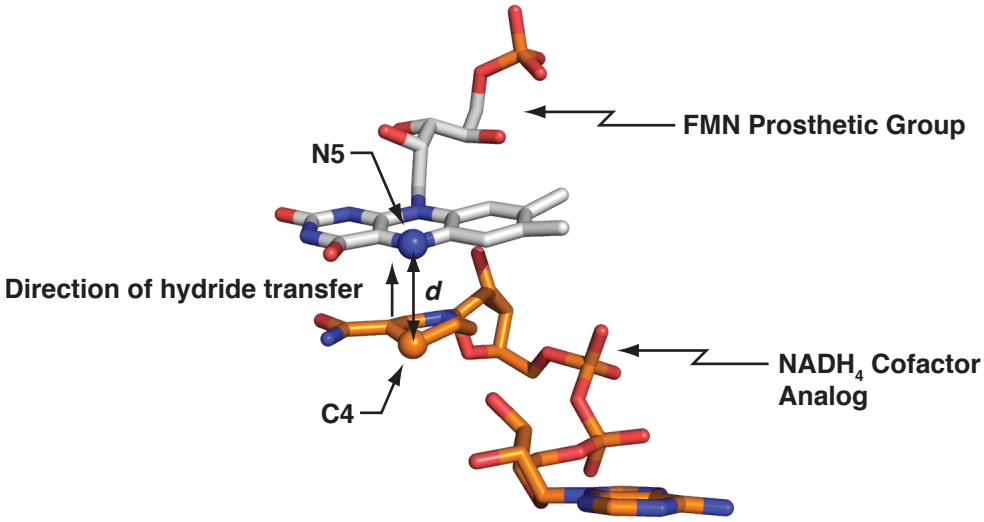


Figure 1.2: The Morphinone Reductase flavin prosthetic group (FMN) and non-oxidizable cofactor analog $NADH_4$ from the crystal structure by Pudney *et al.* [38].

rate (k_D) gives the Kinetic Isotope Effect (KIE) for the reaction. The change of the Kinetic Isotope Effect with pressure and temperature is revealing of the hydride transfer mechanism. The KIE is defined as

$$KIE = \frac{k_H}{k_D}. \tag{1.1}$$

Hay *et al.* [18] demonstrated that the transfer rates of both protium and deuterium ions from the NADH cofactor to the FMN prosthetic group of morphinone reductase increased exponentially with applied pressure.

Semi-classical transition state models assume that proton transfer between reactant and product states occurs by passage of the proton over a potential barrier between the reactant and product states. The rate of passage over the barrier is thus only a function of the height of barrier and of the attempt frequency, not of the barrier width. Semi-

classical transition state theory models predict that the application of high pressure should only affect the vibrational frequencies of atoms involved in proton transfer [39, 40], thus only affecting the attempt frequency for barrier crossing. Different hydrogen isotopes should vibrate at different frequencies due to their different masses, resulting in different transfer rates at room temperature and pressure for different isotopes. Experiments by Isaacs *et al.* [39] indicate that these stretching vibrations are insensitive to pressures lower than a few hundred MPa. Thus, pressures in the range of a few hundred MPa should not affect the Kinetic Isotope Effect ratios of proton transfer reactions that require passage of the proton over a potential barrier.

However, modern transition state theories do predict that in certain reaction geometries, high pressure may increase a hydride transfer rate by increasing the probability of potential barrier penetration by quantum mechanical tunneling through the potential barrier between the reactant and product states [18]. Thus, the transfer rate through the barrier is a function of the barrier width, as well as the barrier height. As different hydrogen isotopes have different masses, and thus different de Broglie wavelengths, the rate of transfer of the two different isotopes will vary with different rates with increasing pressure. Thus, the Kinetic Isotope Ratio of the reaction will be sensitive to the application of high pressure.

Hay *et al*. [18] found that the Kinetic Isotope Effect for the hydride transfer reaction from NADH (or NAD^{2}H) to the FMN group in morphinone reductase varies with temperature and pressure.

Hay *et al*. [18] proposed a numerical quantum mechanical model based upon small, pressure-induced structural deformations to the active site of morphinone reductase

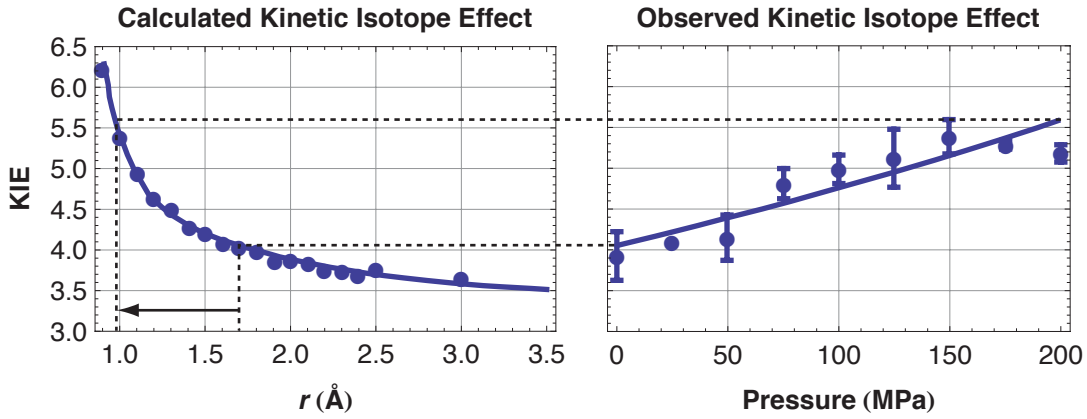


Figure 1.3: Observed and calculated Kinetic Isotope Effects (KIE) for Morphinone Reductase hydride transfer reaction. Observed KIE is shown as a function of pressure and is correlated with the calculated KIE, shown as a function of proton donor to acceptor separation, *r*. Adapted from Hay *et al*. [18].

that would actuate this large, exponential change in hydride transfer rate and kinetic isotope effect with increasing pressure. Hay *et al.* [18] proposed that hydride ions transfer from NADH to the FMN prosthetic group by tunneling through a potential barrier. The width of this barrier, defined by the separation of the proton donor (the C4 atom on the NADH cofactor) and the proton acceptor (the N5 atom on the FMN prosthetic group): *r*, may be reduced by the application of high pressure, thus significantly increasing the probability of tunneling, and thus the proton transfer rate. The direction of hydride transfer and the separation of the $NADH_4$ C4 and FMN N5 atoms, *d*, are shown in figure 1.2. To clarify, *d* is the separation of the C4 and N5 atoms, while *r* is the separation of the proton donor and acceptor sites attached to these atoms. X-ray crystallography can resolve the positions of the C4 and N5 atoms, but not the acceptor sites. However, the proton transfer model by Hay *et al*. [18] is presented in terms of donor to acceptor site distance, *r*.

An X-ray structure of morphinone reductase complexed with the non-oxidizable cofactor analog $NADH_4$ by Pudney *et al.* [38] indicates that the C4 to N5 separation: *r*, is approximately 3.45 Å at room pressure and cryogenic temperatures. This suggests that the proton donor to acceptor separation is approximately 1.7 Å under these conditions [18]. Hay *et al*. [18] speculate that high pressure compresses the distance between the NADH cofactor and FMN prosthetic group, facilitating tunneling of hydride ions from the cofactor to the prosthetic group. A numerical tunneling model by Hay *et al*. [18] predicts that the donor to acceptor distance: *r*, must reduce from 1.7 Å at 0.1 MPa to 1.0 Å at 200 MPa, given an activation enthalpy of 5.2 kJ mol^{-1}, in order to account for the observed change in Kinetic Isotope Effect by this pressure. A plot of the calculated change in Kinetic Isotope Effect corresponding to this decreasing separation is shown alongside the measured change in Kinetic Isotope Effect versus pressure in figure 1.3.

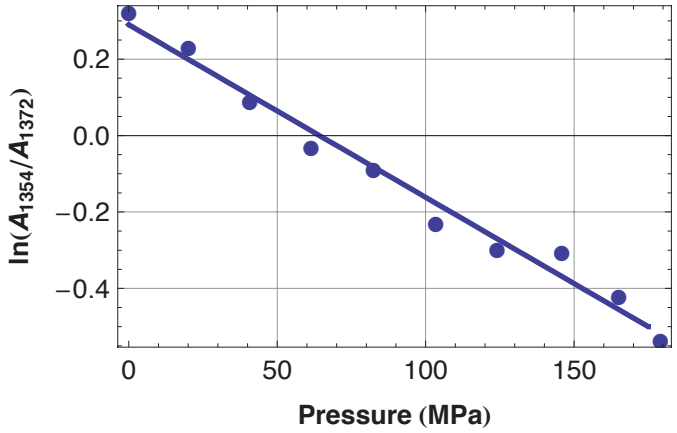
Attempts, not fully successfully, to solve the structure of morphinone reductase under high pressure and confirm the model proposed by Hay *et al*. [18] are outlined in chapter 6 of this thesis.

1.2.2 High Pressure Raman Spectroscopy Links Sub-Angstrom Perturbations of the Heme Group of Myoglobin to an Acceleration of the Carbon Monoxide Association Rate

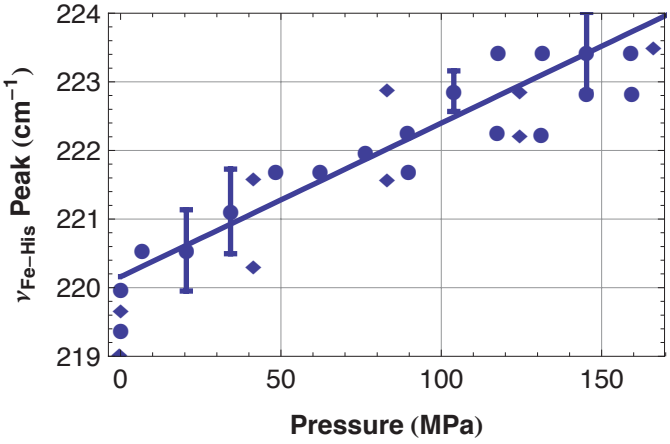
The behavior of the heme proteins hemoglobin and myoglobin under high pressure has received considerable attention. These investigations have allowed a structural explanation of the behavior of ligand re-binding in myoglobin under high pressure to emerge. Galkin *et al*. [33] used high-pressure Raman spectroscopy to infer a correlation between pressure-induced changes in the carbon monoxide re-binding rate

Figure 1.4: A: Ratio of Mb to MbCO absorption markers with increasing pressure, indicating switch in dominant population from Mb at 0.1 MPa to MbCO at 175 MPa. B: Iron-proximal histidine stretch mode band frequency shift with pressure, indicating motion of heme iron into plane of heme group. C: Heme iron charge transfer transition band, indicating motion of heme iron into plane of heme group. Adapted from Galkin *et al*. [33].

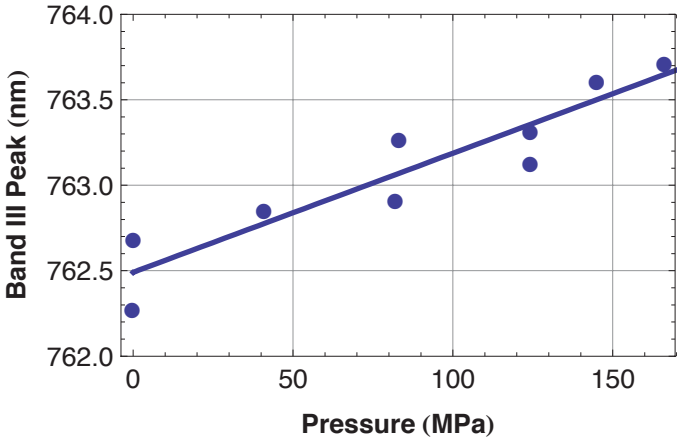
A: Ratio of Mb (1354 nm) and MbCO (1372 nm) Absorption Peaks



B: Iron-Proximal Histindine Stretching Mode Raman Peak



C: Heme Iron $a_{2u} \rightarrow d_{yz}$ Charge Transfer Transition Peak



of horse met-myoglobin and a structural change in and around the ligand-binding heme group of the molecule.

As with hemoglobin, the out of plane position of the myoglobin heme iron is sensitive to the ligand binding state of the molecule. When oxygen, or carbon monoxide is bound to the heme, the heme iron lies in the plane of the heme group. When no ligand is bound to the myoglobin heme, the heme iron puckers out of the plane of the heme group by approximately 0.2 to 0.3 Å [41-43] (the heme iron puckers by ≈ 0.4 Å in hemoglobin). The puckering motion of the heme iron is transmitted to the F helix by the proximal histidine that connects the heme group to the F helix [43].

Galkin *et al.* [33] made steady state measurements of the Raman spectra of a carbon-monoxy-myoglobin sample photolyzed by the Raman excitation beam. The Raman peaks at 1354 and 1372 cm^{-1} are proportional to the populations of photolyzed Mb and recombined MbCO in the sample respectively. The ratio of the 1354 and 1372 cm^{-1} peaks is shown in figure 1.4A. The ratio of these two peak intensities is equal to the population ratio of Mb to MbCO and is thus representative of the rate of recombination of Mb with carbon monoxide. Photolyzed unbound Mb is the dominant species in the sample at atmospheric pressure, while MbCO becomes the dominant fraction at pressures exceeding 70 MPa. This change in the dominant population suggests that the recombination rate of CO to Mb increases with pressure, consistent with the time resolved flash photolysis results of Adachi and Morishima [16] on horse, sperm whale and dog myoglobin.

Galkin *et al*. [33] linked this functional change of myoglobin, an increase in ligand rebinding rate, to a sub-angstrom structural change in the myoglobin molecule by

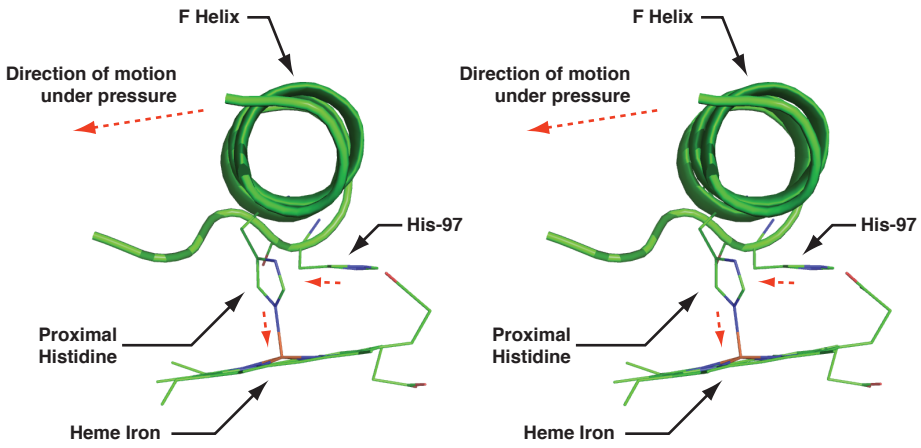


Figure 1.5: Stereo diagram showing proposed pressure-induced structural changes in myoglobin structure. Adapted from Galkin *et al*. [33].

examining peaks in the Raman scattering spectrum of myoglobin. These Raman peaks are highly sensitive indicators of the structural state of the active site.

The most significant change in the myoglobin Raman spectrum is the linear shift of the iron-proximal histidine stretching mode: $\nu_{\mathrm{Fe-His}}$ [44, 45] with increasing pressure. The iron-proximal histidine stretching mode band shifts by approximately 4 cm^{-1} from 0.1 to 175 MPa. The $\nu_{\mathrm{Fe-His}}$ stretch mode frequency versus pressure is shown in figure 1.4B.

Galkin *et al*. [33] also examined the near infrared absorption band III centered at 762 nm at 0.1 MPa. The band III absorption peak is speculated to correspond to the porphyrin a_{2u} Highest Occupied Molecular Orbital (HOMO) to d_{yz} iron orbital charge transfer transition. The position of the band III absorption peak as a function of pressure is shown in figure 1.4C. The band III absorption peak shifts by approximately 1.2 nm from 0.1 to 175 MPa. It is speculated that a shift in the position of the infrared

band III absorption peak is related to the heme iron out of plane position [46, 47] and possibly to the relative orientation of the heme and proximal histidine [48].

Calculations by Stavrov [49] suggest that the pressure-induced shift of the iron-proximal histidine stretch mode peak ($\nu_{\mathrm{Fe-His}}$) corresponds to a motion of the heme iron of approximately 0.015 Å toward the plane of the heme group. Calculations by Galkin *et al.* [33] support the existence of this deformation motion by suggesting that the shift of the band III absorption peak corresponds to a motion of the heme iron of approximately 0.01 Å into the plane of the heme group. This motion of the heme iron towards the plane of heme group places the heme in a structural state more like that observed in ligand-bound myoglobin. A smaller spectral shift of the 343 cm^{-1} Raman peak is speculated to correspond to a motion of the myoglobin helices [33].

Galkin *et al*. [33] speculated that the extremely small heme iron motion suggested by their results is actuated by a small pressure-induced motion of the myoglobin F helix through the proximal histidine linking the heme to the helix. The proposed motion of the F helix and its linkage to the heme is shown in figure 1.5.

Galkin *et al*. [33] speculated that the motion of the proximal histidine induced by high pressure increases the planarity of heme group, reducing the barrier to ligand recombination. As ligand-binding to the heme is the rate-limiting step in the binding of myoglobin to its ligand, this small deformation to the heme group may greatly increase the ligand rebinding rate.

1.2.3 Conformational Substates of Myoglobin Investigated by High-pressure X-ray Crystallography

Ansari *et al*. [50] demonstrated that carbon-monoxy sperm whale myoglobin displays three conformational substates associated with the infrared stretch bands of the heme-bound carbon monoxide. The A_0 band is centered at ≈ 1967 cm^{-1}, the A_1 band at ≈ 1947 cm^{-1} and the A_3 band at ≈ 1929 cm^{-1}. Frauenfelder *et al*. [51] speculated that these substates may possess different reactive properties, giving myoglobin ligand binding and reactive properties not traditionally associated with single subunit, non-allosteric enzymes. Urayama *et al*. [4] noted that an understanding of the structural properties of these conformational substates may shed light on the structural basis of myoglobin's anomalous reactive properties.

The relative populations of the A_0, A_1 and A_3 substates of myoglobin are sensitive to pressure, temperature and pH [12, 50]. At room temperature and pressure, a pH decrease from 6.6 to 5.5 increases the A_0/A_1 population ratio by a factor of approximately 10. Similarly, a pressure increase from 0.1 to 200 MPa at room temperature increases the A_0/A_1 ratio by factor of 2.8. These results suggest a similarity between the pH and pressure modified structures of myoglobin.

Urayama *et al.* [4] solved the structure of carbon-monoxy sperm whale myoglobin using room temperature high-pressure X-ray crystallography techniques and a low-temperature high-pressure cryo-cooling technique [3, 4]. Urayama *et al.* [4] suggested that these high-pressure structures of MbCO may provide a structural explanation for the pressure-dependent infrared spectroscopic and reactive properties of MbCO.

Urayama *et al*. [4] noted similarities between the structure of carbon-monoxy-myoglobin at high-pressure and the structure of the molecule at low pH, concomitant with similar changes in the A_0 to A_1 population ratio under both pressure and pH perturbation. Urayama *et al*. [4] speculated that the small pressure-induced global structural changes of myoglobin corresponded to those associated with the A_0 state. The largest pressure induced displacements in myoglobin occurred in the F helix, AB loop, and CD loop regions. The start of the F helix displaces toward by heme group by approximately 0.25 Å under a pressure of 200 MPa. This displacement of the F helix is particularly significant as this helix is connected to the heme group through the proximal histidine (His93). The displacement of the F-helix suggests a mechanism for pressure to alter the conformation of the heme and hence its spectroscopic and reactive properties.

1.3 Preview of Thesis

The work presented in this thesis builds upon the examples presented by Hay *et al.* [18], Galkin *et al.* [33] and Urayama *et al*. [4], and provides a direct correlation between the directly measured, continuously perturbed X-ray crystal structures of a protein molecule under a range of high pressures and a continuously varying functional change of the molecule.

To systematically investigate the correlation between protein function and small structural deformations, we chose to study the *Aequorea* Yellow Fluorescent Protein (YFP) Citrine [1, 2, 52] under high pressure. Citrine is an extremely bright, intrinsically fluorescent protein whose atomic structure is known to 2.2 Å resolution [1] and displays a fluorescence peak shift of approximately 1 nm per 100 MPa at room temperature [53].

Chapter 2 of this thesis discusses methods for solving the structures of protein molecules under high pressure, and the possibility and reliability of detecting small deformations in the structure of a protein molecule well below X-ray wavelengths and the diffraction resolution limit of protein crystals.

Chapter 3 of this thesis discusses the Citrine molecule, its photophysics, and it a possible mechanism for the response of its fluorescence spectrum to high pressures.

Section 4.2 of chapter 4 details the growth and purification of Citrine. Section 4.3 describes the construction of a safety enclosure for the high-pressure cryocooling apparatus. Section 4.4 describes a micro-spectrophotometer that was used to measure the optical properties of Citrine under high-pressure cryocooling conditions. Section 4.5 of this chapter describes the effects of high pressure on the fluorescence spectrum of Citrine. Section 4.6 describes the crystallization of Citrine.

A closely spaced series of high-pressure structures of Citrine were solved using a novel high-pressure cryo-cooling X-ray crystallography technique developed by Kim *et al*. [3]. Details of data collection from these crystals is described in section 4.7 of chapter 4. Analysis of these high-pressure structures of Citrine reveal a progressive deformation of the chromophore that correlates with a progressive blue-shift of the fluorescence peak of Citrine observed under these high-pressure cryo-cooling techniques that was discussed in section 4.5 of chapter 4. The detailed structural analysis of Citrine's chromophore under high pressure is described in section 4.8 of chapter 4.

The deformation of the Citrine chromophore is actuated by the relative translation and rotation of two clusters of atoms with differing compressibilities that compose the β-barrel scaffold of the Citrine molecule. This relative motion of the two clusters composing the β-barrel scaffold is reminiscent of a bending or buckling of the β-barrel [6]. The identification of these two clusters of atoms, and the mechanism of actuation of the chromophore deformation is discussed in chapter 5.

The series of high-pressure structures of the Citrine molecule also reveal a perturbation of the hydrogen bonding network stabilizing the excited state of the Citrine chromophore. This perturbation is implicated in the reduction of fluorescence intensity of the Citrine molecule under high pressure and is discussed further in chapter 5.

Recent single molecule experiments on enzymes including cholesterol oxidase [54], hairpin ribozyme [55], λ exonuclease [56], lipase [57, 58] and the Yellow Fluorescent Protein [7] indicate that under ambient conditions, single protein molecules switch between a series of conformational states with distinct functional properties [59]. These distinct functional properties may correspond to structural changes on the angstrom to sub-angstrom level in the case of flavin reductase [60], or on the 0.1 to 0.2 Å length-scale in the case of Photoactive Yellow Protein (PYP) [61]. These functional substates are highly reminiscent of the three functional and conformational substates: A_0, A_1 and A_3 of myoglobin that were investigated by Frauenfelder *et al*. [51] and were discussed in section 1.2.3.

Photoactive Yellow Protein (PYP) is a bacterial photo-receptor [62, 63] capable of forming crystals that diffract to extremely high resolution (≈ 0.82 Å) [64]. At

cryogenic temperatures, PYP is capable of executing the first five steps in its photocycle, but becomes trapped in the spectroscopically identifiable state I_1 of the photocycle. This should allow a population of PYP molecules to be completely converted to state I_1. However, based upon optical absorption measurements, Coureux *et al*. [61] concluded that two populations of conformers exist in a sample of flash frozen PYP, despite apparent structural homogeneity in a model of the protein derived from 0.82 Å resolution X-ray diffraction data. One of these populations, comprising approximately 30% of the sample is photoactive, meaning that it can be excited to state I_1. However, the remaining population is photo-inactive. Coureux *et al*. [61] concluded, given an error of approximately 0.1 to 0.2 Å on the PYP structure, that the active and non-active conformers must have structures differ by only 0.1 to 0.2 Å.

Blum *et al*. [7] observed that isolated molecules of Yellow Fluorescent Protein, a close relative of Citrine, transiently convert to a blue-shifted form, with a fluorescence peak at ≈ 514 nm. This blue-shifted fluorescence peak is the same as that observed from Citrine under high-pressure cryo-cooling at 350 MPa. Chapter 5 discusses the possibility of using high-pressure cryo-cooling to gain structural insight into transiently occupied protein states such as those seen for Yellow Fluorescent Protein [7], cholesterol oxidase [54], hairpin ribozyme [55], λ exonuclease [56] and lipase [57, 58]. Possible mutations to the Citrine structure, highlighted by the high-pressure structures of Citrine, that may alter its time dependent behavior to favor the blue-shifted state are discussed in chapter 6.

We attempted to apply the high-pressure X-ray crystallography techniques used to analyze the structure and functional alteration of Citrine under high pressure to Morphinone Reductase. These attempts are outlined in chapter 6.

In chapter 6, we discuss possible future high-pressure X-ray crystallography experiments, and the possibility of using high-pressure cryo-cooling to identify mutation sites in protein structures is discussed.

CHAPTER 2

SOLVING THE STRUCTURE OF PROTEIN MOLECULES AND PRECISION OF PROTEIN STRUCTURES

2.1 Introduction

The 3-dimensional structure of protein molecules is intimately tied to their function. Knowledge of the 3-dimensional structure of a protein molecule is a necessary, but often insufficient, condition for understanding the mechanism of action of the molecule. Computer algorithms exist that predict the 3-dimensional structure of a protein molecule. A well-known technique is template-based modeling, where a known structure of a protein molecule with a similar primary sequence to the target is used to generate a *homology model* [65]. Homology modeling has shown some success in predicting the 3-dimensional structure of protein-molecules from linear amino acid sequences [66, 67]. Assuming a primary sequence identity of 35% between a template and a target structure, the SWISS-MODEL homology modeling server achieves a mean rms (root mean square) deviation of α-carbon atom positions between the homology model and an experimental structure of the molecule of 3.5 Å ± 1.6 Å (supporting material to Arnold *et al*. [66]). For 95% sequence identity, the mean rms deviation between the homology model and experimental structure of a molecule is ≈ 1.2 Å. Similar modeling programs achieve similar levels of accuracy to SWISS-MODEL for sequence identities greater than 35% (supporting material to Arnold *et al*. [66]).

While the rms deviations between homology models and experimental structures are small in comparison to the exterior dimensions of a protein molecule, typically tens of angstroms, they are large in comparison to the structural deviations induced by high pressure. For comparison, the rms deviation between α-carbon atoms of the low temperature, room pressure (Protein Data Bank (PDB) accession code 1A6K) and low

temperature, high pressure (1JP9) structures of myoglobin analyzed by Urayama *et al.* [4] is only 0.132 Å.

Additionally, there are notable examples of protein molecules whose structures may not be amenable to homology modeling. For example, mutation of a single residue in Rat Liver S-Adenosylhomocysteine Hydrolase produces large changes in the conformation and catalytic behavior of the enzyme [68]. More importantly, there appears to be no general rule to estimate the precision of individual results of structural modeling. Given that the function of protein molecules may be sensitive to sub-angstrom structural perturbations, it is presently difficult to accurately predict and understand the catalytic behavior of molecules based only upon computer-simulated models.

As of today, two complementary techniques exist for finding the structures of macromolecules with atomic level precision: X-ray crystallography and Nuclear Magnetic Resonance (NMR). Both X-ray crystallography and NMR have been adapted for high pressure use. Kundrot and Richards [23] developed a high-pressure X-ray transparent beryllium cell for macromolecular crystallography at room temperature. This cell was used to solve the structure of hen egg white lysozyme at a pressure of 100 MPa [24, 69]. Urayama *et al*. [4, 70] developed a high-pressure cryo-cooling technique that is used to freeze protein crystals under high pressure and lock in pressure induced structural deformations. Urayama *et al.* [4] used both the high-pressure cryo-cooling technique and a high-pressure beryllium X-ray cell to solve the structure of sperm whale myoglobin at pressures up to 200 MPa. Kim *et al.* [3] further developed the high-pressure cryo-cooling technique. Kim *et al.* demonstrated that the high-pressure cryo-cooling technique effectively reduces the damage to protein

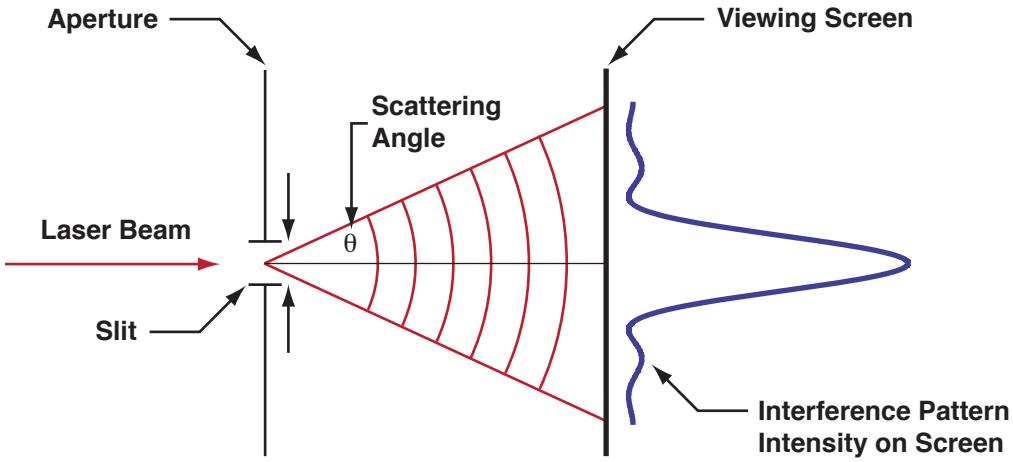
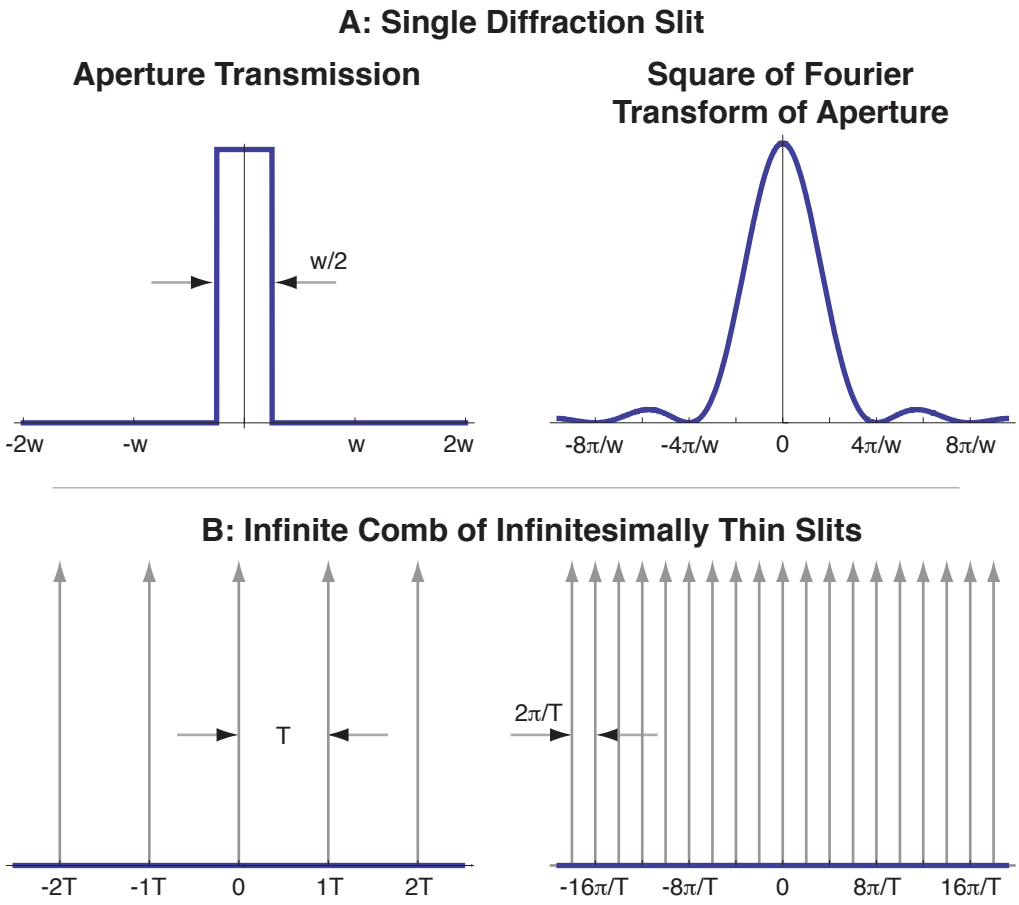


Figure 2.1: Schematic of optical diffraction experiment. A laser beam is passed through a narrow slit, causing interference of the beam. The interference pattern is projected on a screen behind the aperture. The intensity of the interference pattern is given by the Fourier transform of the aperture.

crystals due to freezing to cryogenic temperatures [3]. Fourme and co-workers developed a diamond anvil cell suitable for macromolecular X-ray crystallography at pressures up to 1 GPa [25, 71]. Fourme and co-workers have used the macromolecular diamond anvil cell to solve the structures of hen egg white lysozyme [25], cow pea mosaic virus [72-74] and a short oligonucleotide [75]. Refaee *et al.* developed a high-pressure NMR cell that permits the room temperature, high-pressure solution structure of macromolecules to be solved at pressures up to 200 MPa [26]. The high-pressure cryo-cooling technique was used to solve all new protein atomic structures discussed in this thesis.

Figure 2.2: A: The transmission through a single slit, and the square of the Fourier transform of the slit. B: An infinite comb of delta functions and its Fourier transform. C: An infinite comb of finite width slits, and its Fourier Transform. The scattering angle referred to in the figure is defined in figure 2.1.


A: Single Diffraction Slit
Aperture Transmission
Square of Fourier Transform of Aperture
w/2
-2w
-w
w
2w
-8π/w
-4π/w
0
4π/w
8π/w
B: Infinite Comb of Infinitesimally Thin Slits
T
2π/T
-2T
-1T
0
1T
2T
-16π/T
-8π/T
0
8π/T
16π/T
C: Infinite Comb of Finite Slits
Diffraction spot
-2T
-1T
0
1T
2T
-16π/T
-8π/T
0
8π/T
16π/T
Distance Along Aperture
Reciprocal Distance = Sin(Scattering Angle) / Wavelength

2.2 Introduction to X-ray Crystallography

In order to solve the structure of protein molecules to atomic or close to atomic resolution a probe is required with a wavelength comparable to or smaller than atomic dimensions. Hard X-rays have wavelengths ranging from approximately 10 Å to 0.1 Å, corresponding to photon energies of 1.2 to 120 keV. This wavelength scale makes hard X-rays appropriate for resolving atomic scale features in molecules: solving the structure of the molecule. However, as of the present day, no technique exists to focus X-rays sufficiently to permit imaging with angstrom scale resolution to permit hard X-ray microscopy. In order to solve the structure of a molecule, the diffraction pattern produced by the interaction of X-rays with an atomic scale structure must be analyzed in order to elucidate the structure of a molecule. X-ray crystallography has been used since the first decade of the 20th century to solve the atomic structures of increasingly complex molecules [76, 77].

X-ray crystallography was the first technique successfully applied to solve the 3-dimensional atomic structure of a protein molecule [78, 79]. Over the past five decades, X-ray crystallography of protein molecules has become an increasingly refined and easily applied technique [76, 77].

To gain an intuitive understanding of X-ray crystallography, it is useful to consider optical diffraction: the events that occur when light, for instance a laser beam, passes through, and is scattered by an aperture with dimensions comparable to its wavelength. The interference pattern that results from this scattering is projected on a screen behind the aperture. The interference pattern is the modulus squared of the Fourier transform of the aperture. A diagram illustrating the geometry of this interference experiment is shown in figure 2.1. The transmission of the slit, and the

square modulus of its Fourier Transform are shown in figure 2.2A. In a similar fashion, the X-ray diffraction pattern produced by the interference of X-rays passing through a crystal and projected onto a film, image plate or CCD (charge coupled device) X-ray detector is related to the Fourier transform of the protein crystal.

If the laser beam is passed through a comb of slits of finite width spaced at regular intervals, the resulting interference or diffraction pattern projected onto the screen behind the aperture is also a repeating series of regularly spaced sharp peaks that becomes increasingly dim as one moves away from the center of the pattern. This is shown in figure 2.2C.

The aperture can be thought of as a single finite width aperture convolved with a comb of infinitesimally thin apertures, or delta functions [80]. An infinitely long comb of delta functions is shown in figure 2.2B. The Fourier transform of an infinite comb of delta functions is also an infinite comb of delta functions. The spacing of the delta functions in the Fourier transform is inversely proportional to the spacing of delta functions in real space [80]. For instance, if the delta functions are spaced at intervals of T in real space, the spacing in Fourier space will be proportional to $1/T$. This is shown in figure 2.2B.

An infinite comb of slits of finite width, like that shown in figure 2.2C, can be thought of as the convolution of a single slit of finite width, like that shown in figure 2.2A, with an infinite of comb of delta functions, like that shown in figure 2.2B. The Fourier transform of the convolution of two functions is equal to the product of the Fourier transforms of these two functions [80]. The Fourier transform of the infinite comb of finite width slits is equal to the Fourier transform of a single finite slit multiplied by

the Fourier transform of a train of delta functions. The interference pattern viewed on the screen is equal to the square of the Fourier transform of the infinite comb of finite width slits. The interference pattern appears as a regularly spaced pattern of peaks. The spacing of the peaks is inversely proportional to the spacing of the slits, while the intensity of each peak is determined by the Fourier transform of a single slit. This interference pattern is shown in figure 2.2C.

A crystal contains many identical copies of a molecule arranged into a lattice: a periodic structure. One can think of the lattice as a train of delta functions in 3-dimensions, and the crystal as a single molecule convolved with these delta functions. Thus, the interference, or diffraction pattern from the crystal is related to the Fourier transform of the lattice, multiplied by the Fourier transform of a single molecule.

As only the electrons of the atoms in the molecule interact appreciably with the X-ray beam, the Fourier transform of the molecule is the Fourier transform of the electron distribution in the molecule, often called the *electron density*. As the Fourier transform of the crystal, or the volume of the crystal illuminated by the X-ray beam, is related to the Fourier transform of the crystal lattice, multiplied by the Fourier transform of a single molecule, the diffraction pattern is confined to sharp peaks, called *Bragg spots* [81]. It is important to remember that the optical case described earlier, where a laser beam passes through a slit does not describe all of the features of X-ray diffraction from a crystal. The most important feature at which these two cases diverge is dimensionality. The example of a laser beam passing through an aperture is confined to 1-dimension, whereas a protein crystal is a 3-dimensional structure. For this reason, additional measures are needed to describe the diffraction pattern from the crystal. A representative diffraction pattern from a crystal of the protein Citrine is shown in

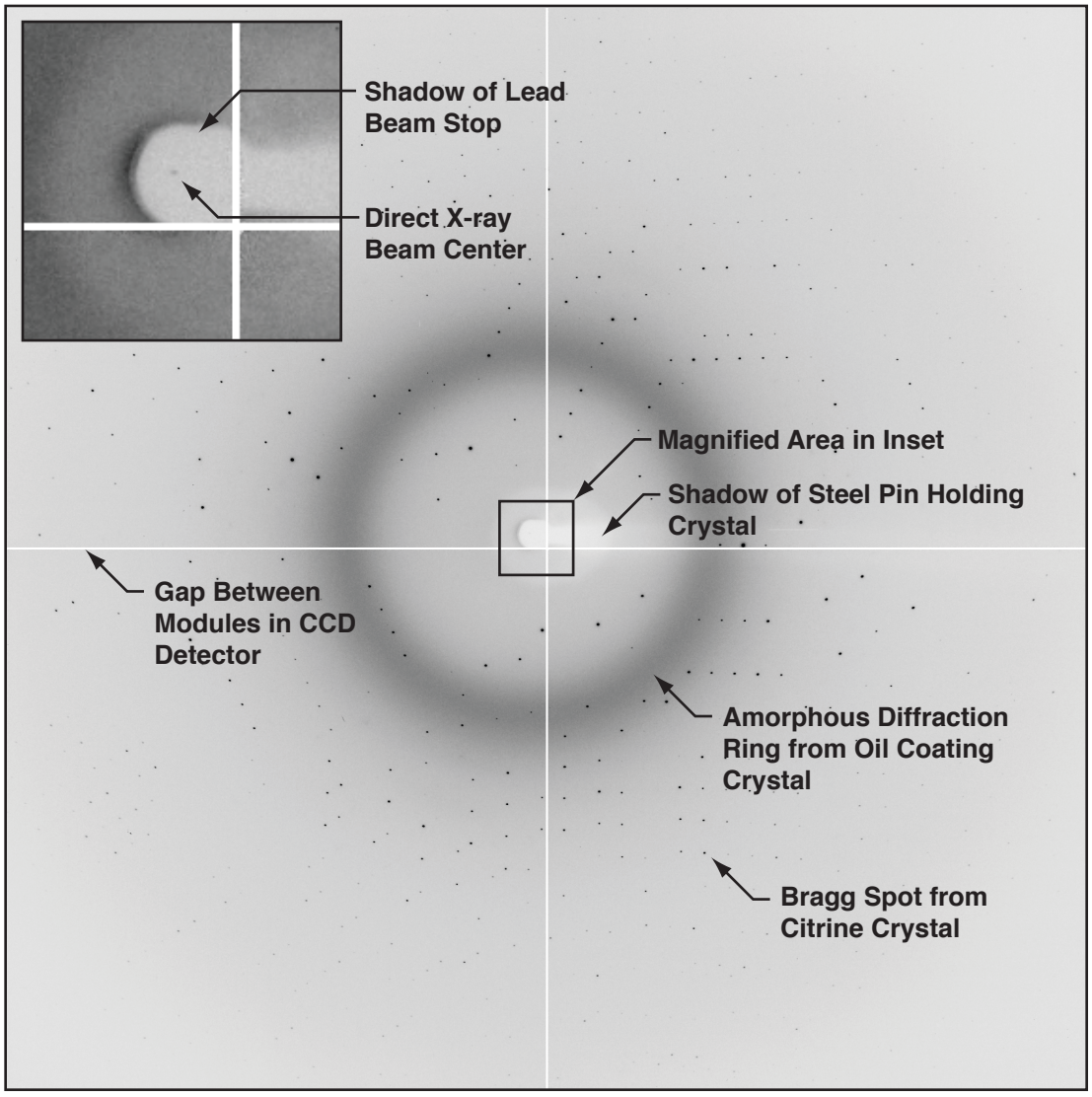


Figure 2.3: X-Ray diffraction pattern from a Citrine crystal taken at Cornell High Energy Synchrotron Source (CHESS) station F2 with and Area Detector Systems Corporation (ADSC, Poway, CA, USA) Quantum 210 CCD X-ray detector.

figure 2.3. The diffraction image also contains a prominent ring that comes from oil used to coat the crystal. As the oil does not have an ordered structure, it cannot cause Bragg spots to appear in the diffraction pattern. The ring diameter is representative of the spacings of carbon atoms within and between oil molecules.

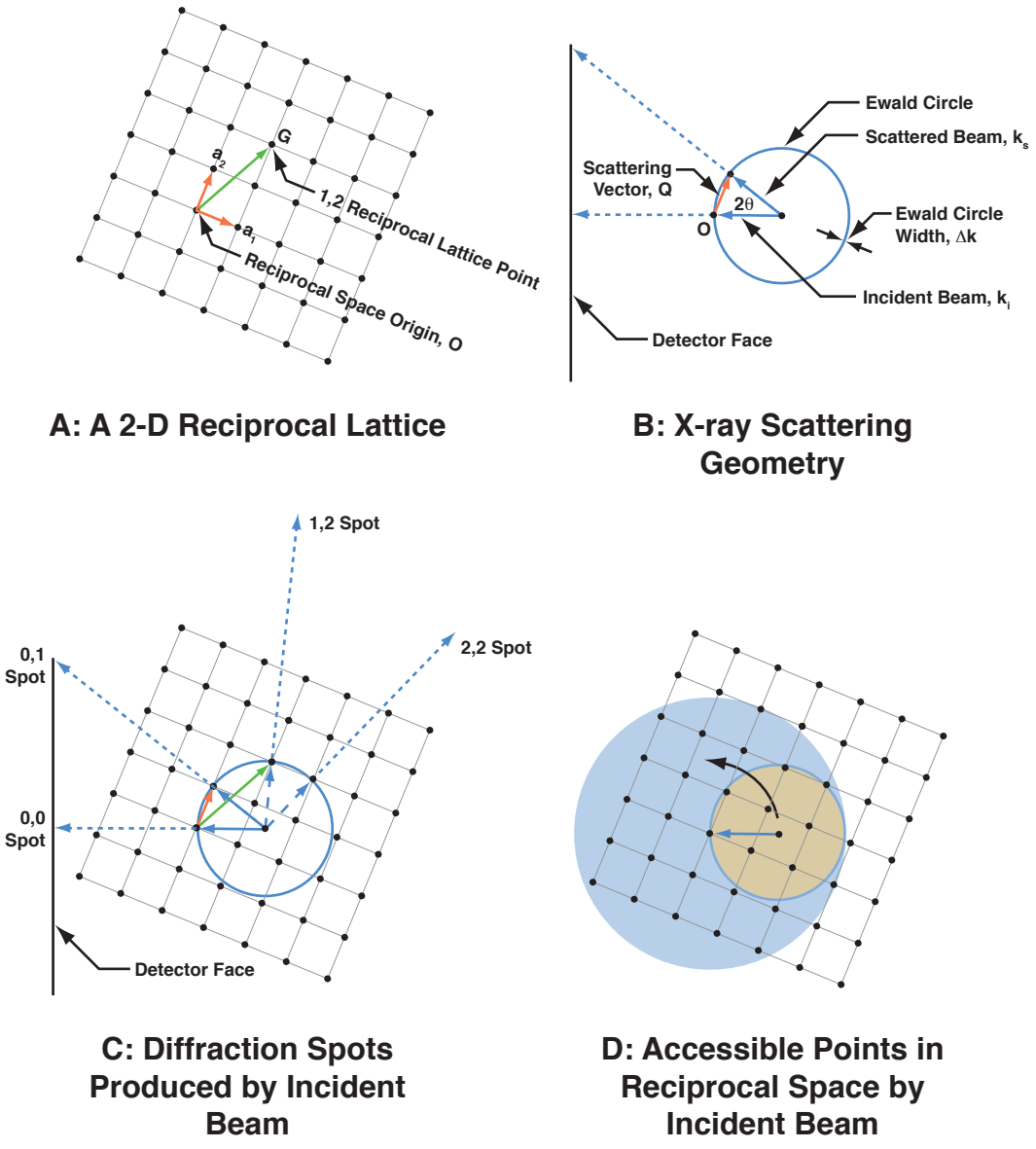


Figure 2.4: Ewald Circle construction. Adapted from Als-Nielsen and McMorrow [82].

A useful construction to describe the X-ray diffraction from a crystal is the *Ewald sphere* (pronounced Ee-vold) [82]. A two-dimensional version of the Ewald sphere, the *Ewald circle*, is shown in figure 2.4. The Ewald construct is often most easily understood, and drawn, in a two-dimensional construct, and can later be generalized to three dimensions. Figure 2.4A shows the *reciprocal lattice* of the crystal: the Fourier transform of delta functions representing the crystal lattice in real space. As in the case of the optical example discussed earlier, the spacing of the reciprocal lattice points is inversely proportional to the spacing of the lattice points in real space. Two vectors, a_1 and a_2 are shown in figure 2.4A that establish a set of basis vectors for this 2-dimensional space. In this example, the basis vectors are orthogonal and of equal length, but need not be, and often are not. Any reciprocal lattice vector can be constructed from a combination of these two reciprocal lattice vectors. A reciprocal lattice vector, *G*, projecting from the origin, *O*, is shown in figure 2.4A.

The Bragg spots in the diffraction pattern from a crystal occur because of constructive interference between X-rays scattered from individual molecules in the crystal. This can only occur at small number of discrete scattering angles. These angles are determined by the crystal lattice. The incident X-ray beam, k_i , and an outgoing scattered X-ray, k_s , are shown in figure 2.4B. The magnitude of the incident and scattered wave-vectors is inversely proportional to the wavelength of the incident and scattered X-ray beams, λ:

$$|k_i| = \frac{2\pi}{\lambda}. \tag{2.1}$$

The end of the incident wave vector, k_i, terminates at the origin of reciprocal space, *O*. The angle of the incident wave vector is the determined by the experimenter, who can

rotate the incident wave vector arbitrarily to explore different parts of reciprocal space. The lengths of these two vectors are inversely proportional to the wavelengths of these X-ray beams. As the X-ray scattering is elastic, no energy is lost during scattering, the lengths of k_i and k_s are equal. The Ewald circle construct assists in understanding when a constructive scattering event will occur. The incident wave vector and outgoing wave vector are both drawn as radii from the center of the Ewald circle. The circumference of the Ewald circle has a finite width, Δk, that reflects the slight spread in wavelengths and divergence present in the incident X-ray beam. The spread in wavelengths used in conventional crystallography experiments is usually very small: the incident beam is essentially monochromatic. A less commonly used type of crystallography experiment, called Laue crystallography, utilizes a much wider spread of incident X-ray wavelengths, and captures many more Bragg spots in a single diffraction image. Typical X-ray beam divergences are less than a few milliradians.

A vector, Q, connecting the incident and scattered wave vectors is shown in figure 2.4B. If Q is equal in length and direction to an inverse lattice vector, G, then constructive scattering will occur. The Ewald circle helps to identify this: if the circumference of the Ewald circle lies upon a reciprocal lattice spot then constructive interference will occur. A beam of scattered X-rays will be directed from the center of the Ewald circle toward the reciprocal lattice point and out of the crystal, and appear on the detector face. Constructive scattering is shown in figure 2.4C. In figure 2.4C, three constructive scattering events occur, producing three Bragg spots. Each spot is numbered by its originating position in the reciprocal lattice. These lattice indices are commonly referred to as *h, k* and *l*. A 2-dimensional version of this numbering scheme is shown in figure 2.4C. In so far as a reciprocal lattice spot is completely within the thickness of the Ewald sphere, the intensity of each Bragg spot will be determined by

the Fourier transform of an individual molecule at the position of the lattice point in reciprocal space. In addition to the three Bragg spots, the incident beam will also appear in the diffraction image. The incident, or direct beam, is labeled as the (0,0,0) reflection, and appears in every diffraction image in the dataset. The direct beam spot is labeled as (0,0) in figure 2.4C. The direct beam is much more intense than the other Bragg spots, and for this reason is heavily attenuated by a small lead beam stop placed directly behind the crystal. The thickness of the beamstop is often chosen so as not to completely block the direct beam, allowing it to be identified in the diffraction image. The direct X-ray beam center is shown in the inset in figure 2.3 in the center of the white shadow of the beam stop.

It is important to remember that the Fourier transform of an individual molecule need not be real: it is almost always complex, having real and imaginary components. Thus, the Fourier transform of a single molecule at a lattice point in reciprocal space (often called the *structure factor*) may be given by the complex number

$$F(h,k,l) = A(h,k,l)\,exp\left(i\phi(h,k,l)\right), \tag{2.2}$$

where $F(h,k,l)$ has an amplitude A, and a phase component, ϕ. The intensity of the Bragg spot associated with this structure factor is given by the structure factor multiplied by its complex conjugate

$$I(h,k,l) = FF^{*} = A^{2}. \tag{2.3}$$

The intensity of the spot, the only quantity that can be practically measured in an experiment, is purely a function of the amplitude of the structure factor, presenting a

complication in interpreting the diffraction pattern of the crystal. The goal of a crystallography experiment is to determine the Fourier transform of an individual molecule, by measuring all of its structure factors, and then to invert the transform to determine the real space structure of the molecule. As the intensity of a Bragg spot is determined solely by the amplitude of its associated structure factor, information on the *phase* of the structure factor is lost. The amplitudes alone are a necessary, but insufficient condition for the solution of the structure of an individual molecule. This loss of information, and the problem it poses for reconstructing the structure of the molecule is called *the phase problem*. For molecules containing less than 100 non-hydrogen atoms, the loss of information associated with the phase problem does not pose an insurmountable problem for the solution of the structure of the molecule. The structure of molecules containing less than 100 atoms may be solved using traditional direct phasing methods [83]. More recently, dual-space recycling direct methods, often called "Shake-and-Bake" methods, have permitted the direct phasing of diffraction patterns from molecules containing well over 200 atoms, and a small number of small protein molecules [83, 84]. However, for larger molecules, and especially for macromolecules, the phase problem, and associated loss of information typically poses too great a challenge for direct methods [83].

The solution to the crystallographic phase problem was the crucial breakthrough in the solution of the structures of protein molecules. Today, several techniques exist to measure the relative phases of the diffraction spots in an X-ray crystallography experiment. For this thesis, no experimental measurement of the phases was required, as closely related structures of the proteins examined in this thesis were already known, permitting phasing by molecular replacement [85].

Figure 2.4D shows the area of reciprocal space that is accessible by an X-ray beam of a given wavelength. By rotating the crystal, and hence the reciprocal lattice about the incident X-ray beam, the Ewald circle can be rotated through reciprocal space, sweeping out a circle with a radius equal to twice the Ewald circle radius. As the length of the incident X-ray wave vector is inversely proportional to incident X-ray wavelength, a shorter X-ray wavelength permits access to more points in reciprocal space, and allows for a more accurate reconstruction of the structure of a single molecule.

It is important to note that to acquire sufficient information to solve the structure of the molecule, the incident wave vector does not have to be swept through 360˚ about the origin as is shown in figure 2.4D. For a 3-dimensional crystal, the diffraction pattern is at least two-fold symmetric. At a maximum, the incident wave vector needs to be rotated by 180˚ to gather sufficient information to reconstruct the structure of a single molecule. However, many crystals display higher symmetry, further reducing the minimum area (or volume) of reciprocal space that must sampled in order to reconstruct the structure of the molecule.

A second important note is that the structure solution process of X-ray crystallography assumes that all of the molecules composing the crystal are identical. This is true at one level: all of the molecules composing the crystal are chemically isomorphous, and have adopted same gross fold. However, fine details of the structure need not, and often are not identical. For instance, a side-chain may appear in two distinct conformations. This can be resolved in the electron density: a residue in the structure of the molecule will appear to have two side chains rather than one. Another type of structural inhomogeneity may not be resolvable given the maximum resolution of the

dataset, and is often modeled by an increase in the *temperature factor* (often also called the B-factor) of the residue. In small molecule crystallography, the temperature factor models the thermal fluctuation in position of the residue. However, in macromolecular crystallography, the temperature factors of residues indicate extremely large positional fluctuations, suggesting conformational inhomogeneity that cannot be resolved into individual positional states.

Finally, the wavelength of X-rays generated by rotating anode and synchrotron sources is approximately 1 Å. For example the copper K-α line produced by a rotating anode X-ray source with a copper target is 1.54 Å, the wavelength of CHESS station F2 when tuned to the selenium edge is 0.9795 Å, and the CHESS station F1 monochromator is tuned to 0.918 Å. These X-ray wavelengths suggest that the highest resolution that should be achievable by X-ray crystallography is approximately 0.5 Å. However, only one dataset out of approximately 54,000 structures reported in the PDB (Protein Data Bank) displays such a high resolution [86]: when using a 0.54 Å synchrotron X-ray beam. The limiting factor in the quality of a protein atomic structure is the quality of the crystal from which it was derived. Typical protein crystallography datasets display maximum resolutions of 1.5 to 4.0 Å. This low maximum resolution is in part due to conformational inhomogeneity inside the protein molecule, but is mostly due to inhomogeneity in the crystal lattice. It is believed that a protein crystal is not composed of a single lattice from edge to edge, but rather contains many sub-domains, that are slightly misaligned from one another. The average degree of misalignment is called the *mosaicity*. The *mosaicity* of the crystal results in a spread of the angular position of each Bragg spot. It is not uncommon to find the angular position of a Bragg spot to be spread across one or more degrees. The mosaicity of the crystal is also enhanced by the divergence of the X-ray beam. The

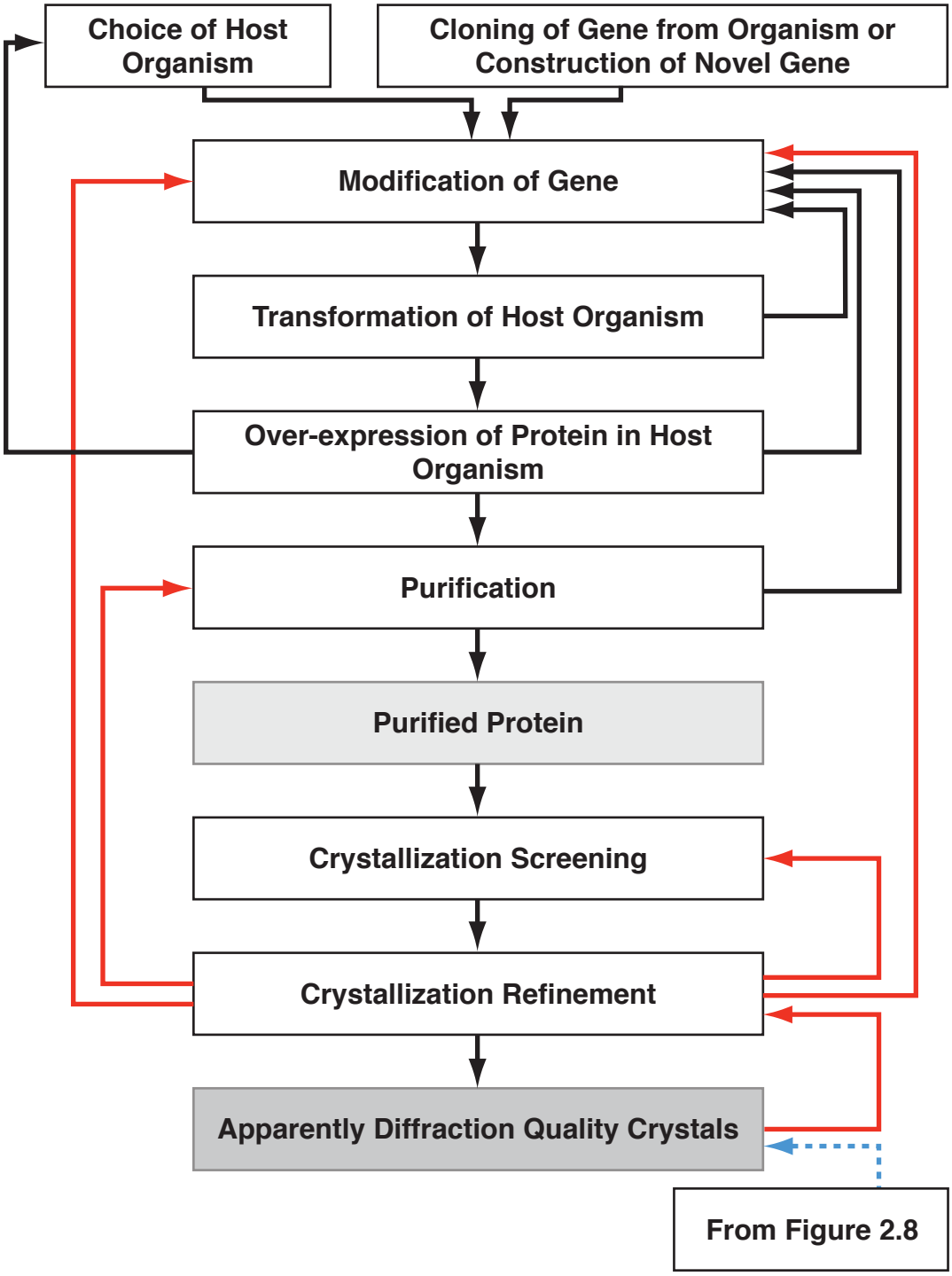


Figure 2.5: Flow chart of protein crystallization procedure. Significant milestones are colored in grey.

crystal mosaicity contributes heavily to the reduction of the maximum possible resolution in the X-ray dataset, and is thus a major determinant of the quality of a model of a protein atomic structure.

2.3 Production of Protein and Preparation of Crystals for X-ray Crystallography

The first two steps of protein crystallography, obtaining sufficient quantities of highly purified protein, and then crystallizing this protein are the two limiting steps in protein crystallography.

The first protein molecules whose structures were solved, sperm whale myoglobin and horse hemoglobin, were chosen because large quantities of the protein were available from natural sources and they readily formed high quality crystals suitable for X-ray diffraction [81]. A flow chart of the steps involved in producing highly purified protein and diffraction quality crystals is shown in figure 2.5. A flow chart detailing the steps towards the solution of the structure of a molecule given diffraction quality crystals is shown in figure 2.8. An important feature of the flow chart shown in figure 2.5 is that the number of arrows pointing forwards is equal to the number pointing backwards. The process of protein expression, purification and expression is often a highly iterative process of trial and error.

Today, the number of proteins that may be produced in large quantities and crystallized is much larger than it was in the 1950s thanks to recombinant DNA technology, modern protein purification technology and a wide selection of commercially available ready-to-use molecular biology kits. Protein crystallization is also considerably easier due a wide body of experience in crystallization and a wide selection of commercially available protein crystallization kits and tools. This, along

with a not inconsiderable amount of hard work, has greatly expanded the scope of crystallization, and has allowed many problems in purification and crystallization that were previously thought intractable to be solved.

Today, unlike in early experiments by Perutz and Kendrew, protein purification rarely begins with proteins derived directly from their parent organism. Today, the gene encoding the protein is usually cloned from the organism and then used to transform microorganisms that can produce much larger quantities of extremely homogeneous protein. This has a number of advantages: firstly, in order to obtain sufficiently large quantities of a protein from its natural source, one is often required to source it from many individual organisms, often reducing the homogeneity of the sample, and its crystallizability.

Secondly, for proteins derived from microorganisms, the number of microorganisms that can be cultured in present day laboratory conditions is extremely small [87]. Some microorganisms do not respond well the culture media used in modern laboratories while others require extremely specialized conditions such as high pressure [88], high ionic strength or high temperature [89] in order to replicate.

Thirdly, recombinant DNA technology allows the rapid manipulation of the gene encoding the protein, allowing the generation of a large number of mutant proteins. This allows the study of many naturally, and non-naturally occurring mutants.

Finally, and perhaps most importantly, cloning of a gene, removes a number of safety and ethical considerations concerning the use of naturally derived proteins. For example, the manipulation of proteins, such as hemoglobin, that have been derived

from human blood carries with it the risk of exposure to a number of blood-borne pathogens such as HIV and hepatitis B and C. For proteins from rare and endangered plants, there is both a collection problem and the ethical dilemma involved with destroying large numbers of individual plants. For human and animal derived proteins, collection of the protein may require a large number of donors. Experiments involving human or live animal subjects often carry with them ethical and safety considerations that should be carefully weighed. For example, the collection of mutant proteins implicated in human disease often carries with it the burden of expectation of treatment. Extraction of proteins from animal species often carries with ethical concerns, especially from endangered species.

In addition to cloning, an important starting point for the production of recombinant protein is the choice of a suitable host microorganism. The most commonly used microorganism is the strain K12 *Escherichia coli*. The K12 strains of *E. coli* have been heavily acclimated to laboratory culture conditions over several decades of laboratory use, and are considered unable to survive outside of laboratory culture conditions, preventing escape from the laboratory [90]. *E. coli* strains are easily transformed with exogenous genetic material using *plasmids*, small circular sections of DNA, usually encoding a small number of genes. The *E. coli* strain can be prompted to produce large quantities of the exogenous protein using a chemical signal that activates transcription from the plasmid. However, prokaryotic cells often lack the protein machinery to properly modify exogenous proteins following translation. In some cases, it is possible to reconstitute these accessory genes into *E. coli*. [91, 92]. However, in other cases it is simpler to use a more complex organism containing post-translational machinery that more closely matches that of the protein's parent organism. Commonly used host

organisms include yeast (*Saccharomyces cerevisiae*) and insect cells from *Drosophila melanogaster* (common fruit flies).

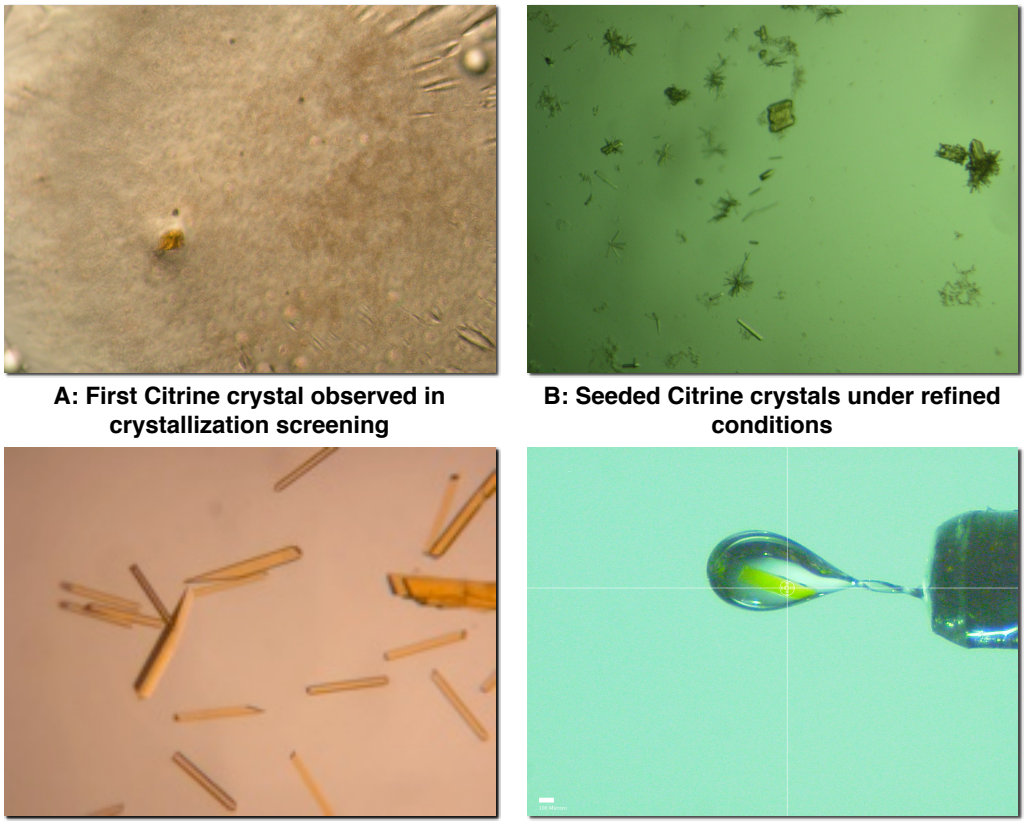
Recent advances in gene synthesis have the potential to render the cloning step redundant. In the last four years, new gene synthesis technology has emerged that allows long genetic sequences to be synthesized artificially with little or no error [93]. Prior to this invention, a long gene would have to be physically transferred between interested parties, usually deposited and dried on a piece of blotting paper. This new invention allows a genetic sequence to be digitally transmitted between parties. The digitally encoded copy of the gene may then be mutated and optimized for the host organism, and sent for synthesis. Companies that offer long gene synthesis are Codon Devices (Cambridge, MA, USA), GeneArt (Regensburg, Germany) and DNA 2.0 (Menlo Park, CA, USA). Recently, microfluidic DNA synthesis was demonstrated, opening of the possibility of gene synthesis on the laboratory bench [94].

Following construction of the gene, the host cell line is transformed with the gene, cultured, and then prompted to over-express the protein encoded by the gene. The cells are then lysed, and the exogenous protein is purified from the cellular debris. It is common to add DNA bases encoding several histidine residues, called an *affinity tag,* to the N or C-terminus of the protein molecule at the gene modification step. The affinity tag is designed to strongly bind to a purification column, permitting the rapid purification of the protein. The affinity tag is often removed by a protease enzyme that attacks the protein sequence between the affinity tag and the start of the desired protein molecule. A typical yield of recombinant protein from several liters of *E. coli* culture is between tens and hundreds of milligrams.

The large quantity of recombinant protein allows the screening of multiple crystallization conditions. Protein crystallization typically requires highly pure protein. Anecdotally, crystallization requires that the protein sample be at least 95% pure. At the time of writing very little ability exists to predict the crystallization conditions for a given protein molecule. For this reason, it is common to attempt to crystallize a protein molecule in a wide variety of conditions, often called a *sparse matrix screen* [95].

The process of refinement of the crystallization conditions for the protein Citrine [1], the topic of this thesis, is shown as a series of photographs in figure 2.6. Figure 2.6A shows the first Citrine crystal observed during a coarse screen of crystallization conditions around those suggested by Griesbeck *et al*. [1]. The crystal shown in figure 2.6A nucleated and was grown in 9% PEG 3350, 50 mM Na acetate, NH_4 acetate, pH 4.5. This crystal was definitively identified as a protein crystal by staining with the dye Izit (Catalog number HR4-710, Hampton Research, Aliso Viejo, CA, USA). The Izit dye binds to the solvent channels present in protein crystals, but not present in salt crystals. Following staining with Izit, protein crystals will retain the dye, while salt crystals that also form under protein crystallization conditions will not.

Following identification of crystallization conditions that produce crystals, it is typical to refine these conditions to produce crystals that appear to be of suitable quality for X-ray diffraction. It is not uncommon at the initial screening and refinement step to revisit the purification step as the crystallization may highlight a need for further purification. Protein crystal *seeds*, produced by tapping the surface of the first Citrine crystal shown in figure 2.6A with a fine needle (Crystal Probe, Catalog Number HR4-217, Hampton Research, Aliso Viejo, CA, USA), were transferred to a fresh



A: First Citrine crystal observed in crystallization screening

B: Seeded Citrine crystals under refined conditions

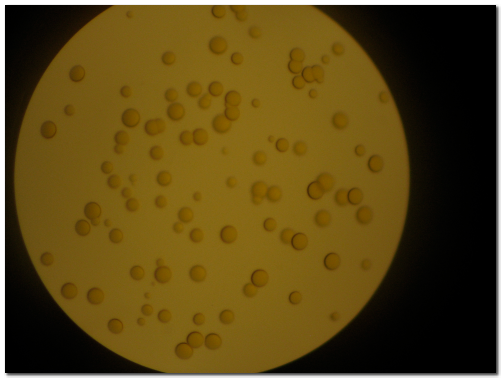
C: Diffraction quality Citrine crystals

D: Citrine crystal at CHESS station F2

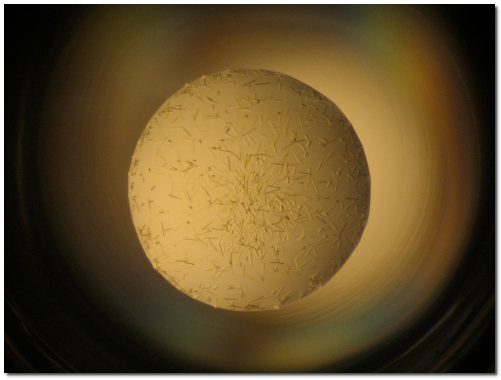
Figure 2.6: Crystals of the yellow fluorescent protein, Citrine, at important stages in the crystallization refinement process.

droplet of protein solution and precipitant. These seeds provided nucleation points for new crystals. This process produced the slightly improved crystals shown in figure 2.6B. This seeding procedure was repeated, and the crystallization conditions refined, eventually producing the high quality crystals shown in figure 2.6C.

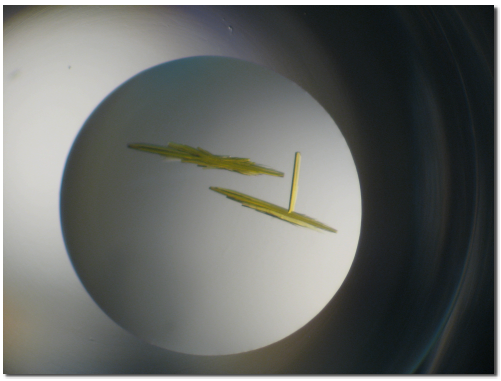
Following the production of apparently diffraction quality crystals, the crystal quality can be tested by X-ray diffraction. One of the high quality crystals shown in figure 2.6C was transferred to a cryo-loop (Hampton Research, Aliso Viejo, CA, USA) and frozen. A frozen crystal mounted on the goniometer at CHESS station F2 for X-ray data collection is shown in figure 2.6D.



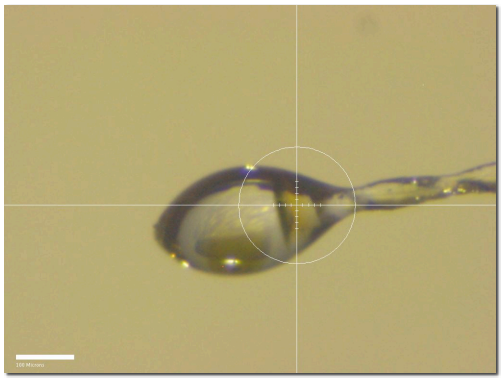
A: Spherulite Morphinone Reductase crystals grown early in screening



B: Shower of rod-like Morphinone Reductase crystals



C: Almost diffraction quality Morphinone reductase crystals

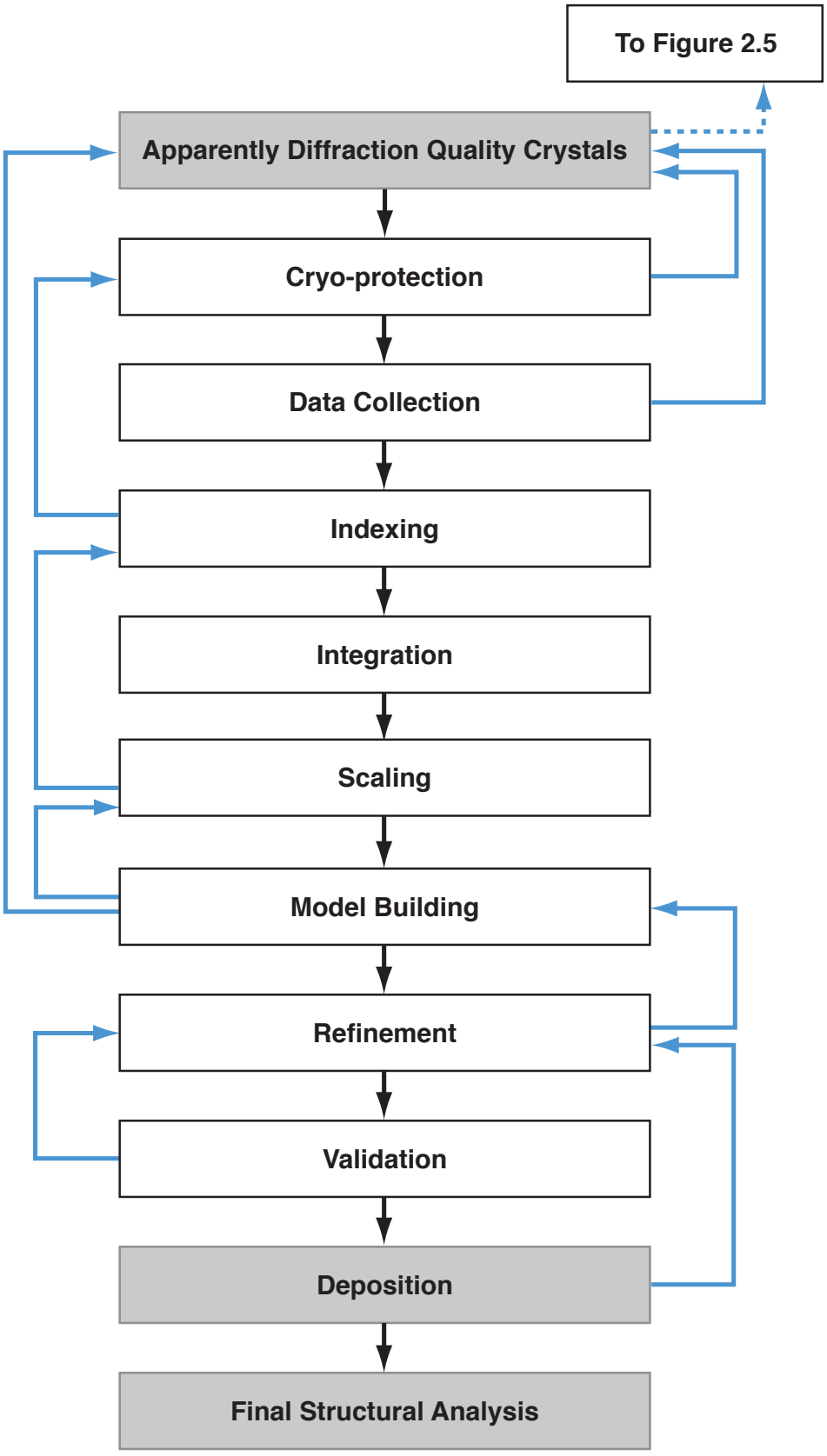


D: Morphinone Reductase crystal soaked in $NADH_4$ at CHESS Station F2

Figure 2.7: Crystals of Morphinone Reductase at important stages in the crystallization refinement process.

Photographs showing the refinement of crystallization conditions for the protein Morphinone reductase are shown in figure 2.7. Initial screening about the crystallization conditions suggested by Pudney *et al.* [38] produced what are believed to be spherical clusters of needle crystals. These spherical clusters are shown in figure 2.7A. Further refinement of the crystallization conditions suggested by Pudney *et al*. [38] was unable to produce usable crystals. This prompted screening around the conditions suggested by Moody *et al*. [96] and Barna *et al*. [97]. The crystals shown in figure 2.7B were produced using the conditions suggested by Moody *et al*. [96]. These conditions were refined, yielding the crystals shown in figure 2.7C. These crystals were soaked with the substrate analog $NADH_4$, mounted in cryo-loops and frozen, and taken to CHESS station F2 for a diffraction data collection.

Figure 2.8: Flow chart of protein crystal data collection and reduction procedure.


To Figure 2.5
Apparently Diffraction Quality Crystals
Cryo-protection
Data Collection
Indexing
Integration
Scaling
Model Building
Refinement
Validation
Deposition
Final Structural Analysis

2.4 X-ray Diffraction Data Collection and Reduction

X-ray diffraction data is collected from a protein crystal at a synchrotron or home X-ray source. A *dataset* of diffraction images is taken from the crystal as the crystal is rotated around the X-ray beam, sweeping the incident beam through the reciprocal lattice space of the crystal. The exposure time for each image varies with source, and can last from < 1 second at a high flux synchrotron beam-line such as F1 at CHESS, to several minutes at a rotating anode source. A flow chart outlining the process of diffraction data collection and reduction is shown in figure 2.8. Data collection is typically a much more straightforward process than that used to arrive at diffraction quality crystals.

One typically chooses crystals for use in diffraction data collection based upon their external appearance. The external appearance is often, but certainly not always, an indicator of diffraction quality. For this reason, it is prudent to *screen* a large selection of crystals, and check for diffraction quality. Screening typically only requires a few diffraction images to establish the diffraction quality of the crystal. Initial screening can often be done by manual inspection of a single diffraction image from the crystal. The author's eye is typically capable of discerning diffraction spots with a signal to noise of approximately 3.0, and so a reasonable estimate of the diffraction limit of the crystal can be obtained manually. If the screening indicates a suitable crystal, then one can collect a full dataset from the crystal.

For each image in the diffraction dataset, the crystal is rotated, or *oscillated*, through typically 1˚. Oscillating the crystal through a small angular range greatly increases the number of reciprocal lattice points that pass through the thickness of the Ewald sphere, and greatly increases the number of diffraction spots captured in the diffraction image.

For the next image, the crystal is rotated through the next 1˚ increment. The number of diffraction images required to solve the structure of the molecule depends upon the symmetry of the crystal. Some crystals have extremely high symmetry, while others have lower symmetry. The maximum number of frames required to solve a structure is 180˚.

Following, or preferably during collection of the full dataset, the first few images of the dataset are *indexed*, to determine the space group, lattice (or unit cell) parameters, and mosaicity of the crystal. An algorithm searches the images for diffraction spots, and calculates the best space group and unit cell parameter match to the position of these spots [98, 99]. Performing this step early in the data collection procedure is preferable, as crystals that display diffraction that is acceptable to the human eye, may not index properly. The most accurate estimate of the crystal mosaicity may be obtained by indexing a small range of frames separated by 90˚, for instance frames 1 to 3 and 88 to 90. For this reason, it is preferable to wait until at least 90˚ of diffraction data has been collected from the crystal before making a final estimate of the mosaicity. The initial indexing algorithm used in this thesis was the Data Processing Suite (DPS) algorithm by Rossmann and van Beek [99]. Indexing parameters were refined using the MOSFLM code [100].

When the full dataset has been collected, the intensities of the diffraction spots in the series of diffraction images are calculated. This process is called *integration*. For the work described in this thesis integration was performed with the MOSFLM code by Leslie [100] that is part of the CCP4 suite (Collaborative Computational Project Number 4, Daresbury Laboratory, Warrington, United Kingdom) [101].

The intensity of the spots is both a function of the structure of the molecule in the crystal, the intensity of the X-ray source, and the volume of the crystal illuminated by X-rays. Following integration, the data reduction process removes the effects of variations in the intensity of the X-ray source and intensity of diffraction due to changes in the volume of the crystal that is illuminated as the crystal is rotated. This procedure is called *scaling* [102]. Scaling produces a set of intensities that are independent of the X-ray intensity and crystal volume illumination. Following scaling, the intensities are *truncated*, to produce amplitudes [103].

The procedure thus far described yields only a set of structure factor amplitudes. If a model structure that is closely related to the molecule of interest's structure is known, then the amplitudes alone may be used to solve the structure of the molecule of interest. A process called *Molecular Replacement* (MR) is first used to establish the orientation of the model of the unit cell of the molecule of interest [104]. Two currently popular programs for molecular replacement are MOLREP [85] and AMORE [105, 106]. MOLREP was used for the work presented here. MR finds the orientation of the model in the unit cell of the crystal. Following the molecular replacement step, the orientation of the model is sometimes further refined by a process called *rigid body refinement*. Rigid body refinement was performed with the REFMAC5 code [107], part of the CCP4 suite of programs.

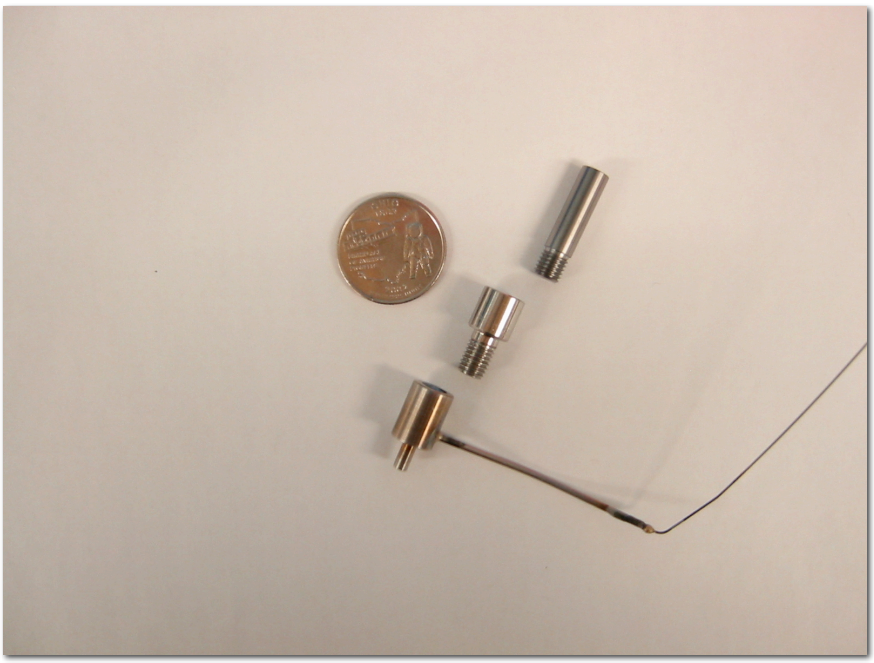
Following the rigid body refinement step, the model is then *refined*. The refinement procedure iteratively adjusts the model of the molecule of interest by rotating bonds, translating atoms, and adding atoms to the model. Modern refinement algorithms minimize the difference between the observed diffraction amplitudes and calculated diffraction amplitudes predicted for the model, and minimize the divergence of the

model's stereochemical parameters such as bond lengths and bond angles from idealized library values. The refinement algorithm used in the work presented in this thesis was REFMAC5 [107], part of the CCP4 suite of programs. Solvent addition was performed with the ARP/wARP (Automated Refinement Procedure/water Automated Refinement Procedure) algorithm [108-113].

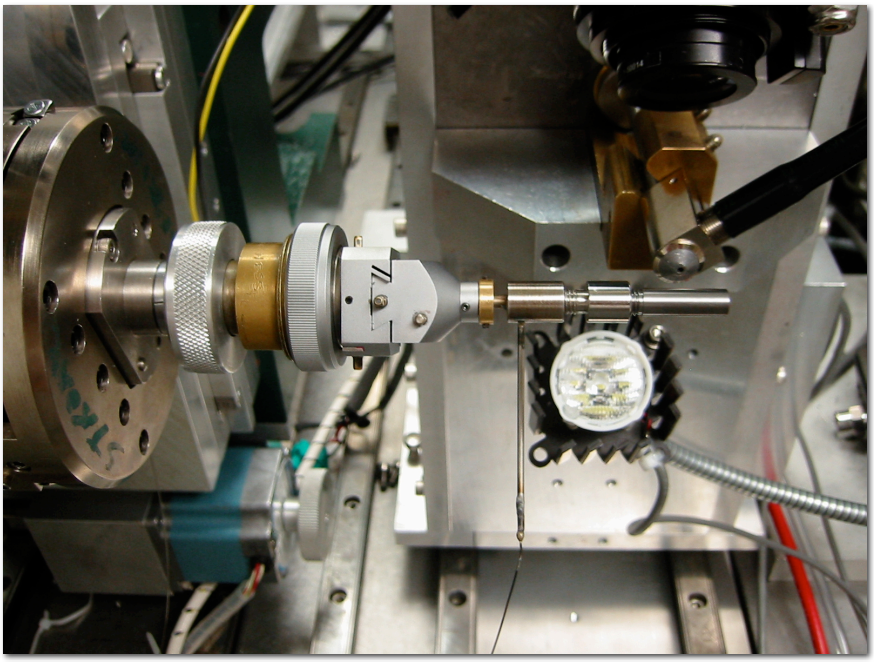
Finally, the model is *validated* by checking its stereochemistry against an independent library of structural constraints, and by examining its quality indicators [114]. At this step, it is often wise to deposit the structure into the Protein Data Bank (PDB): a publicly accessible database of protein structures. Deposition of protein atomic structures discussed in a journal article is typically a precondition of publication. The structural quality necessary for inclusion into the Protein Data Bank often requires revision of a structure that the experimenter may deem acceptable. To avoid conflict between published structural analysis, and the structures available to the public in the PDB, it is recommended that the experimenter deposit their structures into the PDB prior to making a final structural analysis that will be included in a publication.

2.5 Solving the Structure of Proteins Molecules Under High Pressures

Two high-pressure adaptations to X-ray crystallography were used in this thesis to solve the structures of protein molecules under high pressures. The fist technique uses hydrostatic pressure, in which the liquid solvent surrounding a protein crystal is pressurized to apply pressure to protein in a crystal. The second technique: high-pressure cryo-cooling, applies pressure to a protein crystal with pressurized helium gas.



A: Disassembled beryllium high-pressure X-ray cell



B: Beryllium cell mounted on CHESS station F1 goniometer

Figure 2.9: The high-pressure beryllium X-ray cell. A: The beryllium cell in its disassembled state next to a US quarter dollar coin (diameter ≈ 2.4 cm). B: The beryllium high-pressure cell at CHESS station F1.

2.5.1 High-pressure Beryllium X-ray Cell

The beryllium cell for high-pressure X-ray crystallography was developed by Kundrot and Richards [23] and used to solve the structure of hen egg white lysozyme under a pressure of 100 MPa [24, 69]. The original beryllium cell design by Kundrot and Richards was limited to a maximum pressure of 100 MPa. A modification to the seal of the beryllium cell by Urayama increased the maximum pressure of the cell to 200 MPa [70].

The primary advantage of the beryllium cell is that it operates at room temperature. Most high-pressure effects on protein function are initially noticed at room temperature. For catalytic proteins, a fluid, non-frozen solvent is required to perform their function. The beryllium cell permits solution of the structure under these conditions.

Unfortunately, the beryllium cell has a number of disadvantages. Firstly, beryllium metal displays strong X-ray powder diffraction rings starting at 2 Å and extending to higher resolution [23]. These powder diffraction rings obscure diffraction spots from a protein crystal in the cell at resolutions higher than 2 Å [70]. Secondly the cell requires complex assembly at an X-ray beam-line, necessitating that several hours of X-ray beam-time be used for setup. Finally, it is challenging to locate small protein crystals, smaller than 500 μm in largest dimension, in the cell.

A photograph of the disassembled high-pressure beryllium X-ray cell is shown in figure 2.9A. A photograph of the beryllium cell at CHESS station F1 is shown in figure 2.9B.

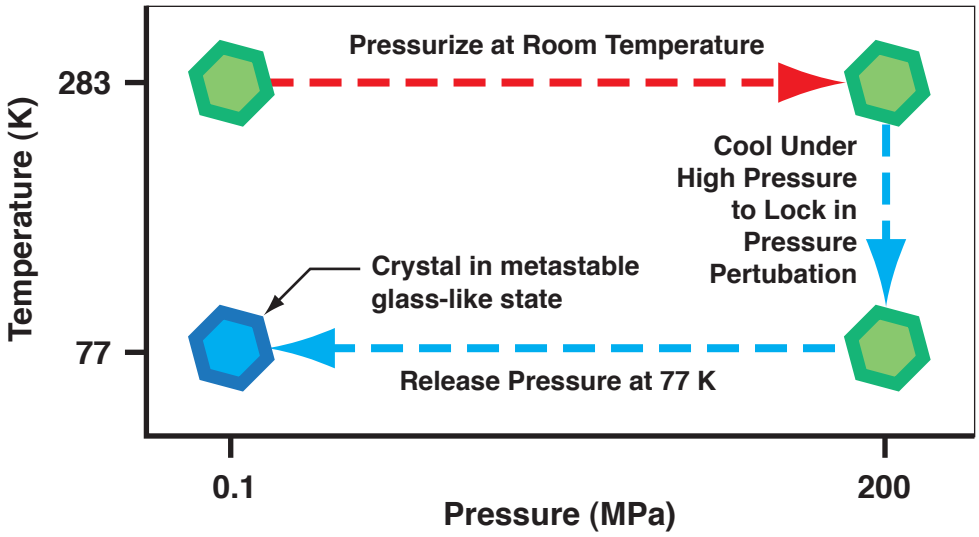


Figure 2.10: Pressure-temperature phase diagram for high-pressure cryo-cooling procedure. Adapted from Kim *et al*. [3].

2.5.2 High-pressure Cryocooling

An alternative to the beryllium cell is the high-pressure cryo-cooling technique developed by Kim *et al*. [3, 12, 115-117]. A schematic of the high-pressure cryo-cooling procedure is shown in figure 2.10. A photograph of the safety enclosure used to surround the high-pressure cryo-cooling apparatus is shown in figure 2.11. A photograph of the assembled high-pressure cryo-cooling apparatus is shown in figure 2.12. The assembled high-pressure cryo-cooling apparatus is enclosed in a room to provide additional shielding to the operator in the event of an unexpected pressure-release. The apparatus is operated remotely from an adjacent room.

A protein crystal is pressurized with helium gas and is then cooled to 77 K, locking in collective pressure-induced structural changes [3, 4]. Following pressure release, the protein molecules composing the crystal will retain many of the collective changes of the pressurized state, on the condition that the crystal's temperature remains well below its glass transition temperature [3, 12, 116].

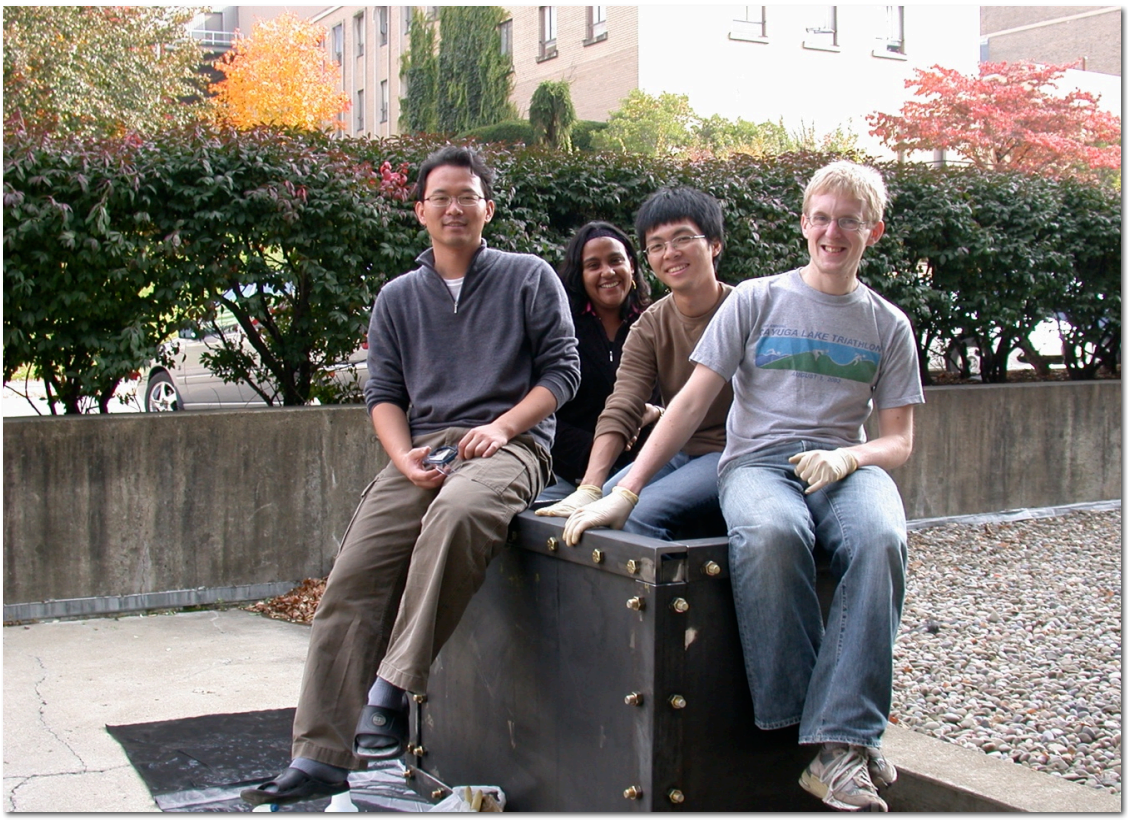


Figure 2.11: Photograph of high pressure cryo-cooling apparatus safety enclosure prior to painting. From the left: Chae Un Kim, Marianne Hromalik, Yi-Fan Chen, Buz Barstow. Photograph by Nozomi Ando.

2.6 Locating Objects with Sub-wavelength Precision

Given that the expected perturbations to protein structure induced by pressure are typically on the order of 0.1 to 1.0 Å it is important to address how precisely these structural deformations may be detected. What is the limit of detectability for small structural motions in proteins? How does one reliably and verifiably visualize structural changes to a protein molecule on the order of 0.1 Å when one only has imaging resolution, through X-Ray crystallography of 1.0 to 2.0 Angstroms? To address this, an optical analogy is in useful.

Figure 2.12: Photograph of high-pressure cryo-cooling apparatus after full assembly.

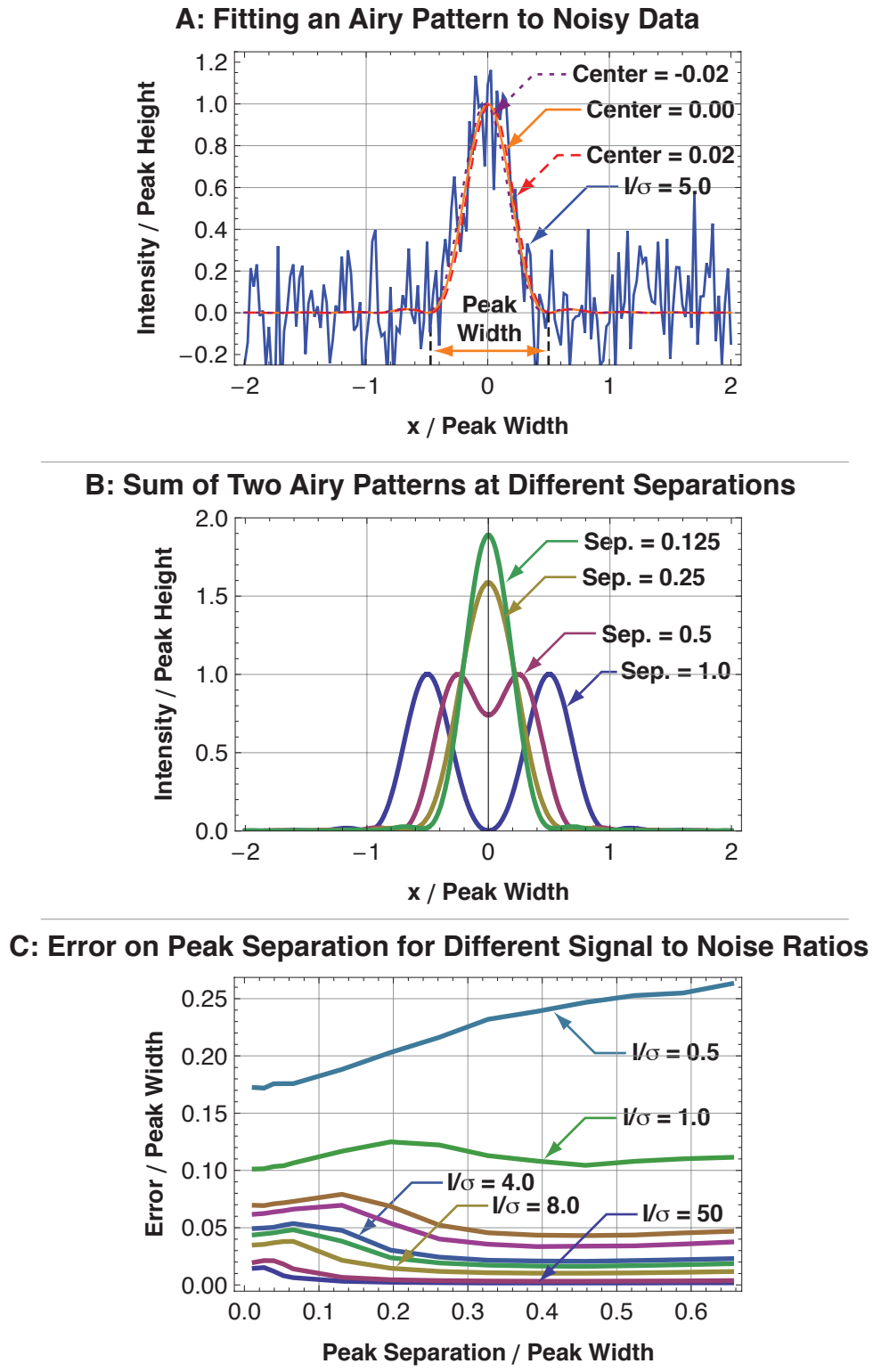


Figure 2.13: Precision possible in determining the center of two closely separated Airy patterns in the presence of random noise. See section 2.6.

In an optical microscopy system, a point source is imaged as an Airy pattern. The width of this Airy pattern is determined by the numerical aperture of the imaging system, and by the wavelength of light used to detect it. It is commonly accepted that the smallest feature that can be resolved by an imaging system, given a perfect set of optics, has a size of approximately half of the wavelength of light used to image it [118]. However, if one carefully measures the intensity distribution of an object imaged by an optical system, one can numerically fit an Airy pattern to this, and determine the center of the Airy pattern. Given sufficient signal to noise, one can determine the center of the Airy pattern and the position of the object with sub-wavelength precision.

The effect of noise on the precision of the measurement of the center of an Airy pattern was explored using a Monte Carlo simulation. The results of this simulation are shown in figure 2.13. Figure 2.13A shows an Airy pattern with simulated Gaussian noise. The standard deviation of the noise is $1/5^{th}$ of the height of the Airy pattern: giving a signal to noise ratio (I/σ) of 5.0. Three Airy patterns have been overlaid on the simulated noisy Airy pattern, each with a slightly different center: one centered at $x = 0.0$, the second at $x = -0.02$ and the third at $x = 0.02$. To the eye, it is difficult to determine which of these patterns bests fits the simulated noisy data. However, with a least squares fitting algorithm, it is easy to distinguish, even though the difference in position of these patterns is very much smaller than the width of the Airy pattern (the distance from first minimum to first minimum is 1.0 in the unit system used in this simulation).

Figure 2.13B shows the several double Airy patterns, formed by the sum of two Airy patterns, with decreasing separations between their peaks. As the individual Airy patterns move closer together it appears increasingly difficult to separate the two peaks, at least to the eye, even without noise. However, given knowledge of the number of individual Airy patterns forming the summed pattern, a numerical fitting algorithm can accurately separate the two peaks.

A library of ten thousand simulated noisy double Airy patterns was generated with random Gaussian noise with a specified signal to noise level and separation. Each simulated pattern in this library was fitted to a clean double Airy pattern, and the separation of the peaks was calculated. The difference between this separation, and the actual separation was calculated. The standard deviation of these differences was calculated, and used to represent the error on the calculation of the separation. The results of this simulation are shown in figure 2.13C.

The results shown in figure 2.13C indicate that with signal to noise ratios greater than 1.0, the separation of the peaks can accurately be resolved to distances much lower than the peak width. For a signal to noise of 1.0, the peak separation can be determined for peak separations greater than ≈ 0.25 times the peak width. The results of this simulation indicate, that at least in principle, given knowledge of how many scattering centers one is viewing, one can accurately locate them relative to one another with a precision much smaller than the viewing wavelength. This sub-wavelength scale location of objects has been demonstrated by several "super-resolution" optical techniques [119-130] that are able to locate objects imaged with an optical microscopy system with nanometer resolution.

2.7 Macromolecular Structural Refinement

The analogy with optical systems presented in section 2.6 suggests that although the length scale associated with the diffraction resolution of protein crystals is very much longer than the expected pressure-induced structural deformations in the molecule, it may be possible to reliably detect these displacements. In order to understand the ability of X-ray crystallography to discern small structural changes in protein structures, especially given data of only limited resolution; an introduction to the process of structural refinement is helpful. This short introduction follows the convention of Prince and Boggs [131].

The purpose of refinement is to improve the quality of an atomic model of a molecule by adjusting the model parameters so as to minimize the difference between the observed X-ray structure factors and the structure factors calculated from the model. In addition, refinement often draws upon libraries of bond lengths and angles in order to improve the model. A refinement procedure that attempts to constrain bond lengths and angles to library values is known as a *restrained refinement*, while one that draws only upon diffraction data is known as an *unrestrained refinement*.

Historically, protein structure refinement has proceeded by the minimization of a least squares residual function. This approach is highly suitable for small molecule crystallography, where the number of diffraction spots is small, and the maximum resolution of the X-ray diffraction data is high. However, least squares refinement is less successful for protein diffraction datasets that contain many diffraction peaks yet display a low maximum resolution. A more successful residual, the log-likelihood residual, is now implemented in several crystallographic refinement packages, including REFMAC5 [107], used in this thesis. However, much of the discussion of

precision in macromolecular X-ray crystallography is rooted in least squares refinement, and so a discussion is appropriate.

The goal of refinement is to find a vector of model parameters, **x**, containing the atomic coordinates, B-factors and occupancies of the atoms composing the protein molecule. The number of model parameters ranges from 5 per atom in the case of an isotropic B-factor model, to 13 in the case of an anisotropic B-factor model. As a protein molecule is often composed of thousands of atoms, the total number of model parameters, p, may be in excess of 10,000.

The calculated model parameters, **x**, should produce a set of calculated diffraction amplitudes and phases, $\mathbf{M}(\mathbf{x})$, that most closely matches the observed diffraction amplitudes and possibly phases, **y**. Each observed diffraction data point is associated with a weight, w_i, that is inversely proportional to its likely precision. The most commonly used residual in least squares refinement is the L_2 norm [132]:

$$L_2(\mathbf{x}) = \sum_{i=1}^{n} w_i \left[y_i - M_i(\mathbf{x}) \right]^2 . \tag{2.4}$$

The best set of model parameters, $\hat{\mathbf{x}}$, that most accurately reflect the real molecule, will minimize L_2 with respect to each model parameter: x_j. These partial derivatives form a set of equations often called the *normal equations*:

$$\frac{\partial L_2(\mathbf{x})}{\partial x_j} = -2 \sum_{i=1}^{n} w_i \left[y_i - M_i(\mathbf{x}) \right] \frac{\partial M_i(\mathbf{x})}{\partial x_j} = 0 . \tag{2.5}$$

In the vicinity of the minimum, where $\partial L_2 / \partial x_j = 0$, the calculated structure factors, $M_i(\mathbf{x})$, may be approximated as

$$M_i(\mathbf{x}) \approx b_i + \sum_{j=1}^{p} A_{ij} x_j \,, \tag{2.6}$$

where the A_{ij} are the elements of a matrix **A** of rank $n \times p$. n is the total number of observations and p is the total number of parameters in the model. The b_i are the elements of a vector **b** of length n. In matrix form

$$\mathbf{M}(\mathbf{x}) \approx \mathbf{b} + \mathbf{A}\mathbf{x}. \tag{2.7}$$

In matrix form, the residual, L_2, may be written as

$$L_2 = \left[(\mathbf{y} - \mathbf{b}) - \mathbf{A}\mathbf{x}\right]^{\mathrm{T}} \mathbf{W} \left[(\mathbf{y} - \mathbf{b}) - \mathbf{A}\mathbf{x}\right], \tag{2.8}$$

Where **W** is a matrix whose diagonal elements are equal to the set of weights w_i

$$W_{ij} = \begin{cases} w_i & i = j \\ 0 & i \neq j \end{cases} \,. \tag{2.9}$$

In this notation, the normal equations may be written as

$$\mathbf{A}^{\mathbf{T}} \mathbf{W} \mathbf{A} \mathbf{x} = \mathbf{A}^{\mathbf{T}} \mathbf{W} (\mathbf{y} - \mathbf{b}). \tag{2.10}$$

In the vicinity of the minimum where $\partial L_2 / \partial x_j = 0$,

$$M_i(x) \approx M_i(\mathbf{x_c}) + \sum_{j=1}^{p} \left. \frac{\partial M_i}{\partial x_j} \right|_{\mathbf{x}=\mathbf{x_c}} (\mathbf{x} - \mathbf{x_c})_j \,. \tag{2.11}$$

Thus,

$$A_{ij} = \left. \frac{\partial M_i(x)}{\partial x_j} \right|_{\mathbf{x}=\mathbf{x_c}} \,. \tag{2.12}$$

The solution to this normal equations, the set of model parameters, $\hat{\mathbf{x}}$, that minimizes the residual L_2, is

$$\hat{\mathbf{x}} = \left(\mathbf{A}^{\mathbf{T}}\mathbf{W}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathbf{T}}\mathbf{W}(\mathbf{y} - \mathbf{b}) \,. \tag{2.13}$$

In the case of the solution of the structure of protein molecule, in which there are many thousands of model parameters, *p*, and many tens of thousands of observations, *n*, the computation of the matrix $\mathbf{H} = \left(\mathbf{A}^{\mathbf{T}}\mathbf{W}\mathbf{A}\right)^{-1} \mathbf{A}^{\mathbf{T}}\mathbf{W}$, of size $p \times n$ **,** and of the inverse of $\mathbf{A}^{\mathbf{T}}\mathbf{W}\mathbf{A}$, of size $p \times p$ **,** are too computationally costly to perform. Inversion of a square ($p \times p$ **)** matrix typically requires a number of computational operations proportional to p^3. The best-fit model parameters, $\hat{\mathbf{x}}$, are found by iterative numerical procedures that minimize the L_2 residual while avoiding explicit computation of **H** and its inverse, usually requiring a number of operations proportional to p^2 or np^2 . As *p*, the number of observations is typically larger than *n*, sometimes by an order of magnitude, these operations show considerable time savings [131].

In order to add constraints to the refinement process, an additional term is added to the least squares residual L_2 . This additional term is either an energy-based term, that

adds an energetic penalty to non-stereochemically correct positions of atoms, or a structure dictionary based penalty term [132].

Cruickshank [133] demonstrated that the coordinate uncertainty on a restrained bond length, *l*, could be estimated by summing the inverses of the uncertainties on the length calculated by diffraction data alone and the uncertainty of the library length of the bond:

$$\frac{1}{\sigma^2_{\text{res}}(l)} = \frac{1}{\sigma^2_{\text{diff}}(l)} + \frac{1}{\sigma^2_{\text{geom}}(l)}. \tag{2.14}$$

In the case where the diffraction component of the uncertainty on the length *l* is large, the uncertainty on restrained length will asymptote to the uncertainty on the library value [133].

More modern refinement algorithms, including REFMAC5 [107], use the maximum likelihood method. The maximum likelihood method is often highly successful with low resolution X-ray diffraction data where the number of observations is only slightly higher than, or often less than the number of model parameters. The maximum likelihood method attempts to find the set of model parameters with the maximum probability or likelihood given a corresponding set of observations. The maximum-likelihood method functions by calculating the joint probability, $P(\mathbf{F_{obs}};\mathbf{F_{calc}})$, of all calculated structure factors given the corresponding observed structure factors.

$$P(\mathbf{F_{obs}};\mathbf{F_{calc}}) = \prod_{i=1}^{n} P\left(F_{\text{obs},i};F_{\text{calc},i}\right). \tag{2.15}$$

As probability is always equal to or less than 1.0, maximizing the probability is equivalent to minimizing its negative logarithm;

$$-\log\left(P\left(\mathbf{F_{obs}};\mathbf{F_{calc}}\right)\right) = -\sum_{i=1}^{n}\log\left(P\left(F_{\mathrm{obs},i};F_{\mathrm{calc},i}\right)\right). \tag{2.16}$$

2.8 Estimation of Errors on Protein Atomic Models

2.8.1 Model Parameter Uncertainty Estimate by Least Squares Matrix Inversion

The most complete picture of the errors in a protein atomic model offers an estimation of the uncertainty on each model parameter. However, this error estimation method requires considerable computation, and thus, even at the time of writing is not regularly performed.

At the end of the refinement the model parameters, **x**, have converged upon a final value, $\mathbf{x_c}$. The normal equations can be applied to the converged model to determine an estimate of the error on each parameter. The true value of each model parameter, $\hat{x}_i$, will vary from the modeled value by a small amount, e_i :

$$\hat{\mathbf{x}} = \mathbf{x_c} + \mathbf{e}. \tag{2.17}$$

The normal equations (equation 2.10) can be used to assess the value of the model error vector **e**. At the end of the refinement procedure, the normal equations will be approximately satisfied:

$$\mathbf{A^T W A}\hat{\mathbf{x}} \approx \mathbf{A^T W}\left(\mathbf{y}-\mathbf{b}\right). \tag{2.18}$$

Assuming that $\hat{\mathbf{x}}$ contains a small error **e**;

$$\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\left(\mathbf{x}_{\mathbf{c}}+\mathbf{e}\right)=\mathbf{A}^{\mathrm{T}}\mathbf{W}\left(\mathbf{y}-\mathbf{b}\right), \tag{2.19}$$

$$\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\left(\mathbf{x}_{\mathbf{c}}+\mathbf{e}\right)=\mathbf{A}^{\mathrm{T}}\mathbf{W}\left(\mathbf{y}-\mathbf{M}\left(\mathbf{x}_{\mathbf{c}}\right)+\mathbf{A}\mathbf{x}_{\mathbf{c}}\right). \tag{2.20}$$

Thus, the error, **e**, on the refined model parameters, $\mathbf{x}_{\mathbf{c}}$, may be estimated as

$$\mathbf{e}=\left(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\left(\mathbf{y}-\mathbf{M}\left(\mathbf{x}_{\mathbf{c}}\right)\right). \tag{2.21}$$

By explicitly inverting the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$ it is possible to find estimates for the uncertainty on all parameters in the model. However, inverting such a large matrix, of size $p \times p$ is computationally intensive. At the time of writing, crystallographic refinement programs do not regularly perform this operation.

Tickle *et al*. [134] performed full-matrix inversion calculations to estimate the coordinate uncertainty in two eye-lens proteins: γB-crystallin, refined at 1.49 Å and βB2-crystallin, refined at 2.1 Å. The estimated coordinate uncertainties for the main chain atoms ranged from 0.05 to 0.27 Å for γB-crystallin and from 0.05 to 0.35 Å for βB2-crystallin.

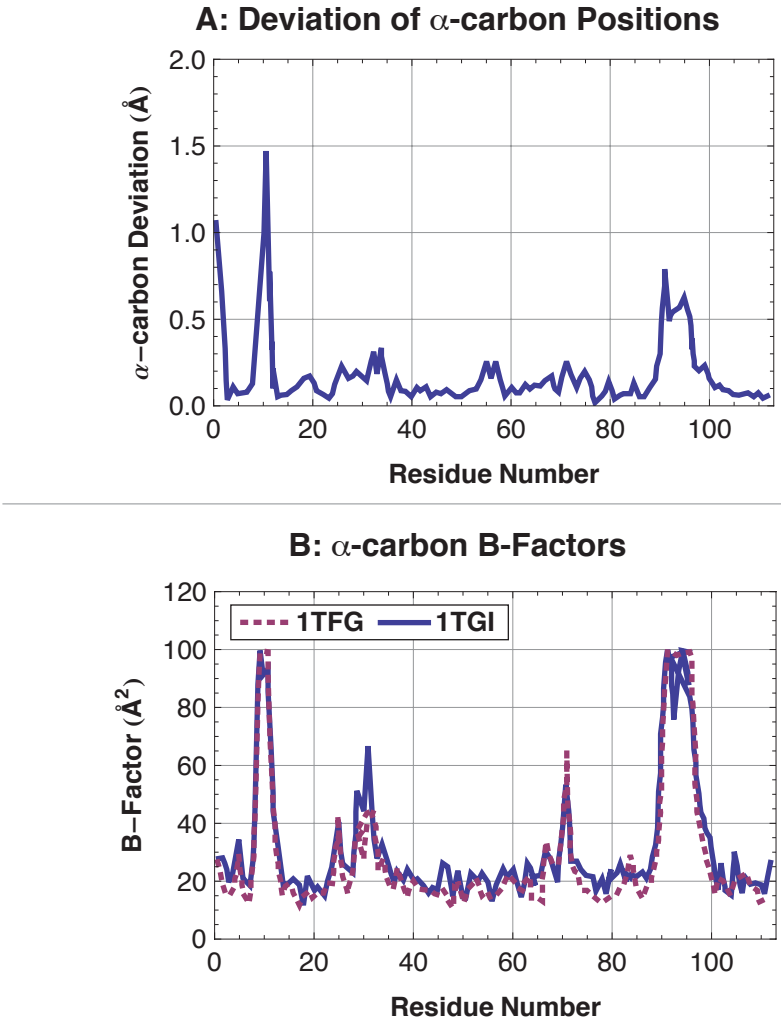


Figure 2.14: A: Plot of distance deviations between positions of α-carbon atoms of ostensibly identical TGF-β2 structures 1TFG and 1TGI. B: Plot of B-factors of α-carbon atoms of 1TFG (dotted line) and 1TGI (solid line). Note that the B-factors correspond strongly with the positional uncertainties of the α-carbon atoms shown above. Adapted from Daopin *et al*. [9].

2.8.2 Approximate Formulas for Coordinate Uncertainty

To address the computational difficulty of inverting the matrix $\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}$, several attempts have been made to produce approximate model parameter error formulas. Cruickshank tested these formulas against full matrix inversion estimates of errors on protein atomic models, and compared them with an experimental study of error in protein structures by Daopin *et al*. [9].

Daopin *et al*. [9] addressed the issue of accuracy and reproducibility in protein structures solved by X-ray crystallography by solving and comparing the structures of two ostensibly identical Transforming Growth Factor-β2 (TGF-β2) molecules expressed by Chinese Hamster Ovary (CHO) cells and *E. coli*. The molecule expressed in CHO cells formed crystals that diffracted to 1.8 Å, while the molecule expressed in *E. coli* formed crystals that diffracted to 1.95 Å. Daopin *et al*. [9] found that the structures of these two ostensibly identical molecules matched extremely well, showing deviations for the most part only at the sub –angstrom level. The deviation between α-carbon atom positions for most of the protein was at the 0.1 Å, with a smaller fraction at the 0.25 Å level, while some apparently disordered regions showed deviations at the 0.5 to 1.0 Å level. The deviation between α-carbon atom positions for each of the residues in the two structures of TGF-β2 is shown in figure 2.14A.

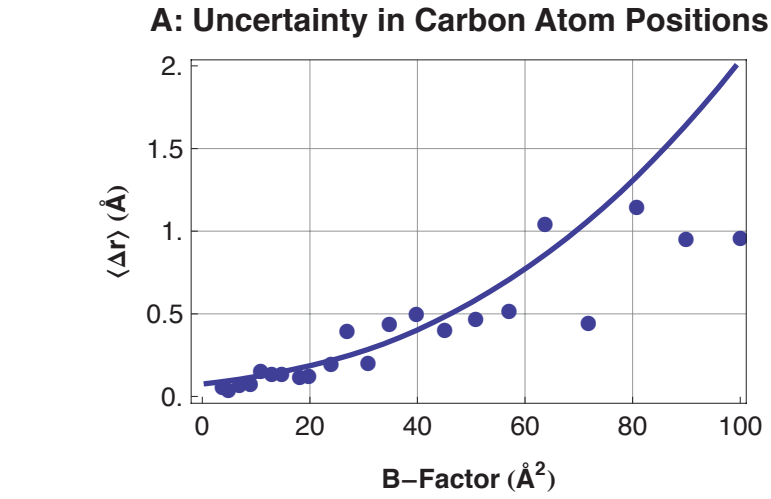
Daopin *et al*. [9] found that the magnitude of the deviation between atoms in the two ostensibly identical structures correlated with the B-factor of the atom. The B-factor is defined as

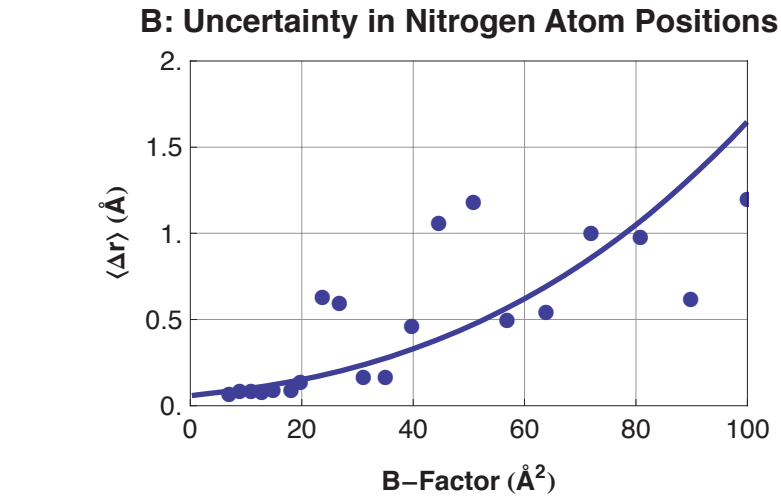
$$B = 8\pi\left\langle u^2 \right\rangle \tag{2.22}$$

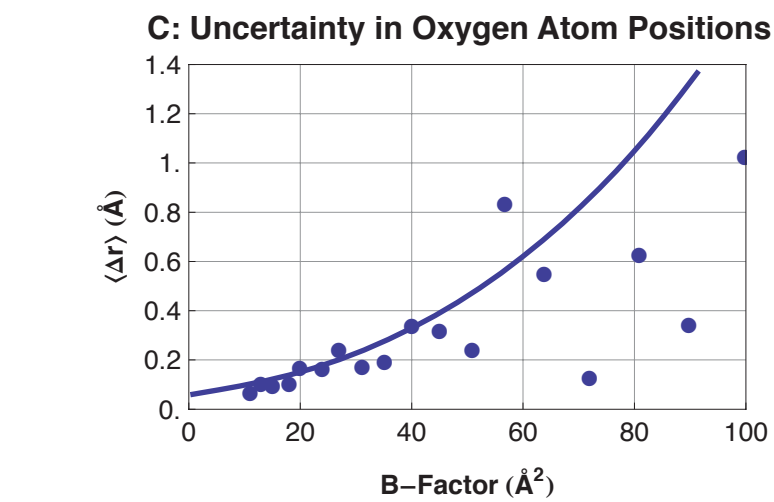
For a B-factor of 80 Å^2, the root mean square (rms) amplitude of the motion $\sqrt{\langle u^2 \rangle}$ of the atom is 1.01 Å, for a B-factor of 40 Å^2, the rms amplitude is 0.71 Å and for 20 Å^2, the rms amplitude is 0.5 Å. These distances are comparable to the lengths of bonds found in protein molecules. The length of a C-N bond is typically ≈ 1.32 to 1.46 Å [135], a C-C bond is typically ≈ 1.44 Å [135] and a C-O bond is typically ≈ 1.24 Å [136]. Additionally, the scattering power of an atom reduces exponentially with increasing B-factor ($\exp\left(-2B\left(\sin(\theta)/\lambda\right)^2\right)$). Thus, it is reasonable to expect that the precision with which an atom may be located reduces with increasing B-factor.

A plot of the B-factors of the α-carbon atoms of the two structures of TGF-β2 are shown in figure 2.14B. Cruickshank provided a qualitative explanation for this dependence of positional uncertainty upon B-factor. The Debye-Waller B-factor expresses the root mean square amplitude of the displacement in the position of an atom. In small molecule crystallography, the B-factor is often small, and can be attributed solely to thermal motion of the atom. However, in protein molecules, the B-factor may be large, and is sometimes attributed to conformational heterogeneity in the protein sample. For instance, in a protein molecule, it may be possible for an atom or an entire residue to assume two or more distinct conformations. Due to the resolution limit of the X-ray diffraction from a crystal of this protein it may not be easy to separate these two conformations, and this conformational heterogeneity may be modeled as a large B-factor.

Figure 2.15: Positional uncertainty of C, N and O atoms in TGF-β2 structures versus B-factors. Trend lines were computed from Cruickshank's positional error estimate formula (equation 2.23). Adapted from Daopin *et al*. [9].


A: Uncertainty in Carbon Atom Positions
2.
1.5
1.
0.5
0.
⟨Δr⟩ (Å)
0
20
40
60
80
100
B−Factor (Å²)


B: Uncertainty in Nitrogen Atom Positions
2.
1.5
1.
0.5
0.
⟨Δr⟩ (Å)
0
20
40
60
80
100
B−Factor (Å²)


C: Uncertainty in Oxygen Atom Positions
1.4
1.2
1.
0.8
0.6
0.4
0.2
0.
⟨Δr⟩ (Å)
0
20
40
60
80
100
B−Factor (Å²)

Cruickshank's first positional precision formula [133, 137] related the positional error of an atom to its B-factor through a Fourier map approach. Cruickshank [133] described the formula qualitatively as

$$\sigma(x) = \sigma(\text{slope})/(\text{atomic peak curvature}). \tag{2.23}$$

Atoms that have a low positional spread will have a tightly constrained electron density with a high curvature, giving a low positional error. Conversely, atoms that have a wide positional spread will have a low electron density peak curvature and a high error in their position.

The $\sigma(\text{slope})$ term is identical for all atoms in the molecule, while the curvature term varies according to the atom type. Both terms are summed over all observations, obs, in the diffraction dataset:

$$\sigma(\text{slope}) = \left(\sum_{\text{obs}} h^2 |\Delta F|^2 \right)^{1/2}, \tag{2.24}$$

$$\text{atomic peak curvature} = \sum_{\text{obs}} h^2 f_i(\sin(\theta)/\lambda) \exp(-B \sin^2(\theta)/\lambda^2)\{m/2\}, \tag{2.25}$$

$$m = \begin{cases} 1 & \text{Centric reflections} \\ 2 & \text{Acentric reflections} \end{cases}. \tag{2.26}$$

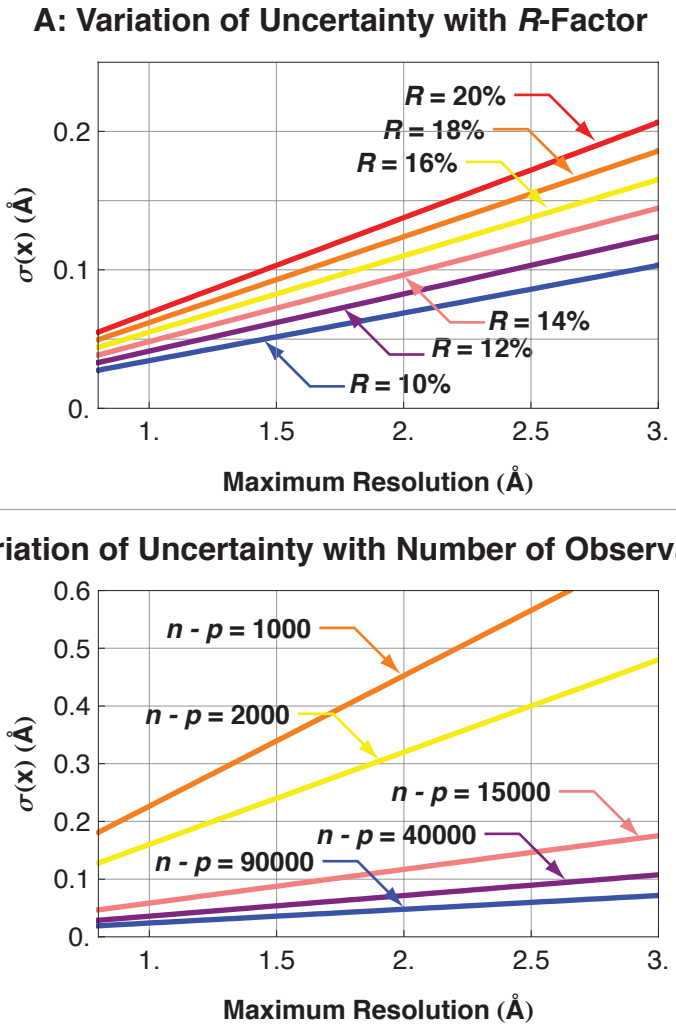


Figure 2.16: Estimated average coordinate uncertainty for C, N and O atoms in protein structures. A: Variation of coordinate uncertainty with maximum resolution for several different *R*-factors. The observation parameter number, *n*, was 15,000, and model parameter number was 8,000. B: Variation of coordinate uncertainty versus maximum resolution for several different observation - parameter (*n* - *p*) numbers. *R*-factor was 0.16.

Centric reflections are those for which $I(h,k,l) = I(-h,-k,-l)$ due to the symmetry of the crystal. Daopin *et al*. [9] concluded that the approximate Cruickshank formula produced an adequate estimate of the uncertainty on the position of atoms in the TGF-β2 molecule. Figure 2.15 shows the estimated uncertainty on the position of carbon, nitrogen and oxygen atoms in the TGF-β2 structure versus B-factor, and the expected uncertainty predicted by the Cruickshank formula shown in equations 2.24 to 2.26. The plots in figure 2.15 indicate that atoms with B-factors lower than 40 Å^2 can be located with sub-angstrom precision.

Cruickshank developed two additional formulas to produce an overall estimate for the coordinate uncertainty in a protein atomic model. By approximating the solution of the equation 2.21, Cruickshank estimated the uncertainty on an atomic coordinate for a single atom in a protein structure derived by unrestrained refinement [133]:

$$\sigma(x_i) = k(N_i/p)^{1/2}\left(g(B_i)/g(B_{\text{avg}})\right)C^{-1/3}Rd_{\text{min}}\,. \qquad (2.27)$$

In this equation, k is ≈ 1.0, B_i is the B-factor of the atom under consideration and B_{avg} is the average B-factor of the fully occupied sites in the atomic model. C is the fractional completeness of the dataset to the resolution d_{min}.

Sheldrick [133] noted that N_i may be approximated as

$$N_i = \sum_j Z_j^{\#2}/Z_i^{\#2}\,, \qquad (2.28)$$

where $Z_j^{\#}$ is the scattering factor of the atom at $\sin\theta/\lambda = 0.3$ Å^{-1}.

The function g is an empirical function of the form

$$g(B) = a_1 + a_2 B + a_3 B^2 . \tag{2.29}$$

However, the coefficients a_1, a_2 and a_3 appear to vary between structures [133], limiting the applicability of equation 2.27.

However, from equation 2.27, Cruickshank was able to derive an approximate formula for the average atomic coordinate uncertainty in an unrestrained atomic model that does not rely upon the use of empirical coefficients:

$$\sigma\left(x, B_{avg}\right) = 1.0\left(N_i / (n-p)\right)^{1/2} C^{-1/3} R d_{\min} . \tag{2.30}$$

To estimate the uncertainty on the coordinates of C, N and O atoms, N_i may be taken as the number of fully occupied atomic sites. For heavy atoms, N_i may be calculated from equation 2.28.

Cruickshank's average, diffraction component only coordinate uncertainty formula predicts that the uncertainty on the position of an atom varies with the number of atomic sites in the molecule, the difference between the number of observations, n and the number of model parameters p, the completeness of the dataset C, and the maximum resolution of the dataset $d_{\min}$.

In the case of low-resolution structures where $n - p$, is negative, Cruickshank proposed an alternative formula based upon the R_{free} factor [138],

$$\sigma\left(x, B_{avg}\right) = 1.0\left(N_i / n_{obs}\right)^{1/2} C^{-1/3} R_{free} d_{min} . \tag{2.31}$$

Cruickshank [133] also noted that the error in position, rather than error in a single coordinate (x, y or z), could be estimated by multiplying the error formulas by $\sqrt{3}$.

$$\sigma\left(r, B_{avg}\right) = \sqrt{3}\sigma\left(x, B_{avg}\right). \tag{2.32}$$

Plots of the behavior of the coordinate uncertainty formulas (equations 2.30 and 2.31) are shown in figure 2.16. It is interesting to note that the approximate coordinate error formulas do not contain an explicit dependence on the average B-factor. However, the B-factor does implicitly enter into the formulas through the R-factor of the atomic model. The R-factor of a protein atomic model may be reduced by increased structural heterogeneity and diminished maximum diffraction resolution associated with an increased B-factor. Additionally, the average coordinate uncertainty formulas of equations 2.30 and 2.31 contain a term for the completeness, C, of the diffraction dataset and a term for the difference between the number of observations, n, in the dataset and the number of parameters, p, in the model derived from the dataset. These terms are however, linked, as a reduction in the total number of observations will result in a reduction in the completeness of the dataset. However, the relationship between C and $n - p$ is complicated, making the inclusion of the two terms appropriate. Variation of the C term alone had negligible effect on the predicted coordinate uncertainty, and so was left at 100% in all of the examples shown.

All of the examples shown in figure 2.16 assume a protein with 2000 fully occupied sites, approximately the number in the Citrine molecule studied in this thesis. Additionally, it was assumed that each atom had 4 associated parameters, 3 coordinates and 1 B-factor, giving a model with 8,000 parameters.

Figure 2.16A shows the estimated coordinate uncertainty versus maximum resolution for several different *R*-factors, ranging from 10% to 20%. The total number of observations was assumed to be 25,0000, typical of Citrine datasets. At a maximum resolution of 2 Å, the coordinate uncertainty ranges from ≈ 0.07 Å at R = 10% to ≈ 0.14 Å at R = 20%.

Figure 2.16B shows the estimated coordinate uncertainty versus maximum resolution for dataset with different numbers of observation – parameter numbers, $n - p$, ranging from 90,000 to 1,000. The variation of $n - p$ produces a large range of coordinate uncertainties at a maximum resolution of 2 Å. For $n - p$ = 90,000, the coordinate uncertainty of the resulting model is predicted to be less than 0.05 Å. For $n - p$ of 1,000, the coordinate uncertainty is predicted to be ≈ 0.45 Å.

These results indicate that even with a maximum diffraction resolution of 2 Å, a high degree of atomic model coordinate precision is possible, even without the use of library constraints. However, the precision of the model is highly dependent upon the number of independent observations in the diffraction dataset used to derive the model.

2.8.3 Estimation of Coordinate Uncertainty due to Refinement Procedure

Urayama *et al.* [4] attempted to quantify the coordinate uncertainty in the structure of sperm whale myoglobin due to differences in macromolecular refinement procedures. Urayama *et al*. [4] compared two atomic models of sperm whale myoglobin at room temperature and pressure that were derived from the same X-ray diffraction dataset, but were refined by different procedures. These two models showed a room mean square deviation in extended main chain atomic positions of 0.095 Å. Urayama *et al.* [4] concluded that the smallest pressure-induced deformation that could be ambiguously identified in sperm whale myoglobin was approximately 0.1 Å.

2.8.4 Conclusions

Several studies have been conducted of the uncertainty on the atomic coordinates of protein atomic models derived by X-ray crystallography. These studies all indicate that for reasonably well ordered atoms, the uncertainty on the coordinates of the atom is well below 1 Å, even when the resolution of the diffraction dataset used to derive the model is well above 1 Å. Intuitively, it is reasonable to expect that fluctuations in the position of an atom due to thermal vibrations or conformational heterogeneity will reduce the precision with which it may be located. Studies by Cruickshank and Daopin *et al.* [9] support this intuition. Similarly, atoms with higher atomic numbers, and consequently higher X-ray scattering powers, may be more precisely located than those with lower scattering powers. The most important determinant of precision in a protein atomic model appears to be the ratio of X-ray diffraction observations to model parameters. The use of constraints in the refinement of an atomic model can significantly reduce the uncertainty on bond lengths and some bond angles in the protein atomic model. However, large numbers of distances in a protein molecule are unconstrained, and must be determined purely by diffraction data.

These results suggest that given X-ray diffraction data with a maximum resolution of 2.0 Å, it is possible, to identify sub-angstrom structural perturbations in a protein molecule, given sufficient observations. With this in mind, chapter 3 examines a protein system where the structure and function of the protein are highly linked, and can be affected by structural changes on the sub-angstrom level.

2.9 Suggested Reading

Bergfors, T.M., *Protein crystallization*. IUL Biotechnology Series. 2008, La Jolla, CA: International University Line.

Perutz, M.F., *Science is Not a Quiet Life: Unravelling the Atomic Mechanism of Haemoglobin*. 1997, London: Imperial College Press.

Guinier, A., *X-ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies*. 1994, New York: Dover.

Warren, B.E., *X-ray Diffraction*. Dover ed. 1990, New York: Dover Publications.

McPherson, A., *Preparation and Analysis of Protein Crystals*. 2nd ed. 1989, Malabar, FL: R.E. Krieger Pub. Co.

Drenth, J. and J. Mesters, *Principles of Protein X-ray Crystallography*. 3rd ed. 2007, New York: Springer.

Rhodes, G., *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. 3rd ed. Complementary science series. 2006, Amsterdam; Boston: Elsevier/Academic Press.

Deutscher, M.P., ed. *Guide to Protein Purification*. Methods in Enzymology. Vol. 182. 1990, Academic Press: San Diego, CA.

Perutz, M.F., *I Wish I'd Made You Angry Earlier: Essays on Science, Scientists, and Humanity*. Expanded ed. 2003, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Perutz, M.F., *Mechanisms of Cooperativity and Allosteric Regulation in Proteins*. 1990, Cambridge, UK: Cambridge University Press.

CHAPTER 3

INTRODUCTION TO CITRINE

3.1 Selection of Citrine

To systematically investigate the correlation between protein function and small structural deformations on the order of 0.1 Å, we sought a model protein that contained an easily probed reporter group, that formed high-quality crystals suitable for X-ray diffraction and was known to display pressure sensitive behavior. Citrine [1], a member of the Yellow Fluorescent Protein (YFP) [52] family of recombinant proteins derived from the Green Fluorescent Protein (GFP) expressed by the jellyfish *Aequorea victoria* [2], satisfied these criteria.

Citrine is an extremely bright, intrinsically fluorescent protein whose atomic structure is known to 2.2 Å resolution at room temperature [1]. The room temperature, room pressure Citrine atomic structure by Griesbeck *et al*. [1] was derived from an X-ray diffraction dataset of approximately 12,200 reflections, has a crystallographic *R*-factor of 0.164 (Protein Data Bank (PDB) accession code 1HUY) and contains approximately 2000 non-hydrogen atoms. Cruickshank's average coordinate uncertainty formula (equation 2.30) suggests that it may be possible to identify structural deformations as small as 0.25 Å in a model of this quality of Citrine at high pressure with 8,000 parameters.

Verkhusha *et al*. [53] observed that the fluorescence peak of EYFP (Enhanced YFP), a close relative of Citrine, shifts towards the red by approximately 4 nm as the hydrostatic pressure applied to it is increased from atmospheric (0.1 MPa) to 300 MPa at room temperature. Citrine displays a similar peak shift of approximately 1 nm per 100 MPa in several buffering solutions. High pressure, room temperature, solution

spectroscopy on Citrine is discussed in section 3.3. The shift of the peak of Citrine's fluorescence spectrum is shown in figure 3.8.

It is appropriate at this time to note that the purpose of the experiments reported in this thesis is not to understand a peak shift in the fluorescence emission from jellyfish. Firstly, no wild-type organism known to exist expresses the Yellow Fluorescent Protein. Secondly, the pressures required to shift the fluorescence peak of Citrine are much higher than those encountered by the *Aequorea* jellyfish on its deepest dive to approximately 200 meters below sea level [139]. The fluorescence peak of Citrine noticeably changes at pressures exceeding 100 MPa, equal to the pressure at the deepest point on the ocean floor, at the bottom of the Marianas Trench, 1100 meters below sea level. The pressure at 200 meters is in contrast only 20 MPa, insufficient to noticeably perturb the fluorescence peak of Citrine.

The purpose of these experiments is to understand the structural basis of the pressure sensitivity of protein function, to provide deeper understanding of the protein structure-function relationship at the sub-Ångstrom level, to highlight problems and opportunities in the engineering of protein molecules, and possibly to suggest means to engineer and optimize protein molecules to perform novel tasks.

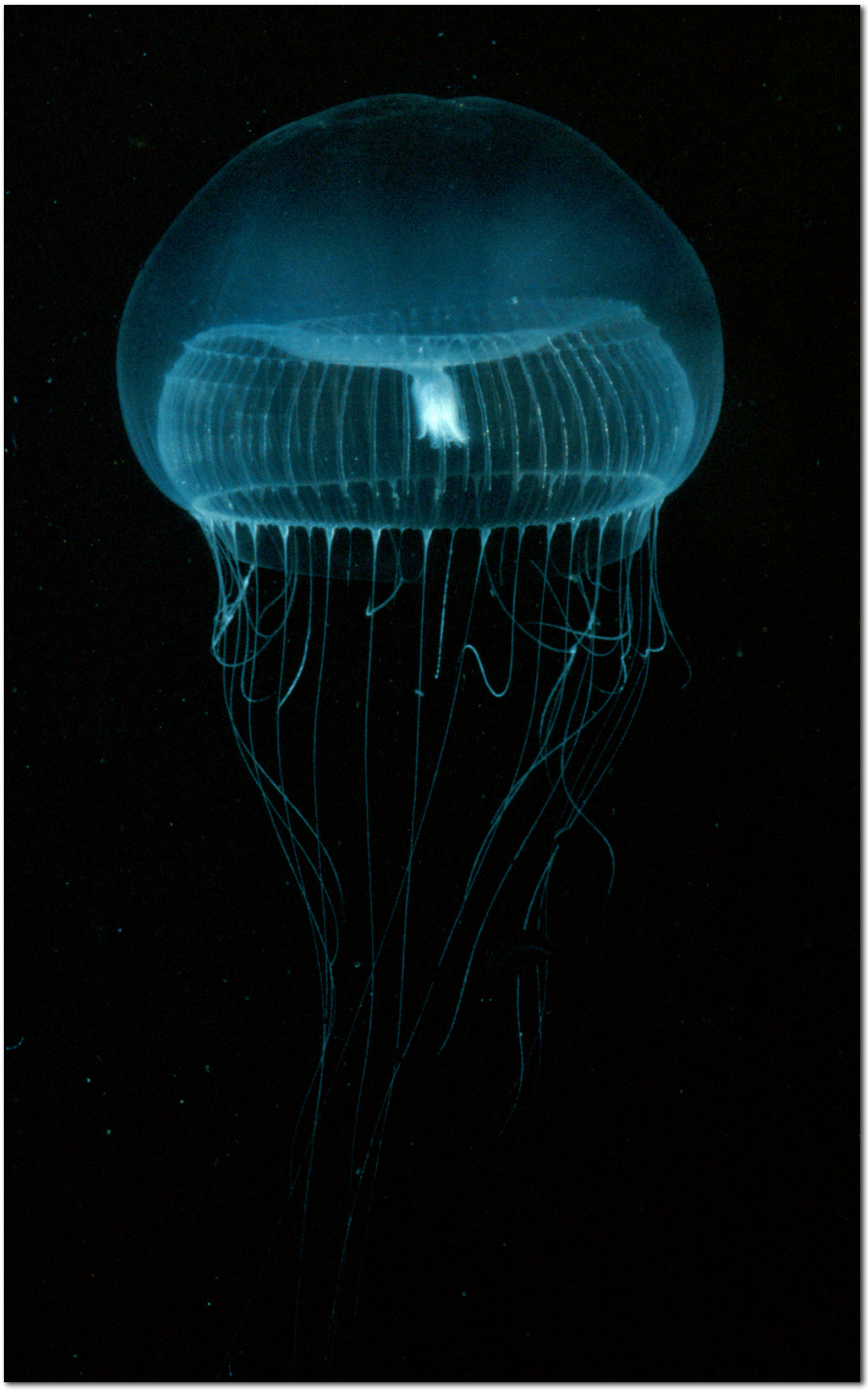


Figure 3.1: Scattered light from the *Aequorea victoria* jellyfish. This scattered light is often confused for fluorescence light. Photograph courtesy of Dr Steven Haddock (Monterey Bay Aquarium, Moss Landing, CA).

3.2 Citrine's Family Background

Citrine is a member of a large family of proteins derived from the wild-type (wt) Green Fluorescent Protein (GFP) (wtGFP) that is naturally expressed in the Pacific Northwestern hydromedusan jellyfish *Aequorea victoria* [2]. The role of GFP in the *Aequorea* jellyfish is to convert, with a very high quantum efficiency, blue light emitted by a photochemical reaction catalyzed by the protein aequorin into green light. A common misconception is that the *Aequorea* jellyfish constantly emits light from large portions of its body. An image of the *Aequorea* jellyfish in figure 3.1 shows light scattered from the jellyfish that is often mistaken for fluorescence light. Fluorescence light is emitted in bursts from small nodules containing aequorin and GFP located around the skirt of the jellyfish. These nodules are shown in figure 3.2.

The biological function of fluorescence from the *Aequorea* jellyfish remains unclear. However, the *Aequorea* jellyfish is known to emit light when touched, leading some to suggest that its purpose may be to act as a "burglar alarm", to ward off predators, or possibly to attract even larger predators to consume the would-be consumer of the jellyfish [140]. The reason why GFP converts blue light to green light remains even more unclear. The evolutionary advantage of emitting green light over blue light, with the resulting small, but significant loss of energy, and the reason why the aequorin enzyme in the jellyfish has not evolved to emit green light without the GFP accessory protein remain even greater mysteries [2].

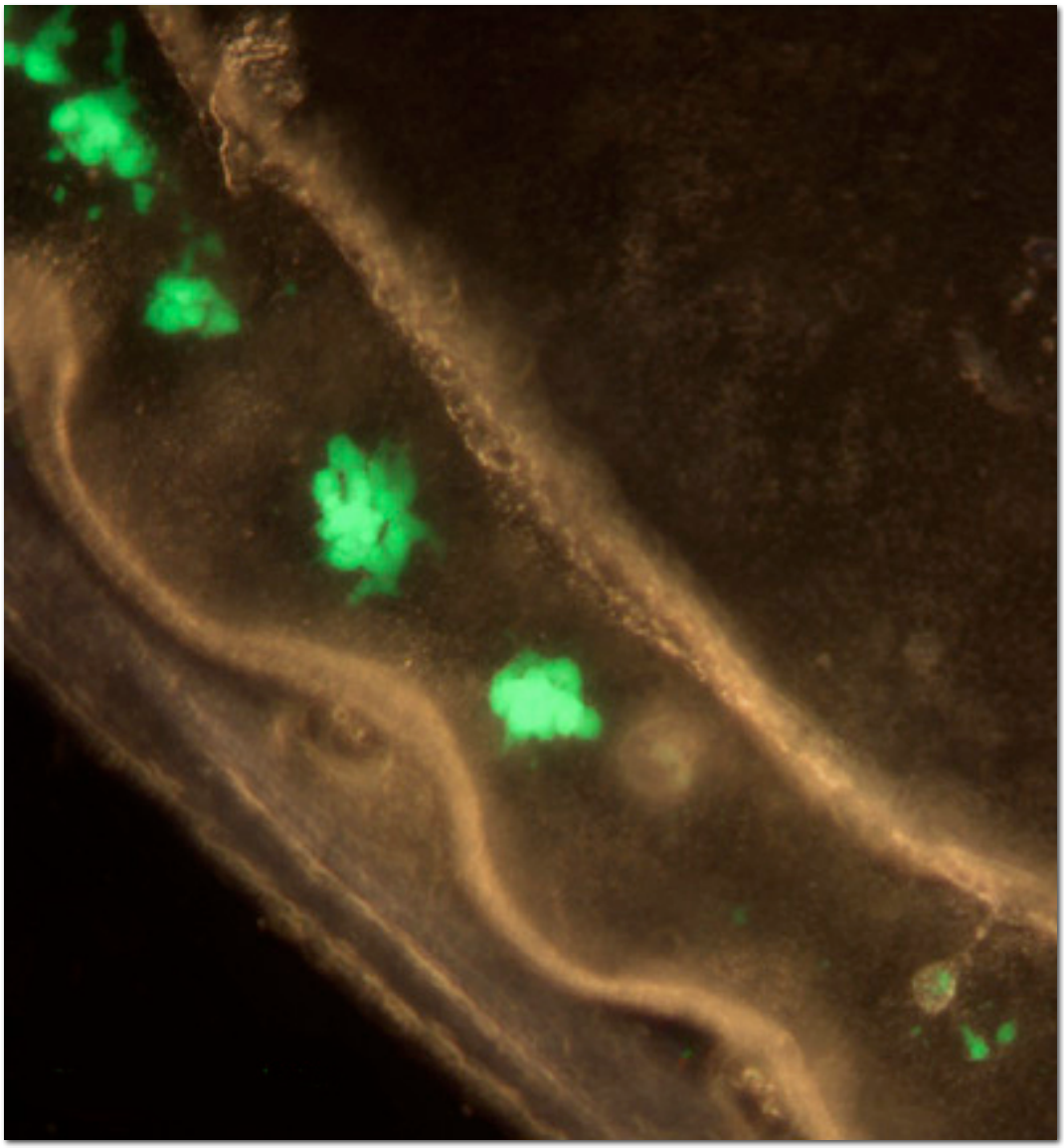


Figure 3.2: Aequorin and GFP containing nodules around the skirt of the *Aequorea victoria* jellyfish. Photograph courtesy of Dr Steven Haddock (Monterey Bay Aquarium, Moss Landing, CA).

In the past 15 years, GFP has acquired a completely new role, separate from its original biological role. GFP is unusual amongst fluorescent proteins in that it requires no prosthetic groups or accessory proteins to become fluorescent. After translation, GFP folds from an unmodified polypeptide chain into its native conformation, a β-barrel, without assistance. The GFP scaffold catalyzes the formation of the light absorbing and emitting element in protein, the *chromophore*, within 90 minutes to 4 hours of folding [141-143].

Another important feature of GFP is that it will acquire its native conformation even with additional residues added to either its N or C terminus. Thus, the *Aequorea* fluorescent proteins may be fused to other proteins and still acquire their native conformation and fluoresce. More importantly, the gene for GFP, once cloned [144] from the *Aequorea* jellyfish, may be added to the start or end of genes encoding other proteins. The new protein complex may be expressed in eukaryotic and prokaryotic organisms [145-147], providing a visible marker for gene expression. This highly desirable and, at the time of its discovery, unique feature of the Green Fluorescent Protein stimulated the creation of mutants of GFP that fluoresce at both longer and shorter wavelengths.

The β-barrel scaffold of all *Aequorea* fluorescent proteins is an extremely stable molecular structure [2]. The start of the polypeptide chain of the *Aequorea* fluorescent proteins folds into a α−helix that lies at the top of the β-barrel. The polypeptide then forms a central 3-10 helix that is wrapped by 11 β-sheets that form the β-barrel. A cartoon representation of the Citrine molecule and its chromophore are shown in figure 3.3. A stereo view of the Citrine molecule is shown in figure 3.4.

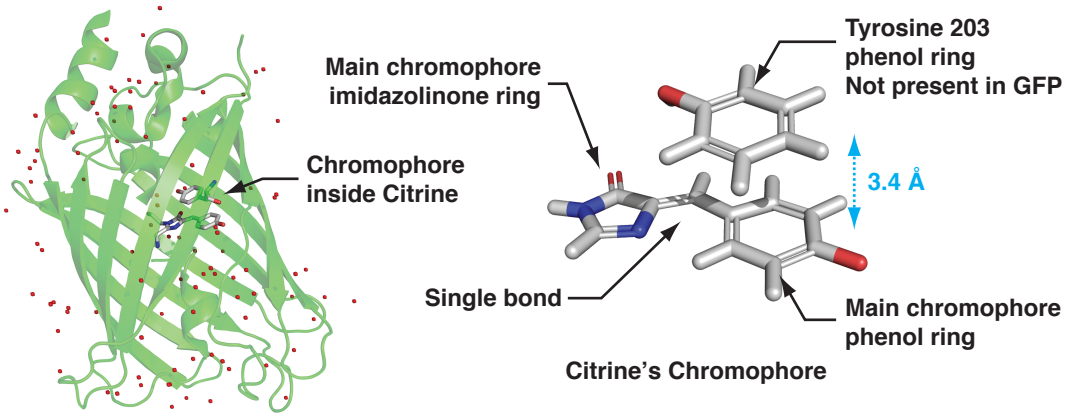


Figure 3.3: The Citrine molecule and Citrine's chromophore.

The β-barrel structure is highly resistant to denaturation and protease action [2]. A temperature of 78 ˚C is required for a sample of wild-type GFP to lose 50% of its fluorescence [148, 149]. Although changes in pH do affect the spectral properties of GFP, there are no reports of denaturation due to excessively acidic or basic conditions. The β-barrel structure is also highly stable under pressure [148]. Wild-type GFP retains its secondary structure up to pressures of 1400 MPa [148]. GFP is highly resistant to chemical denaturants [150]. The β-barrel even resists mechanical attempts to pull it apart [151-154].

The light center, or *chromophore*, of the wild-type GFP auto-catalytically forms from three amino acid residues, Serine 65, Tyrosine 66 and Glycine 67. These three residues lie at the midpoint of the 3-10 helix that runs through the center of the β-barrel. In this thesis, the structure formed by these three residues is referred to as the *main chromophore*. This nomenclature, to my knowledge is not used elsewhere, but is used here to add clarity. The β-barrel forms a cavity that surrounds the main chromophore, sequestering it from the exterior solvent. The solvent protection offered by the β-barrel prevents non-radiative energy emission by the main chromophore and permits it to optically fluoresce.

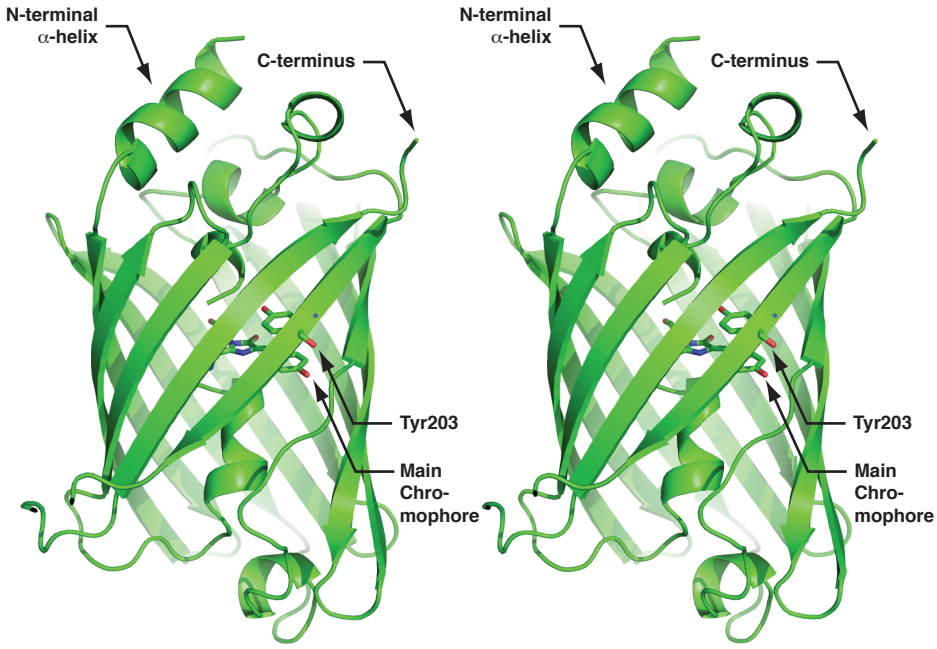


Figure 3.4: Stereo view of the Citrine molecule at a pressure of 0.1 MPa at a temperature of 100 K. Structure taken from Barstow *et al.* [5] (PDB accession code 3DPW).

Niwa *et al.* [155] noted that certain analogs of the *Aequorea* fluorescent protein chromophore (isolated from the β-barrel) absorb light yet are non-fluorescent at room temperature. However, when these chromophore analogs and their solvent are frozen to cryogenic temperatures, they become fluorescent [155]. Additionally, computer simulations on the fluorescence of the *Aequorea* fluorescent protein chromophore indicate that in a vacuum environment the chromophore is fluorescent, while in a liquid environment, where the chromophore is subject to collisions with the solvent molecules, the excited state of the molecule de-excites by quenching rather than by radiative decay [156].

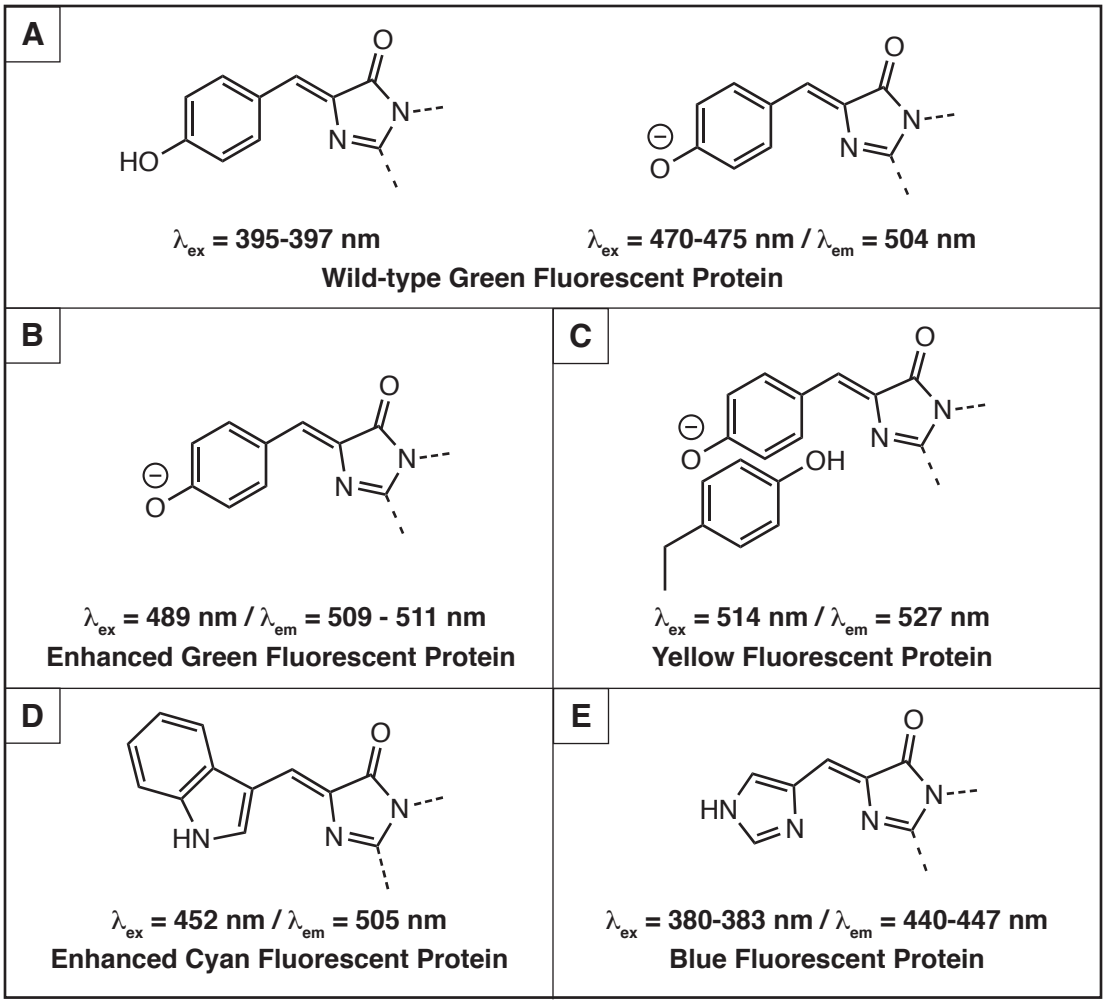


Figure 3.5: *Aequorea* fluorescent protein chromophores. Adapted from Tsien [2].

3.3 Mutants of the Green Fluorescent Protein

The primary determinant of the spectral properties of the *Aequorea* fluorescent proteins is the chemical composition of the main chromophore. Mutation of the three residues that compose the main chromophore, particularly the central residue Tyr66, alters the spectral properties of the protein. The chromophores of several important classes of *Aequorea* fluorescent protein mutants are shown in figure 3.5.

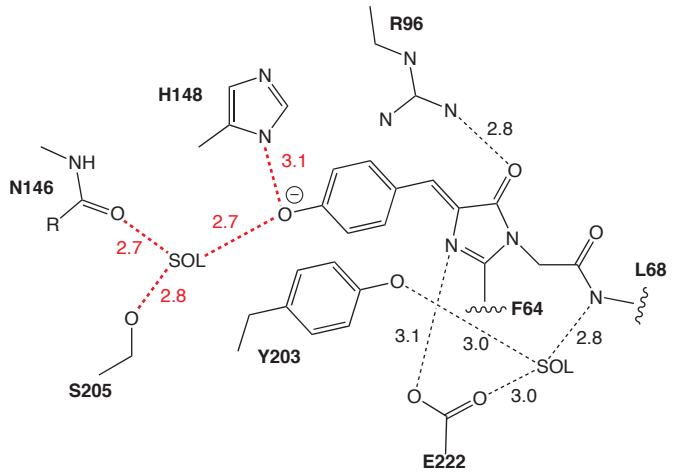


Figure 3.6: Room temperature and pressure Citrine chromophore cavity hydrogen bonding network. Network adapted from Wachter *et al*. [52]. Room temperature, room pressure hydrogen bonding distances from Citrine structure by Griesbeck *et al*. [1] (PDB accession code 1HUY).

The spectral properties of the main chromophore are also tuned by the cavity surrounding it. The main chromophore is embedded in an intricate hydrogen bonding network that runs throughout the cavity in the center of the *Aequorea* fluorescent proteins [2, 52]. Mutation of residues on the walls of this cavity can affect the spectral properties of the *Aequorea* fluorescent protein by perturbing the β-barrel cavity hydrogen bonding network. The effects of these mutations are often more subtle than mutation to the main chromophore [2]. The hydrogen bonding network present in the Citrine chromophore cavity is shown in figure 3.6.

An undesirable feature of the wild-type GFP is that it may be excited at two wavelengths, corresponding to two ionization states of its chromophore. The anionic and neutral ionization states of the wild-type GFP chromophore are shown in figure 3.5A. The neutral form of the chromophore, where the phenolic oxygen is protonated,

absorbs at 397 nm, while the anionic form absorbs at ≈ 470 nm. Upon absorbing a photon, the neutral form of the chromophore becomes de-protonated [157]. The neutral and anionic forms of the chromophore thus both fluoresce at the same wavelength.

An important mutant of GFP is the S65T mutant [158]. The S65T mutant of GFP, where the first residue of the chromophore, serine 65 is replaced by a threonine, alters the hydrogen bonding network in the chromophore cavity, favoring de-protonation of the main chromophore tyrosine [159]. Therefore only one ionization state of the S65T mutant chromophore is present under physiological conditions, and only one band is present in the absorption spectrum of this mutant [2]. Many mutants of the Green Fluorescent Protein contain the S65T mutation [2]. Some mutants, including Citrine, contain the S65G mutation that results in similar spectral behavior.

Blue-shifted mutants of GFP are achieved by mutation of the central residue of the main chromophore. Y66W results in a cyan mutant (CFP), while Y66F results in a blue mutant (BFP). The chromophores of these mutants are shown in figure 3.5. The emission and absorption maxima are shown in table 3.1.

Table 3.1: Absorption and emission maxima for major classes of *Aequorea* fluorescent proteins. Adapted from Tsien [2].

Fluorescent Protein	Absorption, λ_{ex} (nm)	Emission, λ_{em} (nm)	Section in Figure 3.5
Wild-type Green	397, 470	504	A
Enhanced Green	479	511	B
Yellow	510	526	C
Cyan	435	490	D
Blue	380	445	E

Red-shifted mutants of GFP proved harder to fabricate. The X-ray crystal structure of the S65T mutant of GFP [160] and of the wild-type GFP [159, 161] provided information that permitted the rational design of a red-shifted mutant of GFP, called the Yellow Fluorescent Protein (YFP). In the Yellow Fluorescent Protein, residue threonine 203 is mutated to a tyrosine. Residue 203 lies on the wall of the β-barrel and its side-chain projects into the chromophore cavity. The T203Y mutation stacks a phenol ring, the side chain of tyrosine 203, 3.4 Å above the main chromophore found in all *Aequorea* fluorescent proteins. The weak interaction of the main chromophore and the tyrosine 203 phenol is speculated to be responsible for shifting the fluorescence peak of Citrine from green, at 511 nm to yellow at 527 nm [52]. A diagram of Citrine and its chromophore is shown in figure 3.3. The emission and excitation spectra of the Citrine molecule under ambient conditions in solution are shown in figure 3.7. It is notable that absorption and emission spectra of Citrine, and the other YFP molecules are approximately symmetrical about their midpoint.

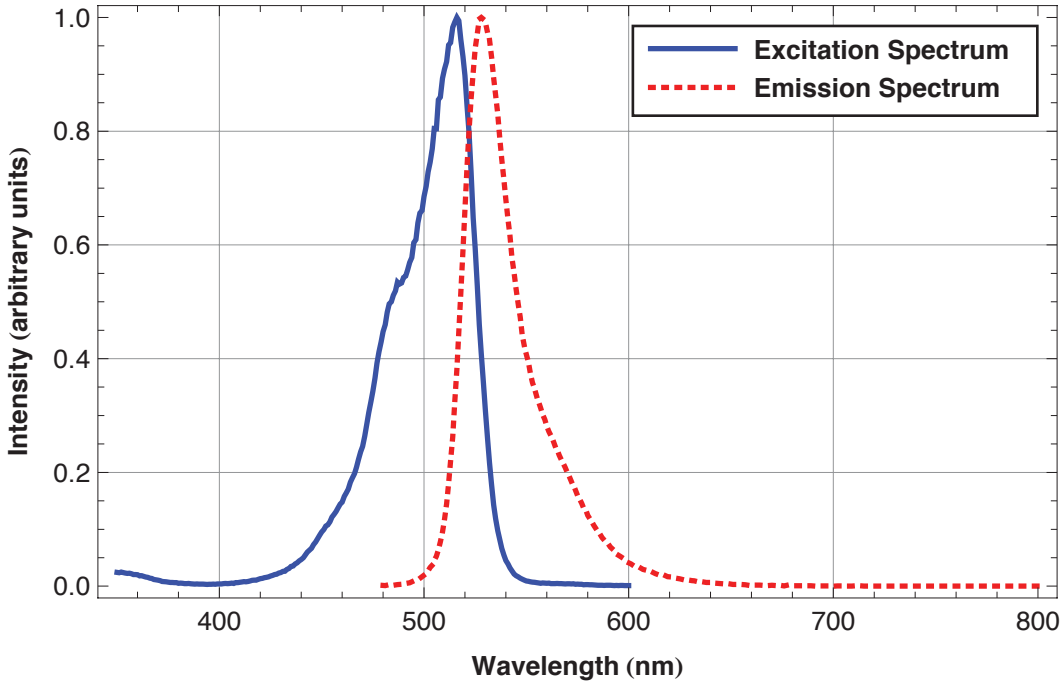


Figure 3.7: Citrine's excitation and emission spectra at room temperature and pressure. For the excitation spectrum, emission was monitored at 529 nm while scanning the excitation beam from 350 to 600 nm. For the emission spectrum, the excitation was fixed at 473 nm. All measurements were made with a Chronos spectrophotometer (ISS, Urbana-Champaign, IL, USA).

One can imagine that a small, pressure induced deformation of the β-barrel scaffold of Citrine could perturb the relative positions of the main part of the chromophore (the main chromophore) and tyrosine 203, changing their weak interaction, and alter the fluorescence properties of Citrine. We sought to demonstrate a direct link between the structure of Citrine at high pressure and a change in its fluorescence spectrum.

The gene encoding Citrine was provided to us by Professor Roger Tsien (University of California at San Diego) in a pRSETB plasmid (Invitrogen, Carlsbad, CA, USA).

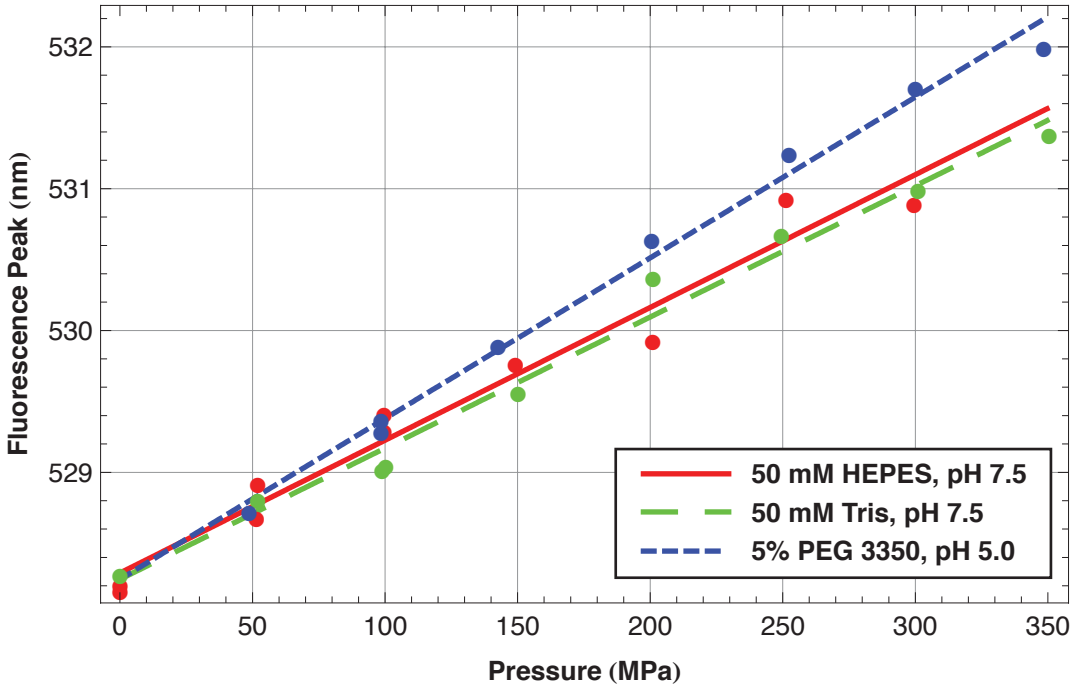


Figure 3.8: Fluorescence peak shift of Citrine in various buffers at room temperature. Peak shift under pressure was measured using a commercially available high-pressure optical cell (ISS, Urbana-Champaign, IL, USA) and a Chronos spectrophotometer (ISS). The complete solution conditions in the final solution are 5% w/v PEG 3350, 50 mM Na acetate, 50 mM NH_4 acetate, pH 5.0.

3.4 Solution Spectroscopy of Citrine at High Pressure and Room Temperature

To confirm the observation by Verkhusha *et al*. [53] that the fluorescence peak of EYFP shifts to the red upon pressurization at room temperature, and of the shift of the fluorescence spectrum of Citrine under high pressure perturbation, we pressurized Citrine in a high-pressure optical cell (ISS, Champaign, IL, USA), and measured its fluorescence spectrum using a spectrophotometer (Chronos, ISS, Champaign, IL, USA). The peak of the fluorescence spectrum was fitted to a polynomial function using the Python module SCIPY (http://www.scipy.org). The polynomial function was analytically differentiated to find the peak.

Citrine displays a linear shift of approximately 1 nm per 100 MPa of its fluorescence peak up to the maximum attainable pressure in the high-pressure spectroscopy system of 400 MPa. This shift was observed in three different buffering solutions: 50 mM HEPES at pH 7.5, 50 mM Tris at pH 7.5 and 50 mM sodium acetate and 50 mM ammonium acetate, 5% w/v PEG 3350 at pH 5.0.

Many buffers display shifts in their buffering pH with pressure, as high pressure often favors the association or dissociation of protons [162]. The buffering properties of Tris and acetate are known to be largely insensitive to pressure [162-164].

Citrine displayed an almost identical shift in fluorescence peak in HEPES at pH 7.5, Tris at pH 7.5 and acetate at pH 5.0. This result implies that the fluorescence peak shift is mechanical in nature, and a direct result of the deformation by pressure of the protein's structure, rather than an effect that is mediated by the protein's solvent. Plots of Citrine's fluorescence peak versus applied hydrostatic pressure at room temperature are shown in figure 3.8. It was very important to clarify the issue of the independence of Citrine's red shift from solvent conditions, so that the fluorescence peak shift could be understood simply in terms of structural changes of the protein, a quantity that is measurable by X-ray crystallography, rather than in terms of changes of the electrostatic environment of Citrine due to changes of pH of the solvent.

3.5 Understanding Citrine's Fluorescence Mechanism

To understand the physical basis of the fluorescence peak shift of the Citrine molecule under high pressure, it is useful to understand the physical basis of Citrine's fluorescence spectrum.

Luminescence is the emission of electromagnetic radiation from an electronically excited state of a *fluorophore* or *chromophore*. Luminescence is typically divided into two categories: *fluorescence* and *phosphorescence*. Fluorescence typically occurs over timescales ranging from 10^{-9} to 10^{-6} seconds, while phosphorescence occurs over much longer timescales; typically 10^{-4} to 10^{2} seconds. This discussion will focus on fluorescence.

Fluorescent fluorophores are typically aromatic molecules. A common feature of many fluorophores, including the Yellow Fluorescent Protein family chromophore, is that the absorption and emission spectra are symmetrical about their midpoint.

Fluorescence is typically initiated by electronic excitation of the fluorophore due to the absorption of electromagnetic radiation. The absorption of photons by the fluorophore occurs by the excitation of a single electron from the ground state of the chromophore to a higher lying unoccupied electronic state. Luminescence is characterized by the spin state of the electrons present in the ground and excited states of the fluorophore. The spins of the two electrons present in the ground and excited states of fluorescent fluorophores are anti-parallel, giving a net spin of 0. The excited state is thus a singlet state. Thus, Pauli repulsion by the electron remaining in the ground state does not impede the rapid return of the photo-excited electron, resulting in the short lifetime of the excited state, and fluorescence emission.

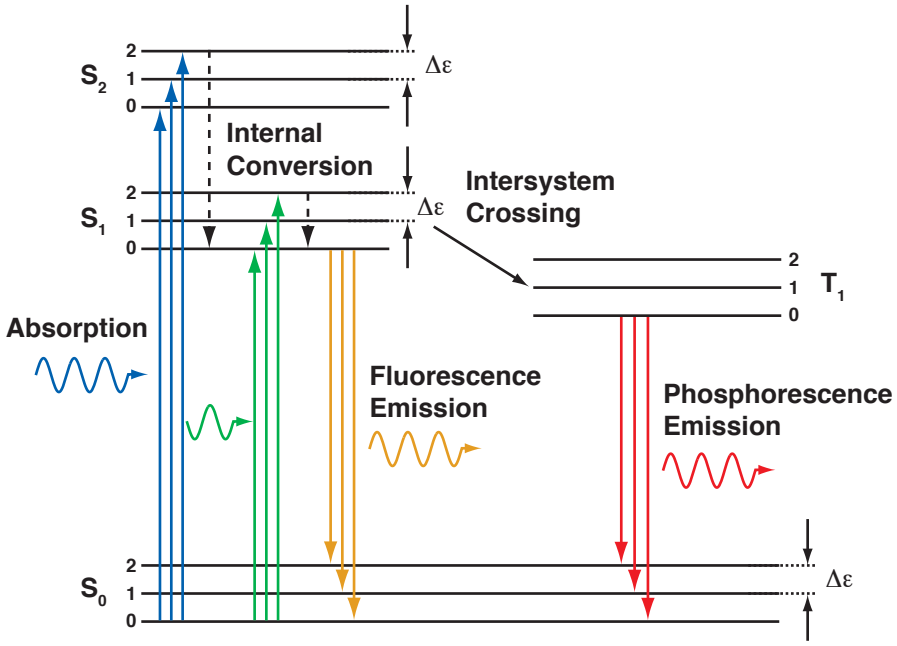


Figure 3.9: Jablonski diagram showing electronic levels involved in fluorescence. The S levels are singlet states with net electron spin of 0, while the T level is a triplet state with a spin of 1. Adapted from Lakowicz [165].

In contrast, in phosphorescent fluorophores, the excited state is a triplet state. Thus, Pauli repulsion impedes the rapid return of the excited electron to the ground state, resulting the long lifetime of phosphorescence.

As the ground and excited states involved in fluorescence emission are singlet states they are often labeled as S states. The ground state is labeled as S_0, while the excited states are labeled from S_1 onwards. The S_0 state is typically the highest occupied molecular orbital (HOMO) of the fluorophore, while the S_1 state is the lowest unoccupied molecular orbital (LUMO) of the fluorophore. Each of the electronic states contain a number of vibrational sub-levels that are due to the vibration of the nuclei of the fluorophore.

Electronic excitation of the fluorophore does not significantly perturb the nuclear geometry, and thus does not significantly perturb the spacing of the vibrational energy levels. Thus, the vibrational sub-structure of each electronic level is largely identical.

A *Jablonski diagram* [165], showing the S_0, S_1 and S_2 states of a generalized fluorophore is shown in figure 3.9. In the absence of illumination, the fluorophore is typically found in the ground vibrational state of the ground electronic state. The absorption peak of the fluorophore is typically close in energy to the band-gap between the S_0 and S_1 states. This band-gap often corresponds to optical or ultra-violet wavelengths.

Light absorption occurs between the S_0 state and the excited states over very short timescales, typically 10^{-15} seconds. This short timescale of excitation does not permit motion of the nuclei of the fluorophore during the excitation process. Thus, the relative probabilities of excitation from the ground state to one of the excited states is determined by the spatial overlap of the ground state and excited state wavefunctions. These relative probabilities determine the shape of the absorption spectrum.

Following light absorption, the fluorophore typically relaxes into the lowest vibrational state of the first excited electronic state, S_1. This internal conversion process typically occurs in approximately 10^{-12} seconds, much shorter than the fluorescence lifetime of 10^{-9} to 10^{-6} seconds. Thus, internal conversion is typically complete before fluorescence emission occurs.

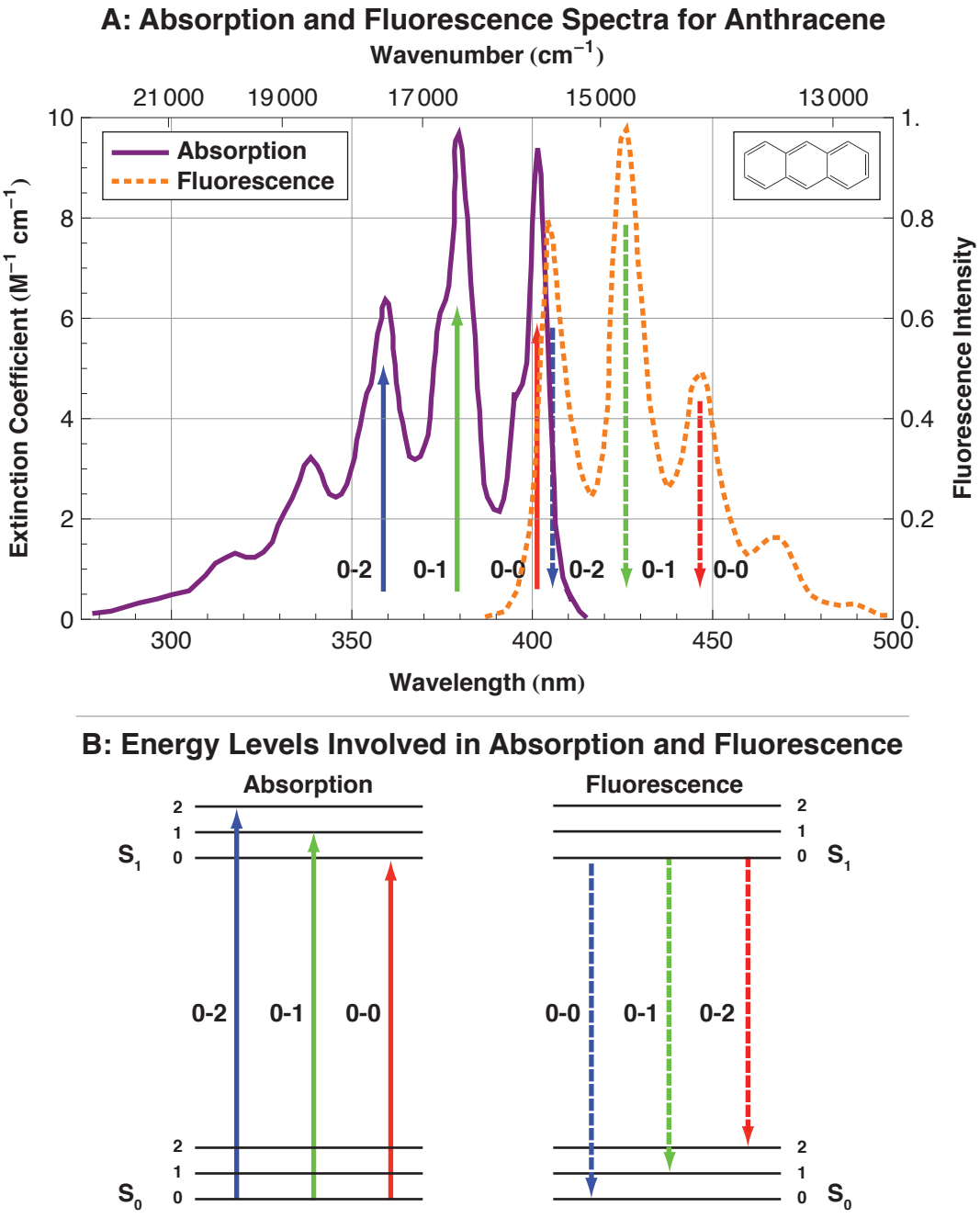


Figure 3.10: Absorption and emission spectra and Jablonski diagram for anthracene. Adapted from Lakowicz [165].

As with absorption, the timescale of fluorescence emission is approximately 10^{-15} seconds. Thus, during the process of fluorescence emission the fluorophore nuclei can be considered stationary. Again, the relative probability of radiative decay from the lowest vibrational state of S_1 to each of the vibrational states of S_0 is determined by the overlap of the wave functions of the ground state and excited state. As the vibrational energy levels of S_0 have the same spacing as those of S_1, the fluorescence spectrum is often the mirror image of the absorption spectrum.

Following radiative decay to one of the ground state vibrational levels, the fluorophore will typically relax to the ground vibrational state of the S_0 state. The loss of energy by internal conversion, and by vibrational relaxation accounts for reduction in energy of emitted photons when compared to absorbed photons.

To illustrate this picture of fluorescence, the absorption and emission spectra of anthracene, and the corresponding Jablonski diagram of the energy levels of anthracene are shown in figure 3.10. The similarity of the absorption and emission spectra of Citrine suggest that this simple framework may also be applied to Citrine as well.

3.6 Computer Simulations of Citrine's Fluorescence Spectrum Under High Pressure

3.6.1 Introduction to Extended Hückel Theory

To gain an intuitive understanding of the spectral shift of Citrine's fluorescence peak at high-pressure, a simple computer model of Citrine's energy levels was constructed using the Extended Hückel Theory (EHT) codes CACAO [166] and YAEHMOP (http://yaehmop.sourceforge.net). CACAO and YAEHMOP both perform identical calculations.

CACAO is known to suffer from a number of bugs but has superior graphical presentation abilities to YAEHMOP. Simulations were performed using both programs and the results were compared. Identical results were obtained in all cases presented here.

Extended Hückel theory (EHT) is one of the simplest formulations for computing the quantum mechanical wave functions of molecules [167]. For a complete description of the theory, the reader should refer to the recent review of quantum chemical modeling techniques by Cramer [167].

The extended Hückel theory is used to calculate the quantum mechanical wave-functions of an extended system, often known as molecular orbitals, from a linear combination of the valence orbitals of the atoms composing the system. Typically, the valence orbitals are the highest occupied electron wave-functions of an atom. Lower lying orbitals are assumed to be unaffected by, and uninvolved in chemical bonding.

As extended Hückel theory neglects the effects of electron-electron interaction, it produces poor estimates of the absolute energies of molecular orbitals. However, the theory is still widely used as given a correct molecular structure, the theory will often correctly identify the general features of the orbitals of the molecule. This computation may be performed extremely rapidly, and thus, may be used to rapidly explore perturbations to the structure of the molecule. The theory often correctly identifies trends in molecular properties given perturbations to the geometry of the molecule [168]. Additionally, extended Hückel provides an extremely intuitive framework in which to describe quantum mechanical basis of the properties of a molecule.

Following Cramer [167], the electronic wave function of a molecule containing N valence orbitals, Ψ, may be written as a superposition of N Slater type atomic orbital wave-functions;

$$\Psi = \sum_{i=1}^{N} a_i \phi_i \,. \tag{3.1}$$

EHT uses a basis set of Slater type s, p and d valence orbitals (YAEHMOP also adds f orbitals in later versions). The angular component of the Slater type orbitals is a spherical harmonic function, $Y_l^m(\theta,\varphi)$, defined by the angular momentum quantum numbers l and m and the spherical angular coordinates θ and φ . The radial component of the orbital is dependent upon the radial coordinate, r, the principal quantum number n, and is scaled by an empirical coefficient ζ, tabulated by Slater [169];

$$\phi_i(r,\theta,\varphi;\zeta,n,l,m) = \frac{[2\zeta]^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} \exp(-\zeta r) Y_l^m(\theta,\varphi). \tag{3.2}$$

The variational principle may be used to determine the energy levels, and wave functions of the molecule. The energy associated with the wave function of the molecule may be calculated using the Hamilton operator, **H**,

$$E = \int_{\text{All space}} d\mathbf{r}\, \Psi^* \mathbf{H} \Psi \Big/ \int_{\text{All space}} d\mathbf{r}\, \Psi^* \Psi \,. \tag{3.3}$$

The energy of the wave function may be estimated by substituting the trial wave function from equation 3.1 into equation 3.3:

$$E = \int_{\text{All space}} d\mathbf{r}\left(\sum_{i=1}^{N} a_i\phi_i\right)\mathbf{H}\left(\sum_{j=1}^{N} a_j\phi_j\right) \Bigg/ \int_{\text{All space}} d\mathbf{r}\left(\sum_{i=1}^{N} a_i\phi_i\right)\left(\sum_{j=1}^{N} a_j\phi_j\right), \tag{3.4}$$

$$E = \sum_{i,j} a_i a_j \int d\mathbf{r}\,\phi_i \mathbf{H}\phi_j \Big/ \sum_{i,j} a_i a_j \int d\mathbf{r}\,\phi_i\phi_j\,. \tag{3.5}$$

Defining a *resonance integral* matrix element

$$H_{ij} = \int d\mathbf{r}\,\phi_i \mathbf{H}\phi_j\,, \tag{3.6}$$

and an *overlap integral* matrix element

$$S_{ij} = \int d\mathbf{r}\,\phi_i\phi_j\,. \tag{3.7}$$

S_{ij} measures the spatial overlap between the atomic orbitals ϕ_i and ϕ_j. It is important to note that this integral is independent of the probability amplitudes associated with these orbitals, a_i and a_j, which will be calculated later.

Thus,

$$E = \sum_{i,j} a_i a_j H_{ij} \Big/ \sum_{i,j} a_i a_j S_{ij}\,. \tag{3.8}$$

The energy *E*, may be minimized with respect to the probability amplitudes, a_k, of the valence orbitals, in order to estimate the energy of the ground state of the molecule:

$$\frac{\partial E}{\partial a_k} = 0 \quad \forall k \,. \tag{3.9}$$

Differentiating *E* with respect to a_k produces a set of *N* secular equations

$$\sum_{i=1}^{N} a_i (H_{ki} - ES_{ki}) = 0 \quad \forall k, \tag{3.10}$$

often written as a secular determinant

$$\begin{vmatrix} H_{11} - ES_{11} & \cdots & H_{1N} - ES_{1N} \\ \vdots & \ddots & \vdots \\ H_{N1} - ES_{N1} & \cdots & H_{NN} - ES_{NN} \end{vmatrix} = 0 \,. \tag{3.11}$$

The set of secular equations has *N* roots, corresponding to *N* energy levels. The set of *N* energies may be substituted into equation 3.10 to solve for the set of probability amplitudes, a_{ij}, giving a set of *N* molecular orbitals

$$\psi_j = \sum_{i=1}^{N} a_{ij} \phi_i \,, \tag{3.12}$$

each corresponding to one of the *N* energy levels.

Each energy level is filled by two electrons, up to the maximum number of valence electrons in the system. This allows identification of the highest occupied (HOMO) and lowest unoccupied molecular orbitals (LUMO) of the molecule.

In order to calculate the energy levels of the molecule, the matrix elements of the secular determinant in equation 3.11 must be specified. In extended Hückel theory, the matrix elements are calculated by a set of simple approximations.

The diagonal elements of the resonance integral matrix, $H_{\mu\mu}$, are approximated by the valence state ionization potential of the electronic orbital in question. For example, for a hydrogen 1s orbital, H = -13.6 eV. The diagonal overlap integral matrix elements, $S_{\mu\mu}$ are equal to 1.0.

The off-diagonal resonance integral matrix elements are approximated as

$$H_{\mu\nu} = \frac{1}{2} C_{\mu\nu} \left(H_{\mu\mu} + H_{\nu\nu} \right) S_{\mu\nu} , \tag{3.13}$$

where $C_{\mu\nu}$ is an adjustable parameter called the Hückel coefficient. $C_{\mu\nu}$ is typically set to 1.75 for all matrix elements. $H_{\mu\mu}$ and $H_{\nu\nu}$ are the ionization potentials for atomic orbitals μ and ν.

The set of probability amplitudes for each energy level, a_{ij} , may be computed by substituting each energy E_i into equation 3.10.

A simple example of the extended Hückel theory is useful in acquiring an intuitive understanding of the theory. A system of two hydrogen 1s orbitals, the H_2 molecule, is a suitably simple example.

The secular determinant for a two orbital system is;

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} \\ H_{21} - ES_{21} & H_{22} - ES_{22} \end{vmatrix} = 0 . \tag{3.14}$$

As the two orbitals are identical, the resonance integrals are identical and equal to *H*. For a hydrogen 1s orbital, *H* = -13.6 eV. Similarly, the overlap integrals S_{12} and S_{21} are both equal to *S*. S_{11} and S_{22} are both equal to 1.0. Thus,

$$\begin{vmatrix} H - E & CHS - ES \\ CHS - ES & H - E \end{vmatrix} = 0 . \tag{3.15}$$

This gives two non-degenerate roots for *E,*

$$E_1 = (H - CHS)/(1 - S), \tag{3.16}$$

$$E_2 = (H + CHS)/(1 + S) . \tag{3.17}$$

Giving an energy gap between the two levels of

$$\Delta E = \frac{2HS(1-C)}{1-S^2} . \tag{3.18}$$

The two corresponding probability amplitudes for this can be found by substituting the two values of *E* into equation 3.10. For E_1,

$$\begin{aligned} a_1 &= -a_2 \\ a_1 &= 1/\sqrt{2} \\ a_2 &= -1/\sqrt{2} \end{aligned} \quad . \tag{3.19}$$

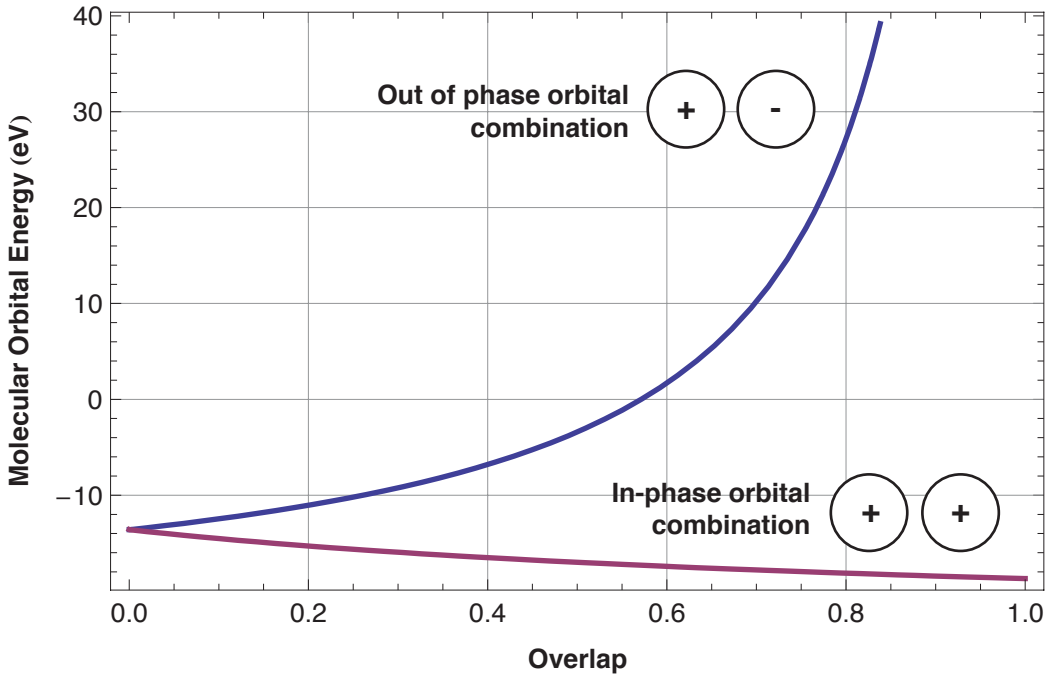


Figure 3.11: Energy of molecular orbitals of H_2 molecule computed by Extended Hückel Theory. Note that molecular orbital composed of an out of phase atomic orbital combination increases in energy with increasing overlap, while the in-phase orbital combination decreases in energy.

The corresponding amplitudes for E_2,

$$a_1 = a_2 = 1/\sqrt{2}\,. \tag{3.20}$$

A plot of E_1 and E_2 with increasing orbital overlap is shown in figure 3.11. It is interesting to note neither E_1 nor E_2 display minima in this plot, highlighting that the extended Hückel Hamiltonian does not include a nuclear repulsion term. However, the plot does highlight an important feature of the theory, generally true, that the overlap of orbitals of opposite phase tends to raise the energy of, or destabilize, a molecular orbital, demonstrated by the increasing energy of molecular orbital 1 with increasing overlap between the two constituent atomic orbitals.

Alternatively, in phase atomic orbital overlap tends to stabilize molecular orbitals, highlighted by the decreasing energy of molecular orbital 2 with the increasing overlap of atomic orbitals 1 and 2.

3.6.2 Justification for Use of Extended Hückel Theory

As the Citrine fluorescence lifetime is very short, ≈ 3.6 ns [170] and its absorption and emission spectra are symmetrical about their overlap point, it is reasonable to believe that the simplified model of fluorescence described in section 3.5 may reasonably describe the fluorescence mechanism of Citrine.

X-ray crystallographic structures of protein molecules under high pressure [4, 24, 26-29] indicate pressures up to a few hundred MPa do not significantly alter covalent bond lengths in protein molecules. Additionally, experiments by Isaacs *et al*. [39] suggest pressures in the same range will also not appreciably alter the vibrational frequencies of the atoms in the Citrine chromophore. Thus, it is not unreasonable to believe that the energy spacing of the vibrational energy levels of the Citrine chromophore will be unaffected by the application of pressures up to a few hundred MPa. For these reasons, it is plausible that the fluorescence peak shift of the Citrine molecule under high pressure may be due to the variation of the energy band-gap between the ground state of the chromophore, the highest occupied molecular orbital (HOMO), and the first excited state of the chromophore: the lowest unoccupied molecular orbital (LUMO).

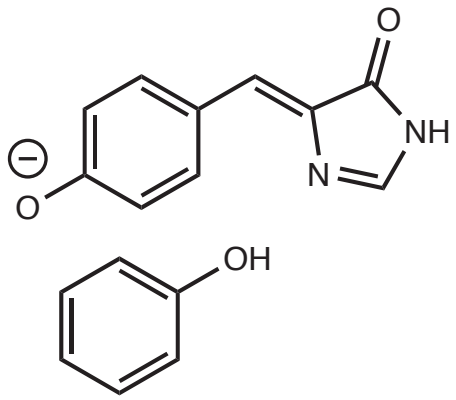


Figure 3.12: Truncated chromophore structure input to Extended Hückel model of Citrine's fluorescence transition.

Although the extended Hückel method is known to perform poorly at computing absolute energy levels, it does allow the identification of trends in the variation of energy levels as a function of the geometry of the system [168]. This may allow an extended Hückel to be used to semi-quantitatively explain the fluorescence shift of the Citrine molecule under high pressure in terms of the deformation of the Citrine chromophore structure by high pressure.

3.6.3 Results of Extended Hückel Simulations of Fluorescence Peak Shift of Citrine

Experiments with an idealized chromophore system give intuitive insight into the structural nature of the chromophore deformation, and the quantum mechanical mechanism of the fluorescence peak shift of the Citrine molecule under pressure.

An idealized chromophore system composed of a truncated main chromophore and the phenol ring of tyrosine 203 was constructed. The initial configuration of this system placed the center of the tyrosine 203 phenol 3.5 Å directly above the center of the main chromophore phenol. The main chromophore and tyrosine 203 phenol ring were coplanar. Hydrogen atoms were added to correct the valence of the input structure.

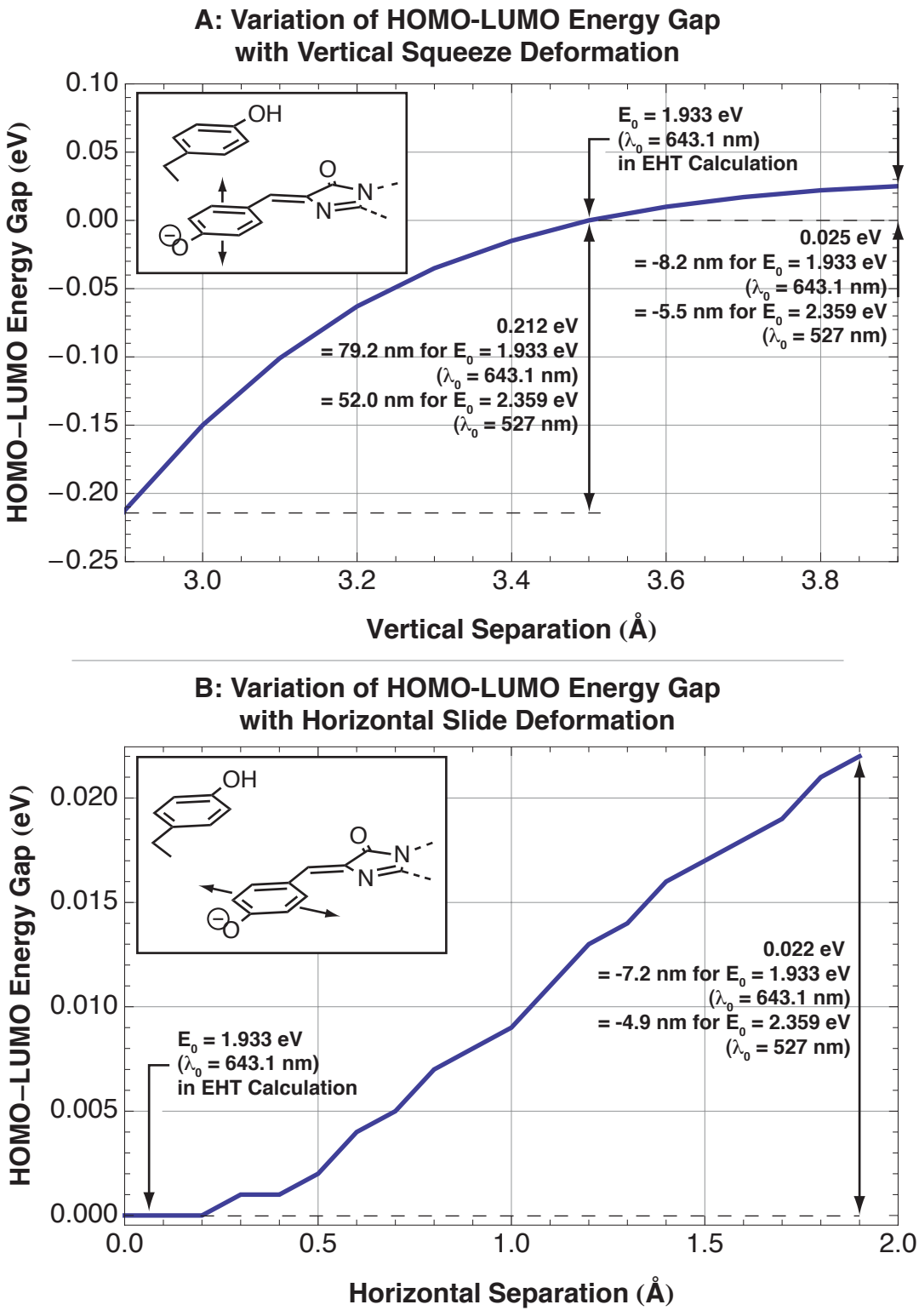
A diagram of this structure is shown in figure 3.12. The structure was placed in the anionic state, thought to be the predominant protonation state of Citrine, where the phenolic oxygen is de-protonated [1, 2], with a charge of -*e*.

The fluorescence of the chromophore is modeled as an electronic transition between the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) of the truncated chromophore structure.

The energy difference between the HOMO and LUMO in the initial configuration was calculated by CACAO and YAEHMOP to be 1.933 eV, corresponding to an emission wavelength of 643 nm. The experimentally measured Citrine emission peak wavelength at room pressure is 527 nm, corresponding to an energy of 2.359 eV.

Two deformation motions of the truncated Citrine chromophore were investigated with the CACAO code. The first motion was a *vertical squeeze*. In this simulation the center of the phenol ring of the chromophore and the center of the phenol ring of tyrosine 203 were placed directly above one another. The vertical separation of the two rings was varied and the molecular orbitals of the combined system were computed at each step. The second motion that was investigated was a *horizontal slide*. In this motion the vertical separation of the two rings was fixed at 3.5 Å and the center of the phenol of tyrosine 203 was moved horizontally away from the center of the phenol of the main chromophore and the molecular orbitals were again computed at each step. Diagrams showing these two deformation motions are shown as insets in figures 3.13A and 3.13B.

Figure 3.13: Results of Extended Hückel Theory (EHT) calculation of HOMO-LUMO energy gap of Citrine. A: Variation of energy gap with horizontal slide deformation. Vertical separation is fixed at 3.5 Å. B: Variation of energy gap with vertical squeeze deformation. Horizontal separation is fixed at 0.0 Å.


A: Variation of HOMO-LUMO Energy Gap
with Vertical Squeeze Deformation
HOMO–LUMO Energy Gap (eV)
0.10
0.05
0.00
−0.05
−0.10
−0.15
−0.20
−0.25
3.0
3.2
3.4
3.6
3.8
Vertical Separation (Å)
OH
O
N
N
O
E0 = 1.933 eV
(λ0 = 643.1 nm)
in EHT Calculation
0.025 eV
= -8.2 nm for E0 = 1.933 eV
(λ0 = 643.1 nm)
= -5.5 nm for E0 = 2.359 eV
(λ0 = 527 nm)
0.212 eV
= 79.2 nm for E0 = 1.933 eV
(λ0 = 643.1 nm)
= 52.0 nm for E0 = 2.359 eV
(λ0 = 527 nm)
B: Variation of HOMO-LUMO Energy Gap
with Horizontal Slide Deformation
HOMO–LUMO Energy Gap (eV)
0.020
0.015
0.010
0.005
0.000
0.0
0.5
1.0
1.5
2.0
Horizontal Separation (Å)
OH
O
N
N
O
0.022 eV
= -7.2 nm for E0 = 1.933 eV
(λ0 = 643.1 nm)
= -4.9 nm for E0 = 2.359 eV
(λ0 = 527 nm)
E0 = 1.933 eV
(λ0 = 643.1 nm)
in EHT Calculation

The variation of the band-gap energy between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) were extracted from the simulations. The shift of the band-gap from the initial orientation band-gap is plotted versus vertical separation in figure 3.13A and plotted versus horizontal separation in figure 3.13B.

The horizontal slide motion produces a small, gentle linear relationship between the energy band-gap between the chromophore HOMO and LUMO and the horizontal separation of the two phenol rings. The band-gap energy gently increases with increasing horizontal separation of the two rings, corresponding to a slight blue-shift in the fluorescence peak. This gentle linear shift resembles the linear relationship seen between applied pressure and wavelength red-shift seen in experiment and shown in figure 3.8.

In the horizontal slide deformation case, a horizontal motion of 1.9 Å from the initial configuration produces an increase in the HOMO-LUMO band-gap energy of 0.022 eV. This corresponds to a blue-shift of the fluorescence peak of 7.2 nm for an initial band-gap of 1.933 eV. This small wavelength shift is of the same magnitude as those observed in high-pressure experiments on Citrine. The corresponding deformation of 1.9 Å is large, but not unreasonably large when compared with previously observed protein structural deformations under high pressure. A deformation motion of 1.0 Å is certainly possible under high pressure, and corresponds to an energy band-gap shift of 0.009 eV, or a 3 nm blue-shift.

By contrast, the vertical squeeze deformation motion produces a band-gap energy shifts that are almost an order of magnitude larger than those seen in the horizontal

slide deformation motion, with a far less linear relationship between deformation and the HOMO-LUMO band-gap energy. These computer simulation results suggest that it is more likely that the predominant deformation motion in the Citrine molecule under high pressure is a horizontal slide of the main chromophore and tyrosine 203 phenol.

It is worth noting that when calculating wavelength shifts from these calculations, if one replaces the initial band-gap energy calculated by simulation with that measured by experiment one produces very reasonable wavelength shift estimates. For instance, in the case of the 1.9 Å horizontal slide motion shown in figure 3.13B, if one replaces the initial configuration energy of 1.933 eV with the experimentally measured value of 2.359 eV, and then calculates the emission wavelength shift, one arrives at a blue-shift of 4.9 nm. We speculate that although the Extended Hückel model may incorrectly calculate the baseline band-gap energy for this system, it may actually correctly calculate the perturbation on the main chromophore by the phenol of tyrosine 203.

3.6.4 Mechanism of Fluorescence Peak Shift

The extended Hückel model also provides intuitive insight into the mechanism of the narrowing of the energy gap between the HOMO and the LUMO. The energy gap between the HOMO and the LUMO is narrowed as the tyrosine ring and the chromophore move closer together due to the destabilization, or raising in energy, of the HOMO. In contrast, the energy of the LUMO remains unperturbed.

Molecular orbital diagrams of the HOMO and LUMO were generated with CACAO and are shown in figure 3.14. Sections of atomic orbitals with negative probability amplitudes are colored in red, while those with positive amplitudes are colored blue.

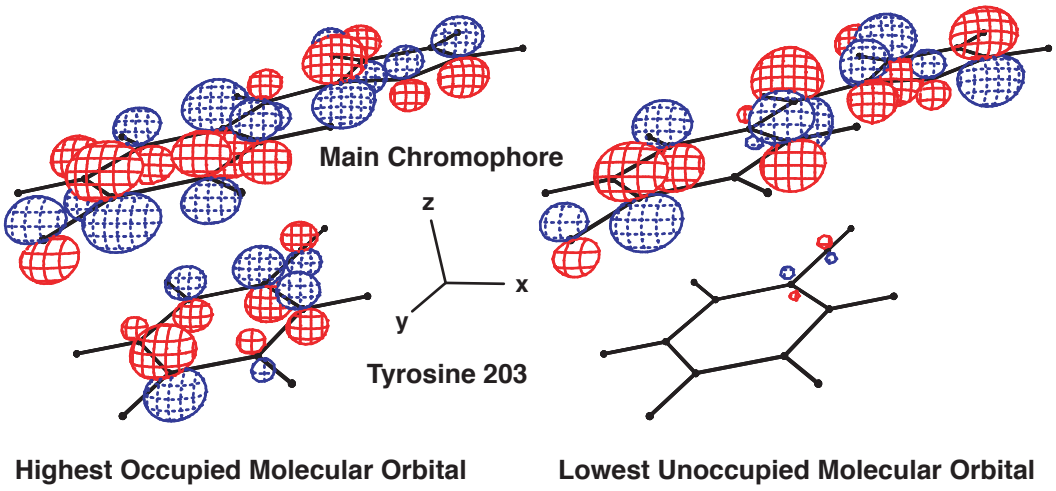


Figure 3.14: Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) of truncated Citrine chromophore calculated and visualized with CACAO. Note that the LUMO takes almost no atomic orbital contribution from the perturbing tyrosine ring. Thus, the energy of the LUMO is insensitive to the relative orientation of the main chromophore and tyrosine 203.

The most striking feature of these models is that while the HOMO takes significant contributions from the atomic orbitals on the phenol of tyrosine 203, the LUMO takes a much smaller contribution. The HOMO is significantly destabilized by the overlap of oppositely phased *p*-orbitals that project from tyrosine 203's phenol ring, raising its energy, while the energy of the LUMO remains unperturbed. This suggests that if one were to alter the orientation of the two rings, one would alter the overlap of the *p*-orbitals on the two rings, alter the energy of the HOMO, and thus alter the energy gap between it and the LUMO, shifting the fluorescence peak.

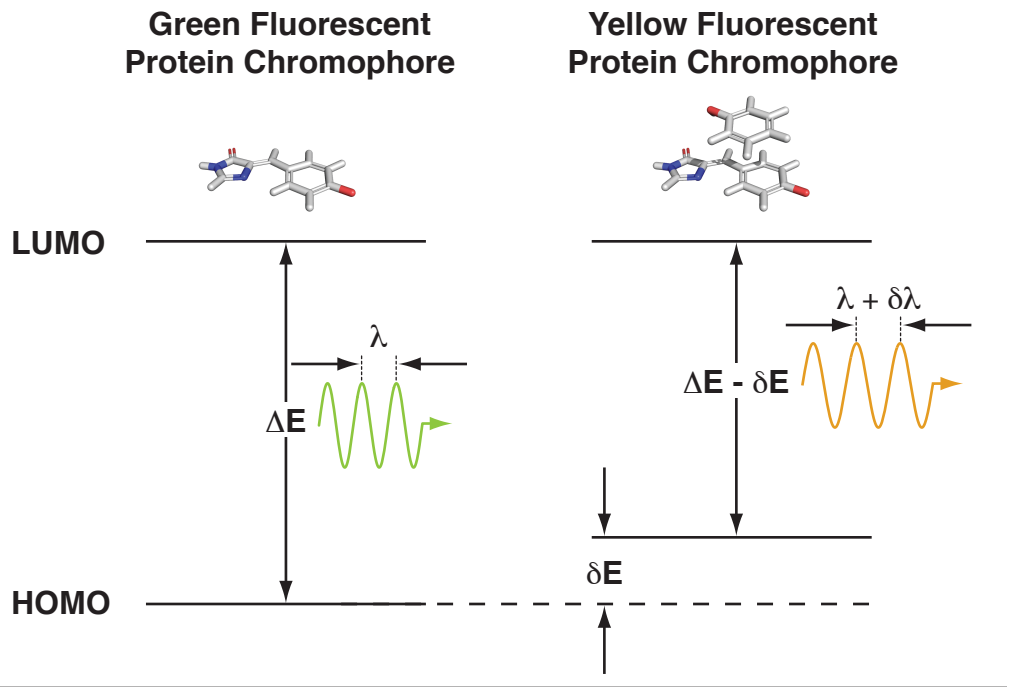


Figure 3.15: Perturbation of Highest Occupied Molecular Orbital (HOMO) in Citrine chromophore by interaction of tyrosine 203 phenol.

A schematic of the narrowing of the band-gap energy between the HOMO and LUMO of the chromophore is shown in figure 3.15.

Chapter 4 discusses the solution of the structure of the Citrine molecule under high pressure using a novel high-pressure cryo-cooling technique [3], and the effect of these high-pressure cryo-cooling conditions on the fluorescence peak.

CHAPTER 4

THE DEFORMATION OF THE CITRINE CHROMOPHORE AND A SPECTRAL PEAK SHIFT UNDER HIGH PRESSURE

4.1 Introduction

In chapter 3, it was established that the peak of the fluorescence spectrum of the Citrine molecule shifts to the red under the application of high hydrostatic pressure at room temperature. A simple quantum mechanical model of the Citrine fluorescence spectrum suggests that the fluorescence peak shift may be due to a sub-angstrom relative sliding motion of the main chromophore and the perturbing tyrosine 203 phenol that compose the fluorescent chromophore of the Citrine molecule.

Section 4.2 of this chapter details the growth and purification of Citrine. Section 4.3 describes the construction of a safety enclosure for the high-pressure cryocooling apparatus. Section 4.4 describes a micro-spectrophotometer that was used to measure the optical properties of Citrine under high-pressure cryocooling conditions. Section 4.5 of this chapter describes the effects of high pressure on the fluorescence spectrum of Citrine[1]. Section 4.6 describes the crystallization of Citrine. Section 4.7 describes the collection of X-ray diffraction data from crystals of Citrine.

To establish a structural basis for the fluorescence peak shift of Citrine under high pressure, the pressurized structure of the Citrine molecule was solved over a range of pressures from 50 to 500 MPa using the high-pressure cryo-cooling technique developed by Kim *et al.* [3] and introduced in chapter 2. These pressurized structures

[1] The results presented in sections 4.5 and 4.8 of this chapter were published in [5]., Barstow, B., N. Ando, C.U. Kim and S.M. Gruner, *Alteration of citrine structure by hydrostatic pressure explains the accompanying spectral shift*. Proceedings of the National Academy of Sciences of the U.S.A., 2008. **105**(36): p. 13362-6. Copyright 2008 National Academy of Sciences, U.S.A.

were analyzed to find deformations to the chromophore of the molecule, and to find deformations to the entire molecule that may actuate any deformation seen at the chromophore. The deformations to the chromophore are discussed in section 4.8 of this chapter. Deformations to the global structure of Citrine will be discussed in chapter 5.

The high-pressure cryo-cooling technique [3, 4] greatly eases high-pressure X-ray crystallography when compared to room temperature high-pressure X-ray crystallography. However, sub-angstrom structural shifts of atoms in protein molecules not only occur in response to high pressure, but also occur due to thermal contraction in response to cooling to cryogenic temperatures. Cooling of the protein crystal under high pressure complicates the analysis of the protein structure, as the deformations locked into the protein molecules composing the crystal by the high-pressure cryo-cooling procedure are a combination of the effects of elevated pressure and of lowered temperature. The result of this combination is that the high-pressure cryo-cooled state is structurally distinct from both the high pressure, room temperature, and low temperature, ambient pressure structures, and is likely to also have distinct functional properties.

The structure of the Citrine molecule at 400 MPa and room temperature is unlikely to be identical to the structure of a Citrine molecule that has been high pressure cryo-cooled at 400 MPa. Thus, the fluorescence properties of Citrine under high pressure at room temperature, and under high-pressure cryo-cooling are likely to differ. For this reason, the high-pressure cryo-cooling technique cannot be immediately used to structurally resolve the cause of pressure-induced modifications to protein function at room temperature.

Attempts to solve the structure of Citrine under high pressure at room temperature using the beryllium high-pressure X-ray crystallographic cell [23, 24, 28, 29, 69] that was introduced in chapter 2 were unsuccessful. We believe this failure to be due to two factors. Firstly, Citrine crystals at high pressure at room temperature appear prone to dissolution and cracking [171]. Secondly, the small size of Citrine crystals results in X-ray diffraction that is weak when compared to the X-ray scattering background from the walls of the high-pressure beryllium cell.

However, with high-pressure cryo-cooling, direct correlations can be drawn between the high-pressure cryo-cooled structure, and the fluorescence properties of high-pressure cryo-cooled samples. With the detailed structure-function relationship that can be derived from high-pressure cryo-cooling experiments, it may be possible to explain the room temperature high-pressure effects on the fluorescence spectrum of Citrine.

This chapter first details the growth and purification of recombinant Citrine molecules and the growth of high-quality crystals of Citrine, suitable for diffraction analysis. The chapter then turns to the design of a micro-spectrophotometer suitable for measuring the fluorescence spectra of high-pressure cryocooled Citrine samples, and the impact of high-pressure cryocooling on the fluorescence spectrum of Citrine. The chapter concludes with details of X-ray data collection from high-pressure cryocooled Citrine crystals, refinement of the high-pressure cryocooled Citrine structures, verification of the existence of pressure-induced deformations to the Citrine structure, and analysis of these deformations.

4.2 Growth and Purification of Citrine

4.2.1 Modification of Citrine Plasmid

The gene encoding Citrine was provided by Professor Roger Tsien (University of California, San Diego) in a pRSETB plasmid (Invitrogen, Carlsbad, CA, USA). Although small quantities of Citrine were expressed in *E. coli* using the pRSETB vector, a few milligrams, this vector has a number of limitations that make it undesirable for large-scale expression of Citrine.

As this project involves the use of recombinant DNA molecules and genetically altered microorganisms, it was registered with the Cornell University Institutional Biosafety Committee (IBC) under Memorandum of Understanding and Agreement (MUA) number 15613.

Manipulation of the Citrine gene and large-scale culture of Citrine producing bacteria was carried out at the Cornell University Protein Facility in the S.T. Olin Laboratory with the help of Dr Cynthia Kinsland.

The pRSETB plasmid has three major limitations. Firstly, the antibiotic resistance scheme used in the pRSETB plasmid, while adequate, would benefit from improvement. In addition to the gene for Citrine, the pRSETB plasmid encodes ampicillin resistance. *E. coli* containing the pRSETB plasmid is cultured in the presence of an antibiotic, ampicillin, to suppress the growth of bacteria that do not contain the plasmid, in an attempt to ensure that all bacteria in the culture are capable of producing the desired protein: Citrine. The ampicillin resistance mechanism operates by extracellular secretion of β-lactamase [172]. β-lactamase degrades the β-lactam ring of ampicillin; neutralizing it. Thus, it is possible for the bacteria

possessing the plasmid to produce sufficient enzyme to neutralize all of the ampicillin in the culture vessel, allowing opportunistic bacteria that do not contain the plasmid to flourish. These non-protein producing bacteria can overwhelm the population of protein-producing bacteria and substantially reduce the yield of protein produced by the culture.

To address this problem of antibiotic neutralization and opportunistic growth, the ampicillin resistance gene was substituted for a kanamycin resistance gene [172, 173]. Kanamycin interferes with the 30S subunit of the bacterial ribosome, disrupting protein synthesis. The kanamycin resistance gene encodes an enzyme that degrades kanamycin and is confined to the interior of the bacterium. Thus, in order to survive, each bacterium must contain the gene that encodes for kanamycin resistance, and the Citrine gene that is attached to it.

The second shortcoming of the pRSETB plasmid is a disappointing protein yield. The protein yield from the pRSETB vector was only approximately 10 milligrams of Citrine per liter of bacterial culture. Typically, one might expect several tens of milligrams of over-expressed protein per liter of bacterial culture.

Finally, and most importantly, the affinity tag scheme used in the pRSETB plasmid is undesirably expensive. The Citrine protein encoded by the pRSETB gene contains an affinity tag, composed of six histidine residues, that is used to selectively bind the protein to a nickel column, facilitating rapid protein purification. Before the protein is crystallized, the affinity tag must be removed so that it does not interfere with crystallization contacts. The affinity tag is attached to the protein by a short amino acid sequence: the cut sequence. The cut sequence is recognized by a protease that

cleaves the polypeptide chain at a point in the cut sequence, separating the tag from the desired protein. Unfortunately, the cut sequence encoded in the pRSETB Citrine can only be recognized and cleaved by the proprietary enzyme EKMax (Invitrogen, Carlsbad, CA, USA)

EKMax proved prohibitively expensive, and required an operating temperature of ≈ 37 ˚C, that promoted the aggregation of Citrine. This aggregation reduced the final yield of Citrine from 10 milligrams per liter of culture to less than 1 mg. Given the losses of protein that are incurred in solvent exchange and filtration, this quantity of protein was insufficient for the crystal growth condition screening necessary to produce crystals of sufficient quality for the highly detailed X-Ray structure determination experiments presented in this thesis.

Thus, it was decided to design a new vector to address the shortcomings of the pRSETB plasmid, based around the Novagen (A division of Merck KGaA, Darmstadt, Germany) pET-24 vector. The new vector would incorporate a Citrine gene with a poly-histidine affinity tag, but with a TEV protease cut site rather than an EKMax cut site, and would encode kanamycin rather than ampicillin resistance. The vector would also contain a modified promoter sequence that would yield a higher level of protein expression than pRSETB.

TEV (Tobacco Etch Virus) protease is produced at the Protein Facility, and as such is available at low cost and in large quantities. Unfortunately, the use of a TEV cut site requires that the cleaved Citrine have an additional glycine residue at its N-terminus. Following crystallization of this protein, it was realized that this additional N-terminal glycine may promote the formation of a plate-like form of Citrine crystals. However,

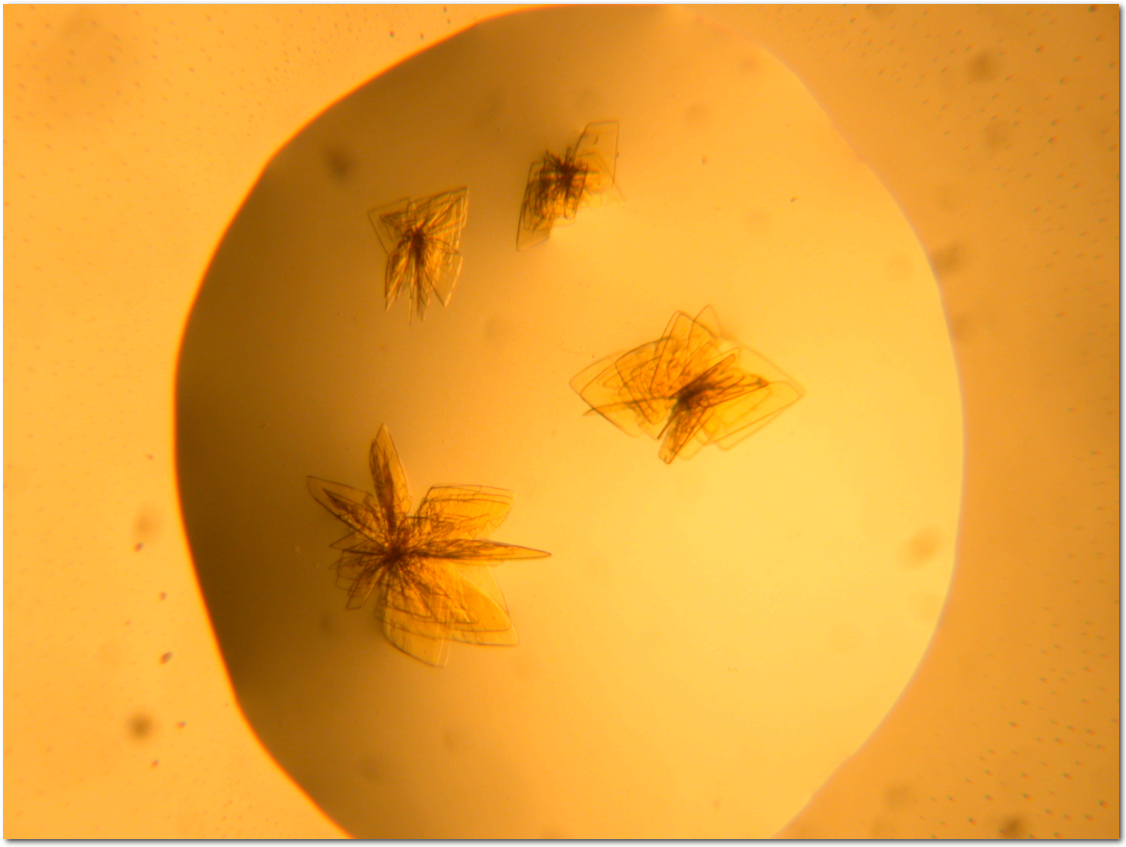


Figure 4.1: Plate-like Citrine crystals grown in 5% PEG 3350, 50 mM sodium acetate, 50 mM ammonium acetate, pH 5.0.

the dramatically improved yield of Citrine from this new growth method more than compensates for any reduction in crystallizability. A photograph of the plate-like form of Citrine crystals is shown in figure 4.1.

The Citrine gene was extracted from the pRSETB plasmid by amplification using the Polymerase Chain Reaction (PCR). By using PCR primers that were complementary to the desired sequence of Citrine, rather than the actual sequence, this step was also used to modify the start of the Citrine gene to incorporate the new protease cut site and add BamHI and NcoI restriction enzyme cut sites to facilitate ligation of the gene into the pET-24 vector. The modified Citrine gene was amplified and then incorporated

into a temporary plasmid with a TOPO Cloning Kit (Invitrogen, Carlsbad, CA, USA) and amplified using MACH1 strain *E. coli*. After growth of the cell culture, the cells were lysed and the plasmid was purified from the cellular debris using a plasmid extraction kit (Plasmid Kit, Qiagen. Qiagen sells a variety of plasmid extraction kits suitable for different purification volumes. Qiagen is a multi-national corporation headquartered in Hilden, Germany; Germantown, MD, USA and Venlo, The Netherlands). Incorporation of the Citrine gene into the plasmid was verified by BamHI and NcoI restriction enzyme digest of the plasmid, followed by agarose gel electrophoresis. The digested gene was extracted from the agarose gel with a gel extraction kit (Qiagen).

The extracted Citrine gene was then ligated into a pET-24 vector. The pET-24 vector was then used to transform non-mutating TOP10F′ *E. coli*, and grown in small-scale culture to amplify the plasmid. The pET-24 vector was extracted and sequenced to confirm that the Citrine gene had not been mutated. The pET-24 vector was then purified from the TOP10F′ strain, and was used to transform BL21 strain *E. coli* for large-scale production of Citrine. A small culture of BL21 carrying the Citrine gene was produced, and the cells were pelleted, mixed with glycerol and frozen at -80˚ C for later use.

4.2.2 Over-expression of Citrine

Small quantities of cells were removed from the frozen BL21 cell cultures with a sterile toothpick and used to inoculate several 5 mL cultures of LB (Luria-Bertani) broth. These cultures were grown overnight at 37˚ C in LB broth with a kanamycin concentration of 100 mg/L. The following morning these cultures were each used to

inoculate a culture flask containing 2L of LB media with 100 mg/L kanamycin. In total, six flasks were inoculated, giving twelve liters of cell culture.

The culture flasks were shaken in a rotary shaker at 37 ˚C. The optical density (OD) of the cell culture was monitored every 30 minutes. When the cell culture had reached an optical density of 0.6, protein over-expression was induced by adding IPTG (Isopropyl β-D-1-thiogalactopyranoside; IPTG is a molecular mimic of allactose, a lactose metabolite that induces the *lac* operon. The promoter for the Citrine gene is IPTG-inducible.) to a final concentration of 100 mM. Protein over-expression was allowed to proceed for 6 hours. The cell cultures were then transferred to 1L centrifuge tubes and spun at 14,000 × *g* for approximately 30 minutes to pellet the cells. The cell pellets were frozen in a -80 ˚C freezer for overnight storage. The following day, the cells were thawed and were resuspended in a high-salt alkaline lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0, 0.22 μm filtered) with a hand-held homogenizer. Each liter of culture produced a cell pellet of approximately 30 mL in volume. The cell pellet was very dense and had the consistency and feel of clay. Normally, cell pellets that have expressed non-colored proteins will have a beige, clay-like color. However, the cell pellets containing over-expressed Citrine had a distinctive algal-green color.

The resuspended cells were then lysed in a French Press style cell disruptor (Constant Systems, Low March, Daventry, Northants, United Kingdom). The lysate was then spun at 20,000 × *g* at 4 ˚C for approximately 30 minutes to pellet the cellular debris and separate it from the cytosolic proteins, including Citrine. The supernatant had the consistency, and because of the presence of Citrine, color of motor oil due to the presence of uncoiled genomic DNA that had been liberated from the lysed cells. The

uncoiled genomic DNA was precipitated from the supernatant by slowing stirring in polyethylenimine to a final concentration of 0.5% v/v. The supernatant volume at this step was several hundred milliliters.

After precipitation of the genomic DNA, the Citrine was purified by affinity chromatography using an FPLC (Fast Protein Liquid Chromatography) apparatus (ÄKTA FPLC, GE Healthcare, Uppsala, Sweden) with a Nickel-NTA column (HisTrap HP, Catalog Number 17-5248-01, GE Healthcare, Uppsala, Sweden). Affinity chromatography often renders the protein 95% pure in a single step.

Prior to use, the Nickel-NTA column was equilibrated with several column volumes (5 mL) of lysis buffer. The supernatant was loaded onto the Ni-NTA column with a suction pump. As a precautionary measure, the flow through was collected. The protein concentration in the flow through was monitored by a UV absorption spectrophotometer incorporated into the exit tubing of the FPLC. Binding of the Citrine protein to the Ni-NTA column could be confirmed visually due to its bright color.

Contaminants that were weakly bound to the column were removed by washing the column with a *wash buffer* with a low imidazole concentration (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0). The imidazole competes for binding sites in the Ni-NTA column, and will displace weakly bound contaminants. The FPLC is capable of mixing the wash buffer and *elution buffer* with a much higher concentration of imidazole (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0) in a user-selected ratio to produce a solution with a variable imidazole concentration. The imidazole concentration passing through the Ni-NTA column was slowly raised and

the protein elution from the column was simultaneously monitored to ensure complete removal of contaminants. Finally, the imidazole concentration was raised to ≈ 250 mM to elute the Citrine bound to the column. Fortunately, Citrine binds strongly to the column and is highly colored, so it easy to verify that the protein remains bound to the column during the wash process.

After purification by affinity chromatography, the poly-histidine affinity tag on the Citrine molecules was removed by TEV protease digest. The Citrine solution was dialyzed overnight into a TEV protease reaction buffer (50 mM Tris-HCl, 0.5 mM EDTA, 1mM DTT, pH 8.0) in dialysis tubing with a 10 kDa molecular weight cutoff (Snakeskin Dialysis Tubing, Catalog number 68100, Pierce Biotech, Rockford, IL, USA). Following overnight dialysis, the TEV protease was added to the dialysis tubing. Typically, the TEV concentration used is approximately $1/100^{th}$ of the concentration of the protein to be digested. The cleavage reaction, or *digestion*, was allowed to proceed for approximately 24 hours at 4˚ C.

The cleaved Citrine was separated from the cleaved his-tags, TEV protease (which is itself his-tagged) and uncleaved Citrine passing it through a second Ni-NTA column. This step binds the uncleaved Citrine, his-tags and TEV protease to the column, while the cleaved Citrine flows through. As a final check, the cleaved Citrine was passed through a gel filtration column to ensure its purity, and was then filtered using a Whatman 0.2 μm Anotop 10 Plus syringe filter (Whatman PLC, Maidstone, Kent, United Kingdom) to remove any bacterial contamination.

Following the final purification step the protein was concentrated to 20 mg/mL with a Pall centrifugal concentrator (Catalog number OD010C37, Pall Separation Systems,

East Hills, NY, USA), and exchanged into a *storage buffer* (50 mM HEPES, pH 7.5). The protein concentration was verified by measuring the absorption of the protein at 514 nm with a Nanodrop ND-1000 absorption spectrophotometer (Nanodrop, Wilmington, DE, USA). The absorption coefficient of Citrine at 512 nm is ≈ 80,000 M^{-1} cm^{-1} [170].

As Citrine does not cold denature, large aliquots of the protein may be slow frozen and thawed. 50 µL aliquots of Citrine solution were frozen at -70 ˚C for long term storage. Citrine appears to keep quite well at 4 ˚C for at least two years. The pET-24 vector yielded approximately 80 mg per liter of culture, giving approximately 1 gram of Citrine.

4.3 Redesign of High-Pressure Cryo-cooling Apparatus

4.3.1 Energy Stored in High-Pressure Cryocooling Apparatus

Prior to the collection of the final set of structural and spectroscopic data on Citrine under high-pressure cryocooling conditions, it was realized that the high-pressure cryocooling apparatus may present a safety hazard to its operator. A simple model of a worst-case scenario explosion inside the high-pressure cryocooling safety enclosure was developed to assess the adequacy of existing enclosure, and to specify the design of a new safety enclosure.

As helium is close in behavior to an ideal gas, the energy released in an adiabatic expansion may be estimated with the ideal gas equation of state. The energy released from the expansion of a gas, W, may be calculated by integrating the product of the pressure and volume change from the initial high pressure, P_0 to a final ambient pressure, P_f

$$W = \int_{V_0}^{V_f} dV\, P\,. \tag{4.1}$$

For an adiabatic expansion of an ideal gas;

$$PV^{\gamma} = \text{constant} = k\,. \tag{4.2}$$

Thus,

$$PV^{\gamma} = P_0^{\gamma} V_0^{\gamma}, \tag{4.3}$$

$$P = P_0^{\gamma} V_0^{\gamma} V^{-\gamma}. \tag{4.4}$$

Thus,

$$W = \int_{V_0}^{V_f} dV\, P_0^{\gamma} V_0^{\gamma} V^{-\gamma}\,, \tag{4.5}$$

$$W = \frac{1}{\gamma - 1}\left[P_0 V_0 - P_0 V_0 V_f^{-\gamma+1}\right]. \tag{4.6}$$

V may be found by appealing to the constancy of PV^{γ}

$$V_f^{\gamma} = P_0 V_0^{\gamma} P_f^{-1}\,. \tag{4.7}$$

For an ideal gas $\gamma = 5/3$. The internal volume of the high-pressure cryocooling apparatus is ≈ 15 mL, and the maximum working pressure is either 200 or 400 MPa. The energy liberated, W, in an adiabatic expansion is 4500 J from 200 MPa and 9000 J from 400 MPa.

For illustrative purposes, if these energies were directed into a single steel fragment with a volume of 1 cm^3, and a mass of approximately 8 grams (the density of steel is approximately 8 gram cm^3) the corresponding velocities would be 1061 ms^{-1} for 200 MPa and 1500 ms^{-1} for 400 MPa.

These velocities, while high compared to the speed of sound in air (330 ms^{-1}), are comparable to the velocity of rifle bullets; (the mass of a bullet without its cartridge is ≈ 5 g), so these velocities may be realistically taken as very high upper limits on the speed of ejecta from a ruptured high-pressure cryocooling system. If, on the other hand, all of the energy of the explosion, were channeled into the original high-pressure cryocooling system enclosure (often referred to as "Frankenstein"), with a mass of approximately 200 Kg, it would travel at a speed of 9.5 ms^{-1} (21.3 mph) for a working pressure of 400 MPa, and 6.7 ms^{-1} (15 mph) for a working pressure of 200 MPa.

For comparison, the energy liberated by the rupture of a compressed gas cylinder was also calculated. Although the pressure inside a gas cylinder is lower: 4000 psi or 28 MPa, the volume is much higher ≈ 100 L. The liberated energy is 4.2×10^6 J, 3 orders of magnitude higher than the energy stored in the high pressure cooling system.

The energy released by a van der Waals (vdW) gas was also calculated numerically for the case of helium. It was found that at a starting pressure of 400 MPa, the energy liberated $W = 8.16 \times 10^3$ J, slightly lower than the ideal gas case: 9×10^3 J. However, at 200 MPa, the liberated energy is higher than the ideal gas case: $W = 6.5 \times 10^3$ J versus 4.5×10^3 J.

Table 4.1: Calculated penetration depths and minimum barrier thicknesses of cubic, 1 cm^3 steel projectiles. For steel, the penetration constant, K was taken at the maximum listed value 0.4×10^{-3} ft^3 lb^{-1}. For concrete, K was taken to be 8×10^{-3} ft^3 lb^{-1}.

Barrier Material	Projectile Velocity (ms^{-1})	Penetration Depth (“)	Minimum Barrier (“)
Steel	330	0.064	0.191
Steel	1275	0.151	0.452
Steel	1428	0.158	0.475
Concrete	330	1.272	3.816
Concrete	1275	3.01	9.03
Concrete	1428	3.163	9.488

4.3.2 Calculation of Penetration Depth of Projectiles

An empirical formula, proposed by Petry [174], was used to calculate the penetration depth of projectiles into infinitely thick steel and concrete barriers, given a projectile velocity, size and mass. The penetration depth d in feet, is approximated by

$$d = K\frac{m}{A}\log_{10}\left(1+\frac{v^2}{215000}\right) \tag{4.8}$$

In this equation, the mass m, is entered in pounds, the maximum cross sectional area A, is entered in ft^2, and the empirical constant K is entered in ft^3 lb^{-1}. For steel K varies between 0.26×10^{-3} and 0.4×10^{-3}. For concrete, $K = 8 \times 10^{-3}$. In order to produce a reliable estimate of the penetration depth, it has been found that the barrier thickness must be at least 3 penetration depths. If this criterion is not adhered to, the penetration depth will be longer. The penetration depths into steel and concrete targets were calculated for projectiles traveling various velocities are shown in table 4.1.

4.3.3 Effect of Explosive Wave on Shielding

In addition to projectiles ejected by a rupture in the pressure system, another hazard comes from the expanding gas. Professor Arthur Ruoff proposed a calculation, based upon a model by Timoshenko [175], for the strain on the edges of a circular plate, clamped at the edges, from a load applied to the center of the plate. Centrally loading the plate will produce a higher maximum stress, experienced at the edges of the plate, than uniformly loading the plate, so this calculation can be considered an upper limit on the stress produced by the expanding gas from a rupture of the high-pressure cryocooling apparatus.

The energy released from the rupture from the pressure system is assumed to be deposited into the plate as strain energy. From Timoshenko [176], the strain energy, U deposited into a plate, by a point force F, causing a central deflection δ

$$U = F\delta/2\,. \tag{4.9}$$

The central deflection of the plate, of radius a, with a flexural rigidity D

$$\delta = \frac{Fa^2}{32\pi D}. \tag{4.10}$$

The flexural rigidity D, is defined by Timoshenko in equation 74 of reference [175] as a function of the Young's modulus E, Poisson ratio μ and thickness, h of the plate;

$$D = \frac{Eh^3}{12\left(1-\mu^2\right)}. \tag{4.11}$$

Thus, the force applied to the plate

$$F = \frac{4}{a}\sqrt{2\pi DU}\,. \tag{4.12}$$

The maximum stress occurs at the clamped edges of the plate, and is given by Timoshenko in equation 106 of reference [175],

$$s = \frac{3F}{2\pi h^2}. \tag{4.13}$$

As a function of the strain energy

$$s = \frac{6}{ah^2}\sqrt{\frac{2DU}{\pi}}\,. \tag{4.14}$$

As the stress is inversely proportional to the radius of the circular plate, it is prudent to consider the stress on a circle that inscribes the smallest rectangular plate on the high pressure cooling system. The smallest plate is the side plate, with a dimension of 26 × 14 inches (0.66 × 0.36 m). This gives the radius of the smallest inscribed circle as 0.18 m. The thickness of the steel shielding around the high pressure cooling system is 0.5 inches, its Young's modulus, $E = 2 \times 10^{11}$ Pa, and its Poisson ratio $\mu = 0.28$.

The maximum stresses for this circular plate were calculated for four cases; where the liberated energy from a rupture at 200 and 400 MPa was projected onto a single plate, and where it was shared between the 6 faces of a cube. The results are tabulated in table 4.2. All of the maximum stresses listed in these scenarios exceed the elastic limit of hard steel, of 0.2 GPa, and that of more ductile 1020 steel, with an elastic limit of 0.35 GPa.

Table 4.2: Maximum stresses and forces exerted on a circular plate inscribing the smallest plate of the shielding surrounding the high-pressure cryocooling system by an explosive wave from a rupture at 200 and 400 MPa, and scenarios where the energy is incident on only one plate, or distributed amongst all 6 shielding plates.

Scenario	Force on Plate (10^6 N)	Maximum Stress (GPa)
400 MPa, vdW Gas, 1 Plate	0.97	2.87
400 MPa, vdW Gas, 6 Plates	0.4	1.17
200 MPa, vdW Gas, 1 Plate	0.86	2.56
200 MPa, vdW Gas, 6 Plates	0.35	1.04

However, while these calculations indicate that the applied force exceeds the elastic yield stress of the shielding plate, it was derived in the elastic limit, and so will not be an accurate predictor of the maximum stress on the plate. Another way to calculate the capacity of the plates to absorb energy is to consider the energy absorption through plastic deformation. Examination of the stress-strain curve for steel indicates that the plastic deformation of steel can absorb considerably more energy than the elastic deformation. At the elastic limit of 0.2 GPa, the strain of hard steel, $\varepsilon = 0.001$, given its Young's modulus, $E = 200$ GPa.

The strain of ductile 1020 steel, with a higher elastic limit $Y = 0.35$ GPa is 0.00175, given that it has the same Young's modulus as hard steel. The energy stored by elastic deformation in a volume V, which for the smallest rectangular plate in the system is $14 \times 26 \times 0.5$ inch^3 $= 3.0 \times 10^{-3}$ m^3 is

$$E_{\text{Elastic}} = Y\varepsilon V/2, \tag{4.15}$$

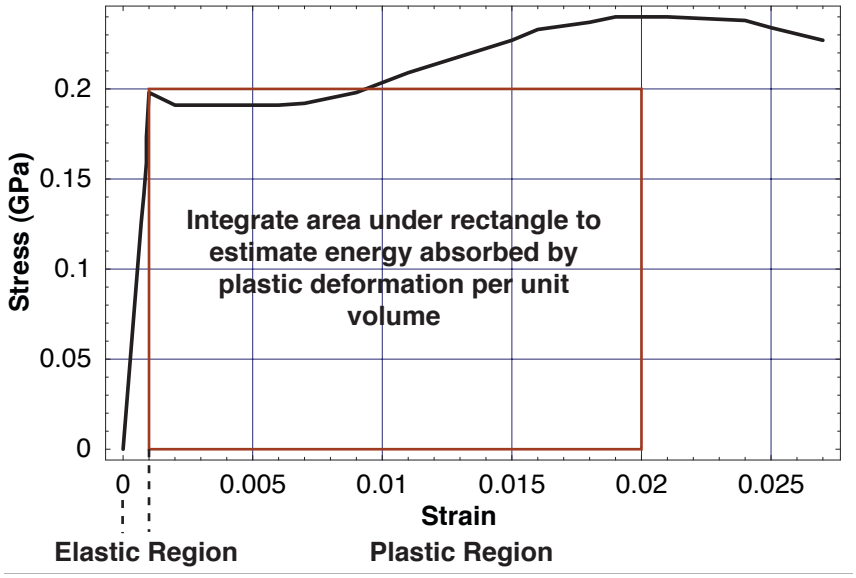


Figure 4.2: Stress-strain curve for steel showing integration area for estimate of energy absorption by plastic deformation.

$$E_{\text{Elastic. Hard}} = 298 \text{ J}, \tag{4.16}$$

$$E_{\text{Elastic. 1020}} = 913 \text{ J}. \tag{4.17}$$

However, the strain at failure of hard steel is 10 to 20 times the strain at the elastic limit, ε = 0.01 - 0.02, so the energy absorbed by plastic deformation can be much greater. The strain at failure of 1020 steel is even greater, $\varepsilon \approx$ 0.15 - 0.20. The energy absorbed can be estimated by integrating a rectangle under the stress-strain curve for steel, from the elastic strain to the ultimate strain, and capping the stress at the elastic stress. The area for this approximate integration is shown schematically in figure 4.2.

Thus, the energy absorbed by plastic deformation of the plate under consideration, with an elastic limit, *Y*, and assuming a strain at rupture of 0.01 rather than 0.02, is;

$$E_{\text{Plastic}} = Y\varepsilon V, \tag{4.18}$$

$$E_{\text{Plastic. Hard}} = 5965\,\text{J}, \tag{4.19}$$

$$E_{\text{Plastic. 1020}} = 156578\,\text{J}. \tag{4.20}$$

For comparison, assuming that the ultimate strain of concrete is 0.001, and that its elastic yield stress is 2 MPa, the energy absorbing capacity of a 14×26 inch$^2 \times 1$ ft concrete cube is only 143 J.

4.3.4 Conclusions

These calculations highlighted a number of shortcomings in the design of the original safety enclosure surrounding the high-pressure cryocooling apparatus. Firstly, the original safety shield featured an open face on one side to allow high-pressure tubing to access the cryocooling apparatus from a nearby high-pressure pump. This open face also allowed a string into the apparatus that was used to pull three magnets away from the high-pressure apparatus, allowing the samples to drop from the top of the cryocooling apparatus to the bottom where they would be frozen under high pressure. It was assumed that this arrangement was safe as the apparatus was isolated from the operator in a concrete-walled room. However, this arrangement would offer no protection to valuable equipment in the room such as the high-pressure pump, nor would it prevent acceleration of the enclosure by the escaping gas.

Additionally, these calculations indicate that the maximum stresses on the safety enclosure that would be experienced in a worst-case scenario explosion would occur at the edges of the faces of the enclosure. In the original safety enclosure design, the

angle brackets used to fix the walls together are very thin, and are attached to the heavy shielding plates by thin machine screws. Finally, the door to the high-pressure safety enclosure through which the operator accesses the apparatus is hinged and is secured by a thin latch. This door would almost certainly open in the event of a rupture of the high-pressure system. It was decided that the high-pressure cryocooling enclosure suffered from at least four serious deficiencies. A new design was constructed that addressed these deficiencies.

Firstly, these calculations indicate that the concrete walls of the room are of sufficient thickness to prevent the penetration of a single steel object, of mass 8 grams, that receives all of the energy stored inside the high pressure cooling system. However, it was decided that a new high-pressure cryocooling enclosure would attempt to prevent the release of any debris from an internal explosion. For a working pressure of 400 MPa, a minimum wall thickness of 0.475” (table 4.1) is required to prevent the escape of a single explosively propelled fragment.

Similarly, calculations of the energy absorbing capacity of ductile 1020 steel by plastic deformation indicate that a 0.5” wall thickness would be adequate to absorb the energy of a pressure wave incident on one of the faces of the enclosure. A schematic of the enclosure by Martin Novak is shown in figures 4.3 and 4.4.

It was decided that the angle bracket used to fix the walls of the enclosure together should also be 0.5” thick. It was decided that it would be imprudent to weld the faces of the apparatus together, as deficiencies in the weld joints on a single side may result in the failure of that joint.

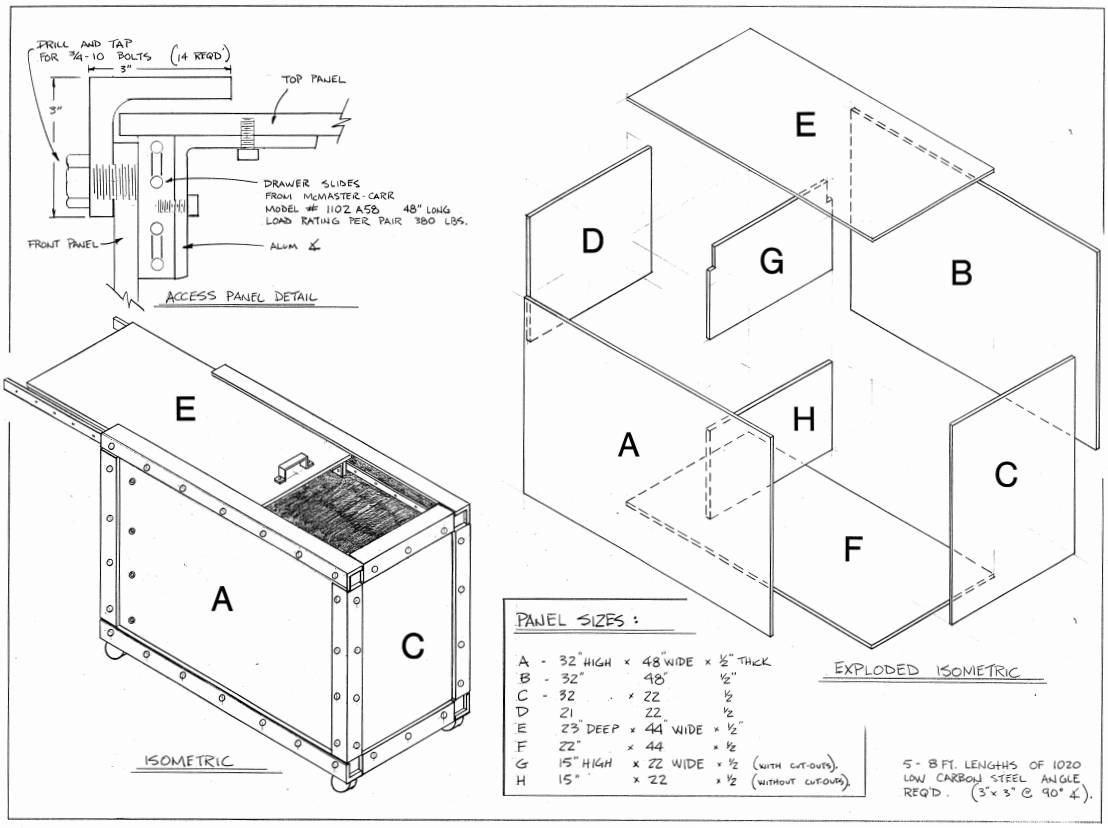


Figure 4.3: Isometric view and expanded isometric schematic of new high-pressure cryocooling apparatus safety enclosure. Design and illustration by Martin Novak.

For this reason, it was decided that nuts and bolts should be used to affix faces of the enclosure together, as a small amount of looseness in the screws would allow even redistribution of the strain due to an explosion across the edges of a face. From table 4.2, the maximum force, $F_{\max}$, that will be experienced by a single plate would be 0.97×10^6 N. This force would be distributed across N bolts, each with an area A. Thus, the stress on the bolts will be

$$s = \frac{F_{\max}}{AN} \tag{4.21}$$

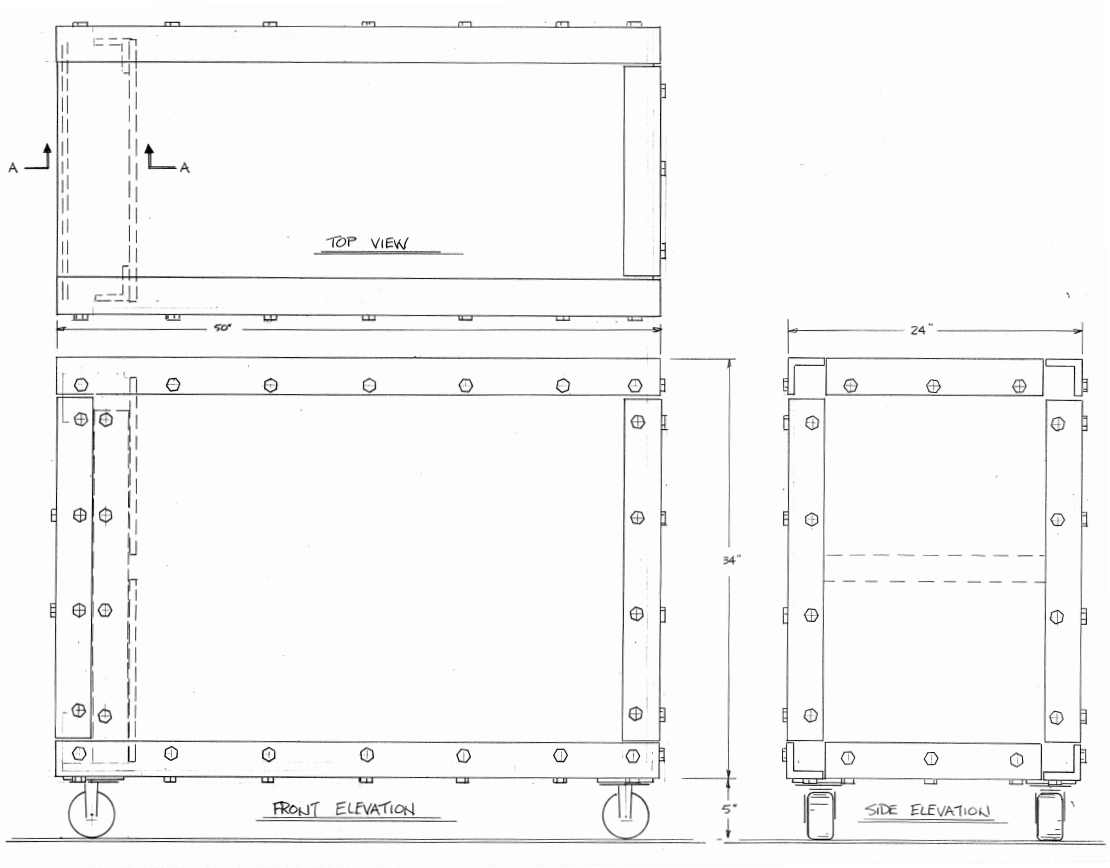
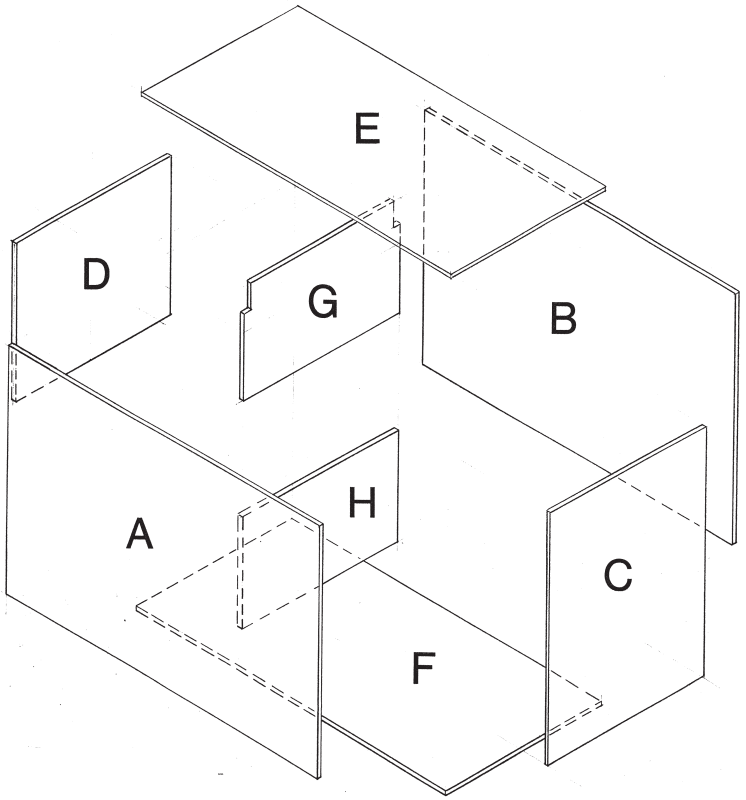


Figure 4.4: Orthogonal view of the new high-pressure cryocooling safety enclosure. Design and illustration by Martin Novak.

A grade 8 bolt has a proof strength of 0.83 GPa and a minor diameter of 0.401" (0.5" major diameter), with a corresponding area of 8.15×10^{-5} m^2. In order for the strain on the bolts to be less than 0.83 GPa, at least 14 bolts should be used per face. A schematic of the face-plates and instructions for the placement of drill or punch holes is shown in figure 4.5. A schematic of the angle brackets and instructions for the placement of holes is shown in figure 4.6.

Figure 4.5: Expanded view of plates of high-pressure cryocooling apparatus, and instructions for drilling holes. Illustration by Martin Novak, hole specifications by Buz Barstow.

Exploded Isometric View of Plates



Instructions for Drilling Holes in Angles and Plates

Through Holes in All Plates Except A and B: All plates should contain through holes for 3/4”-10 screws, except for plates A and B. The diameter for the through holes should be as close to 49/64” as possible to allow the plates to retain maximum strength.

Tapped Holes in Plates A and B: The holes on the top edges of plates A and B will be tapped for 3/4”-10 screws after delivery. Please cut 21/32” holes here.

All other holes on plates A and B should be through holes for 3/4”-10 screws with a diameter of 49/64”.

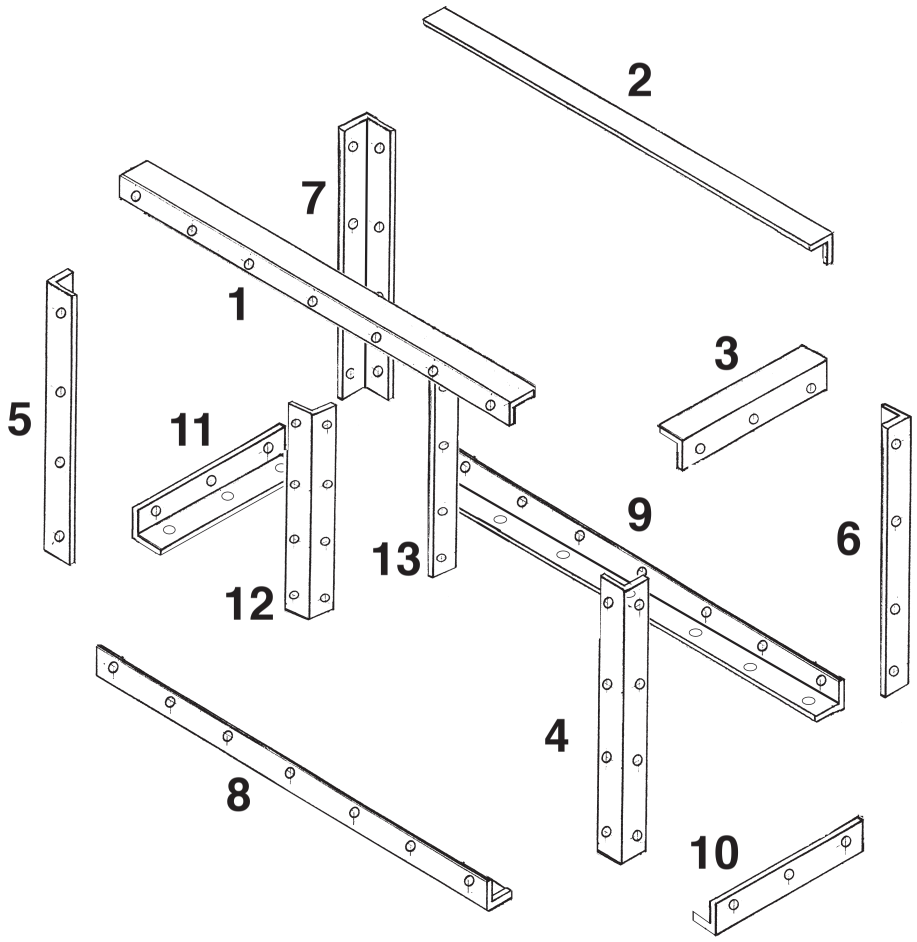
Positioning of Holes Along Length of Edge: The holes should be evenly distributed along the length of each edge.

Positioning of Holes From Edge of Plates: The holes should be positioned so that the edge of each plate can come as close as possible to the radius of the angle.

Procedure for Drilling Holes: When drilling holes, please ensure that a hole through a plate, and the corresponding hole in an angle are as closely matched as possible. If possible, please drill or punch these holes at the same time.

Figure 4.6: Expanded isometric view of angle brackets with instructions for holes in high-pressure cryocooling apparatus. Illustration by Martin Novak, hole specifications by Buz Barstow.

Exploded Isometric View of Angles



List of Angles

Angle Material: 1020 Steel
Angle Dimensions: 3” x 3” x 1/2”

All angles should contain through holes for 3/4”-10 screws. The hole diameter should be as close to 49/64” as possible to allow the angle to retain maximum strength.

1 - 50” long. 7 holes.
2 - 50” long. 7 holes.
3 - 18” long. 3 holes.
4 - 28” long. 8 holes.
5 - 28” long. 8 holes.
6 - 28” long. 8 holes.
7 - 28” long. 8 holes.
8 - 50” long. 7 holes.
9 - 50” long. 7 holes.
10 - 18” long. 6 holes.
11 - 18” long. 6 holes.
12 - 28.5” long. 8 holes.
13 - 28.5” long. 8 holes.

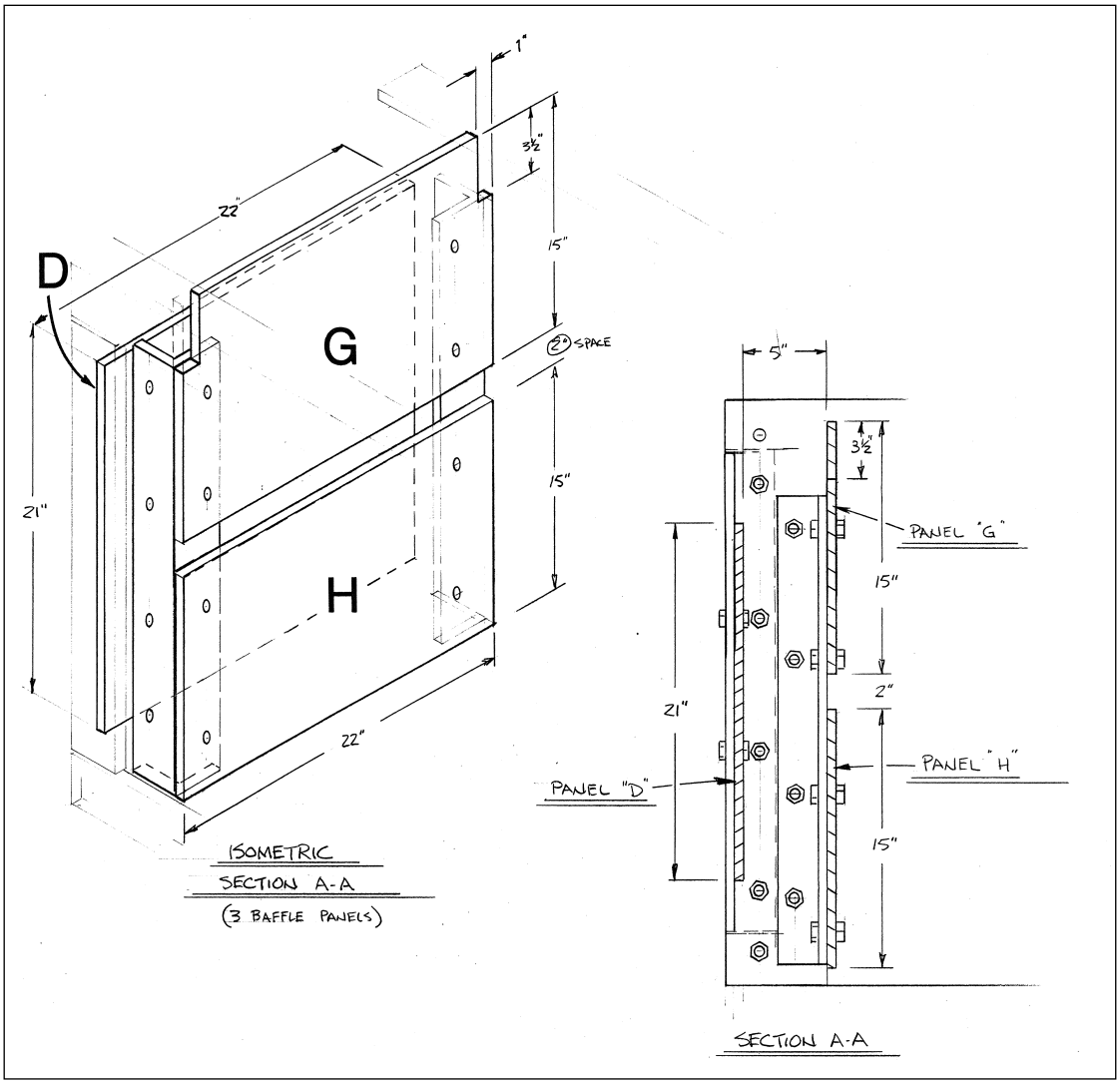


Figure 4.7: Expanded isometric view of gas exit port. Plates G and H face towards the interior of the high-pressure cryocooling enclosure. Gas may escape from the interior through the aperture between plates G and H, and is directed up and down by plate D, which also serves to capture any debris accelerated by an explosion in the interior of the enclosure. Illustration by Martin Novak.

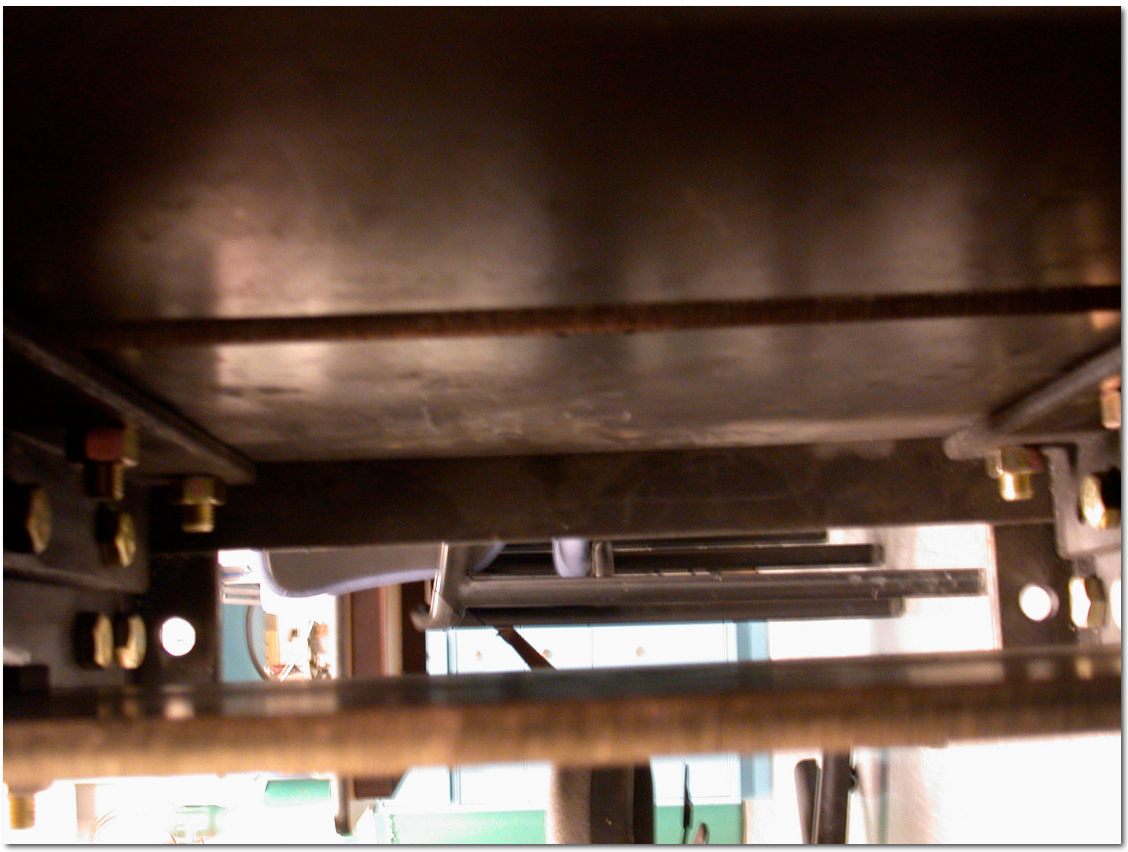


Figure 4.8: View of the gas exit port from the outside of the high-pressure cryocooling safety enclosure. Plate D (figure 4.7) is at the bottom of the photograph, and plates G and H are at the top of photograph.

The calculations also indicate the need to protect against an errant high pressure cooling system. Although it seems unlikely that the high pressure cooling system could be accelerated to 9.5 mph, the danger associated with this outcome was considered sufficiently high that precautions be taken against it. Originally, it was proposed that energy absorbing material be placed around the apparatus. However, this material would pose an inconvenience for the operator, and Professor Sol Gruner proposed that the enclosure should contain a gas exit port direct escaping gas up and down, resulting in zero net thrust from a gas release. A schematic of the gas exit port is shown in figure 4.7 and photographs are shown in figures 4.8 and 4.9.

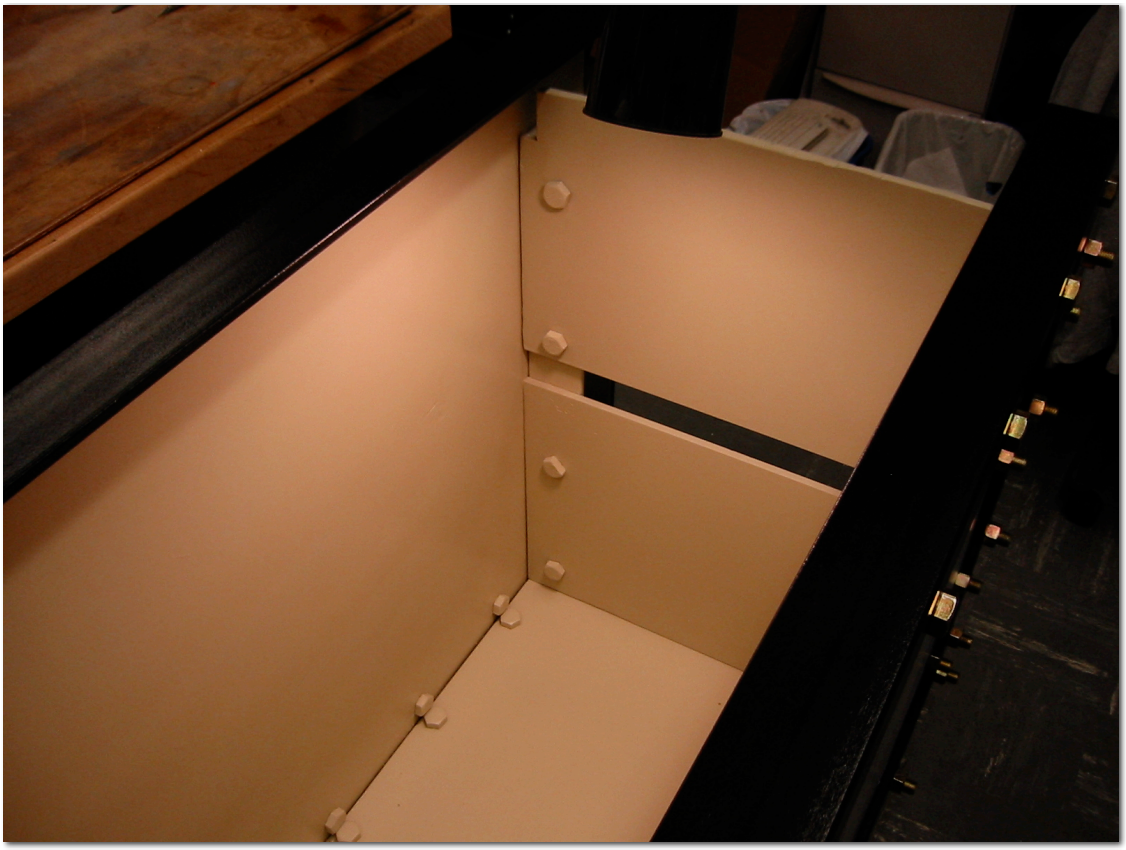


Figure 4.9: View of gas exit port (plates G and H in figure 4.7) from inside the high-pressure cryocooling safety enclosure.

Finally, to prevent opening of the door to the safety enclosure in the event of a gas release, the apparatus would feature a sliding door. The door would be secured by angle brackets, and would slide along an internal rail. The angle brackets will clamp the two parallel edges of the door, and prevent it from escaping.

These calculations illustrate the dramatic changes in energy absorption capacity that different types of steel may have. A small 0.5” thick hard steel shielding plate considered in these calculations can only absorb ≈ 6000 J of energy from an explosive wave prior to failure. This is less than the energy released from rupture of high pressure cooling system operating at 400 atmospheres: 8160 J. However, if this steel is

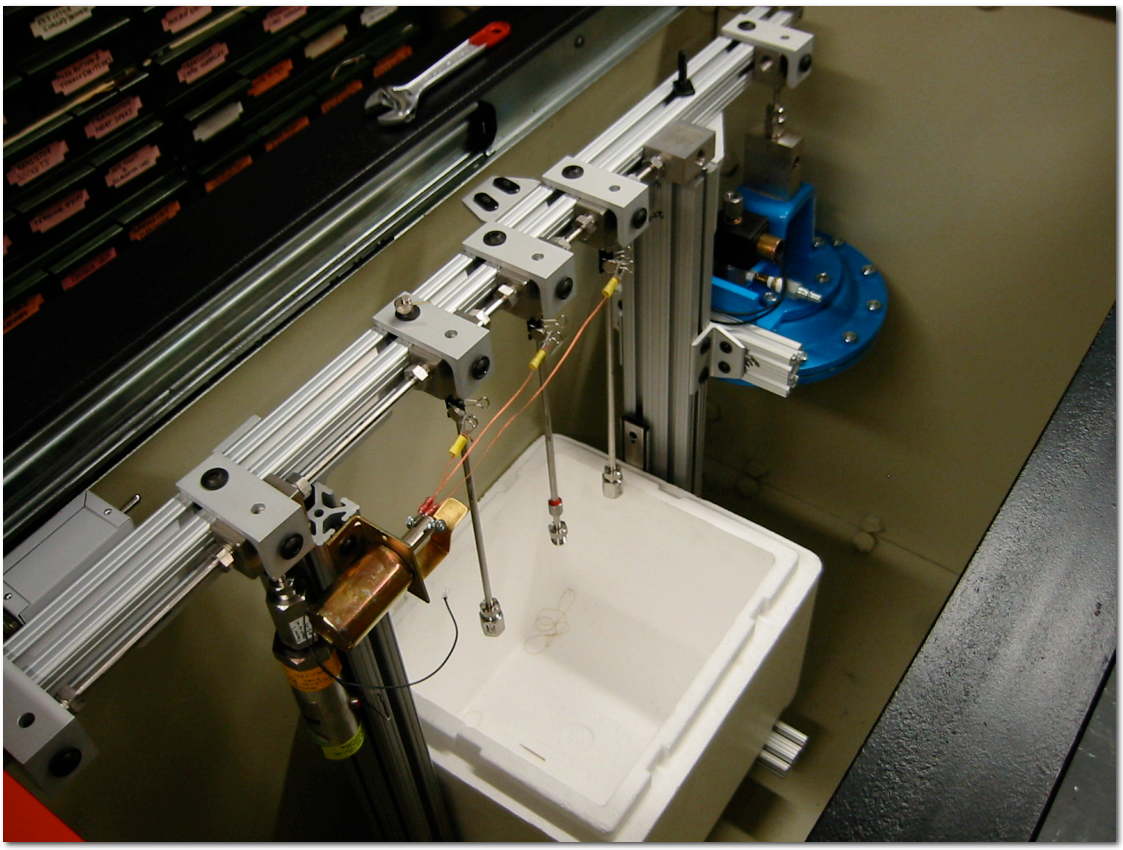


Figure 4.10: Electrically actuated sample release mechanism in the high-pressure cryocooling system. The blue oblate in the top right of the photograph is a pneumatically actuated valve.

replaced with more ductile steel, for instance 1020 or 835, it can absorb ≈ 160,000 J prior to failure.

It should be noted that if the high-pressure cryocooling system were ever to fail, the shielding surround it should be replaced, as the ability of even ductile steel to absorb energy elastically is less than 1000 J. It is also important to note, that an estimate of the energy absorbing capacity of concrete indicates that it could not absorb all of the energy from a rupture of the high-pressure system prior to failure.

The new apparatus was custom fabricated and assembled by AccuFab (Ithaca, NY, USA). Although the construction material was originally specified as 1020 steel, high steel prices at the time of construction necessitated the use of a cheaper equivalent: 835 steel. Calculations indicate that 835 steel has similar energy absorbing properties to 1020 steel. Martin Novak at Cornell University performed final assembly of the apparatus, including installation of the sliding door and painting.

In order to reduce the number of openings to the exterior from the high-pressure cryocooling enclosure, the magnets that hold the protein samples in the high-pressure tubing above the cold reservoir would no longer be removed by strings, but would be electrically removed. A photograph of the electrically actuated sample release system is shown in figure 4.10. The electrical sample release system was designed by Yi-Fan Chen and Mark Tate. Additionally, the screw actuated high-pressure valves in the original high-pressure cryocooling apparatus were replaced with pneumatically actuated valves (the blue oblates in figure 4.10).

The pneumatically actuated valves require only the entry of a pneumatic line into the high-pressure enclosure, rather than an electrically driven screw. The high-pressure cryocooling apparatus in its final assembled state is shown in figure 4.11.

Figure 4.11: High-pressure cryocooling apparatus in its final assembled state.

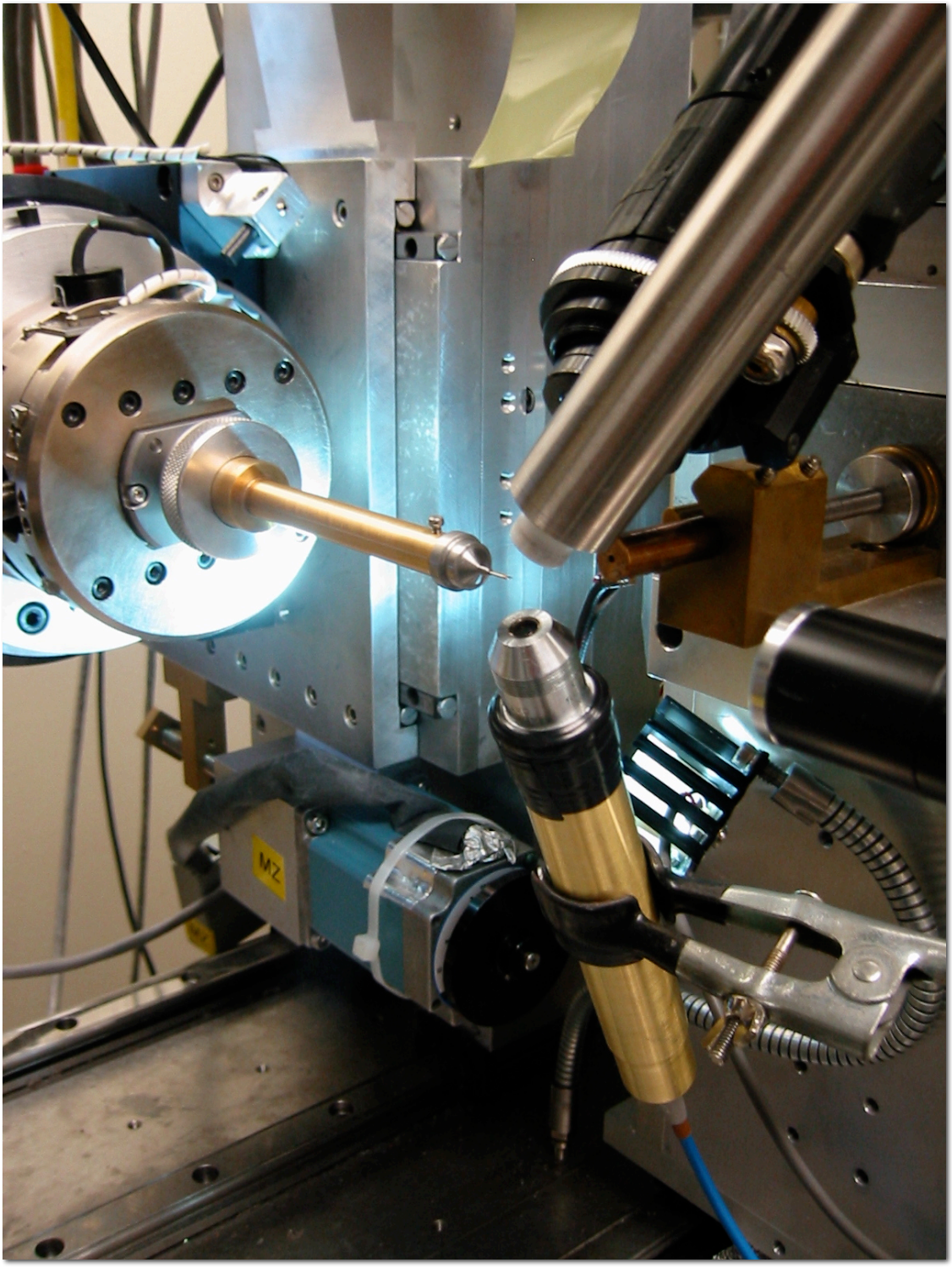


Figure 4.12: First generation crystal fluorescence monitor at CHESS station F2.

4.4 Micro-spectrophotometer for Low Temperature Measurement of Citrine's Fluorescence Spectrum

4.4.1 Introduction

As the thermal contraction of a protein molecule between room temperature and 100 K is comparable to the changes due to pressurization to several hundred MPa [177], it was important to measure the fluorescence spectra of high-pressure cryocooled samples of Citrine to establish a direct link between structure of the molecule and its fluorescence properties.

4.4.2 First and Second Generation Micro-spectrophotometers

During the course of this thesis work, we used three micro-spectrophotometers to measure the fluorescence spectra of high-pressure cryocooled Citrine samples. The first micro-spectrophotometer was constructed at CHESS (Cornell High Energy Synchrotron Source) station F2. Photographs of this apparatus are shown in figures 4.12 and 4.13. This apparatus was built at station F2 due to the availability of a well-maintained cryogenic nitrogen stream. The apparatus was used to measure the fluorescence spectra of high-pressure cryocooled Citrine crystals, in order to find a correlation between the freezing pressure and the peak of the fluorescence spectrum.

In the first generation apparatus, fluorescence excitation light was supplied to a Citrine crystal through a microscope objective, coupled by an optical fiber to a portable xenon lamp (PX-2 Pulsed Xenon Light Source, Ocean Optics, Dunedin, FL, USA). The spectral width of the excitation lamp was narrowed with an optical filter. The fluorescence light was collected through a telescope and passed through an optical fiber to a USB2000 spectrometer (Ocean Optics, Dunedin, FL, USA). Alignment of the excitation optics was performed by hand, by moving a clamp that held the

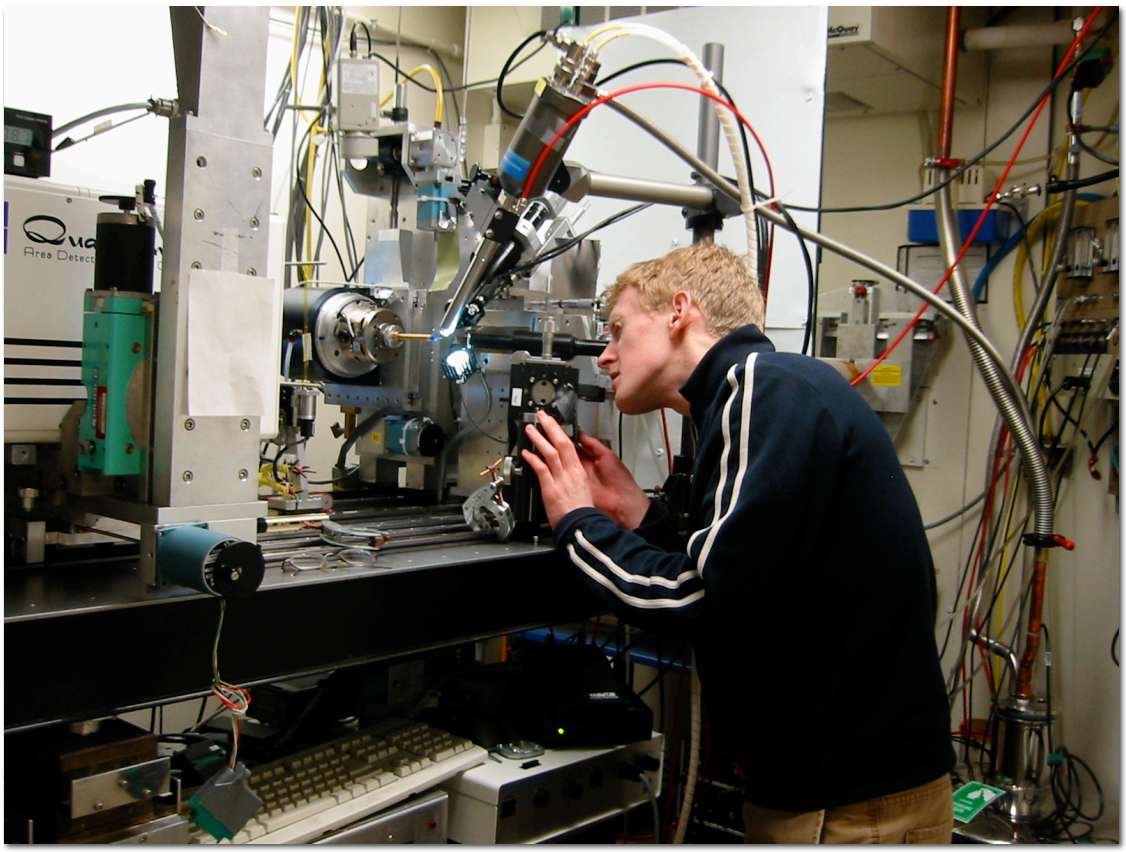


Figure 4.13: First generation micro-spectrophotometer at CHESS station F2. The author is aligning the fluorescence collection optics on the sample, which also served as a fluorescence collection optic.

excitation source microscope objective. This was unwieldy, but proved acceptable. The fluorescence excitation optics are shown in figure 4.12. Alignment of the fluorescence collection optics was considerably more difficult. The fluorescence collection telescope was aligned on the crystal by eye, by swapping the optical fiber coupler at the rear of the fluorescence collection telescope for a telescope eyepiece lens. Once the crystal was aligned in the telescope, the eyepiece lens was swapped for the fiber optic couple. The author is seen aligning the telescope on a crystal in figure 4.13.

Although this apparatus was unwieldy, it did highlight some of the complications that would arise when attempting to measure the spectra of high-pressure cryocooled samples of Citrine. Firstly, this apparatus highlighted the need for good separation of the excitation light and the fluorescence spectrum of the Citrine sample. In the case of the first generation micro-spectrophotometer, the excitation light was monochromated with a filter (Linear Variable Filter, Ocean Optics, Dunedin, FL, USA), and still had a considerable spectral width after passing through the filter. Scattering of the excitation light impinged on the fluorescence spectrum of the crystal, and considerably reduced our confidence in our ability to accurately measure the peak of the spectrum. Secondly, the stability of the optics in this first generation apparatus were poor, making alignment difficult, and sample changes long. The lack of easy adjustability of the optics also contributed to the length of the sample changes. Finally, and most importantly, we were unable to find any obvious correlation between the freezing pressure of a Citrine crystal and the peak of its fluorescence spectrum. However, we did notice that the fluorescence peak of Citrine crystals was consistently higher than that seen in dilute solutions, typically ranging from 530 to 540 nm, rather than the solution peak of 527 nm. This was perplexing, as early crystal structures of Citrine under pressure, when viewed in the light of the simple quantum mechanical model of Citrine's fluorescence spectrum presented in section 3.6.4, indicated that the fluorescence peak should shift to the blue with increasing freezing pressure.

In order to more fully investigate the fluorescence spectra of high-pressure cryocooled samples, we moved our experiments away from CHESS to a location where longer experiments could be conducted. Professor Brian Crane in the Cornell department of Chemistry and Chemical Biology allowed us to modify an absorption micro-spectrophotometer that he had constructed that was modeled upon the absorption

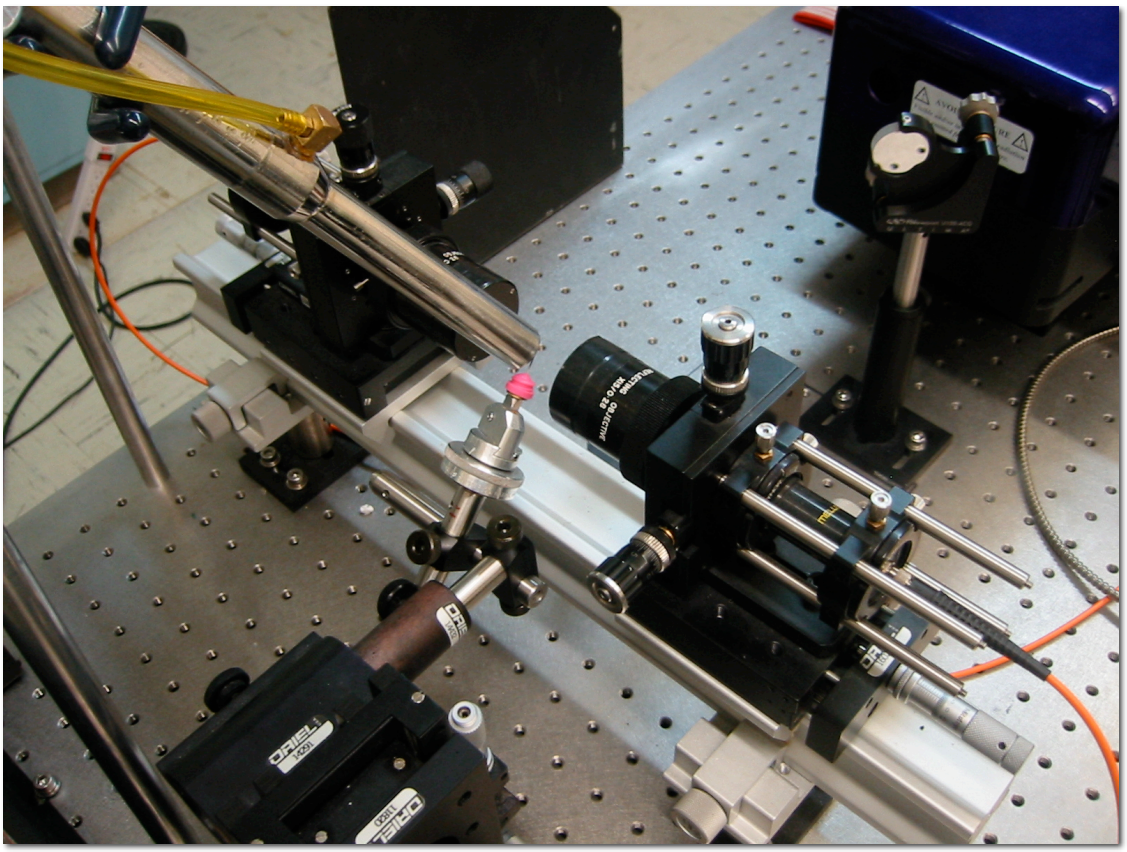


Figure 4.14: Second generation micro-spectrophotometer in Professor Brian Crane's laboratory. The fluorescence collection optics are in the center left of the photograph. The sample is mounted on a bright pink Hampton CryoBase in the center-right of the photograph. Note the mirror in the top right of the photograph used to direct the excitation beam onto the sample.

micro-spectrophotometer designed by Hadfield and Hajdu [178]. A photograph of this apparatus is shown in figure 4.14. A close-up photograph of the sample stage, sample, cryostream and fluorescence collection optics is shown in figure 4.15. Fluorescence excitation was provided by the 488 nm line of a continuous wave krypton-argon laser that was selected with a 488 nm interference filter with a 1 nm bandpass (Catalog number FL488-1, Thor Labs, North Newton, NJ, USA). The excitation laser beam was

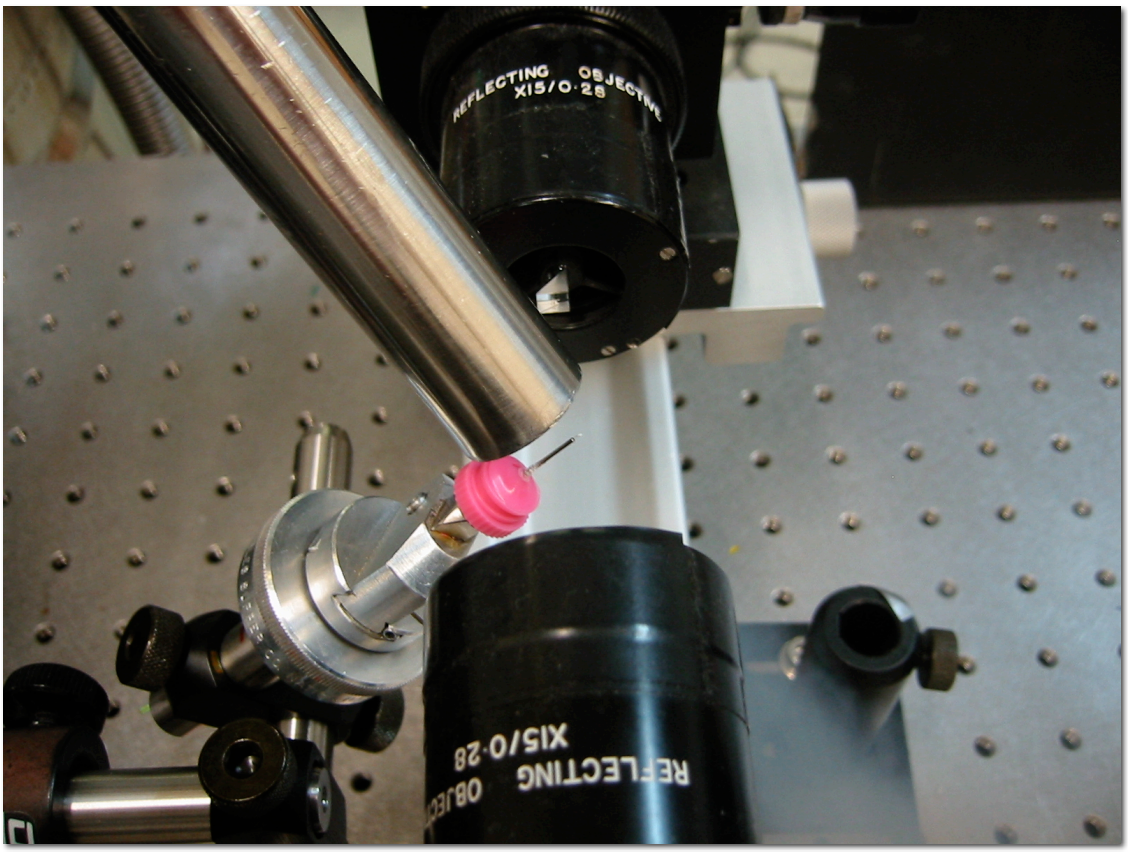
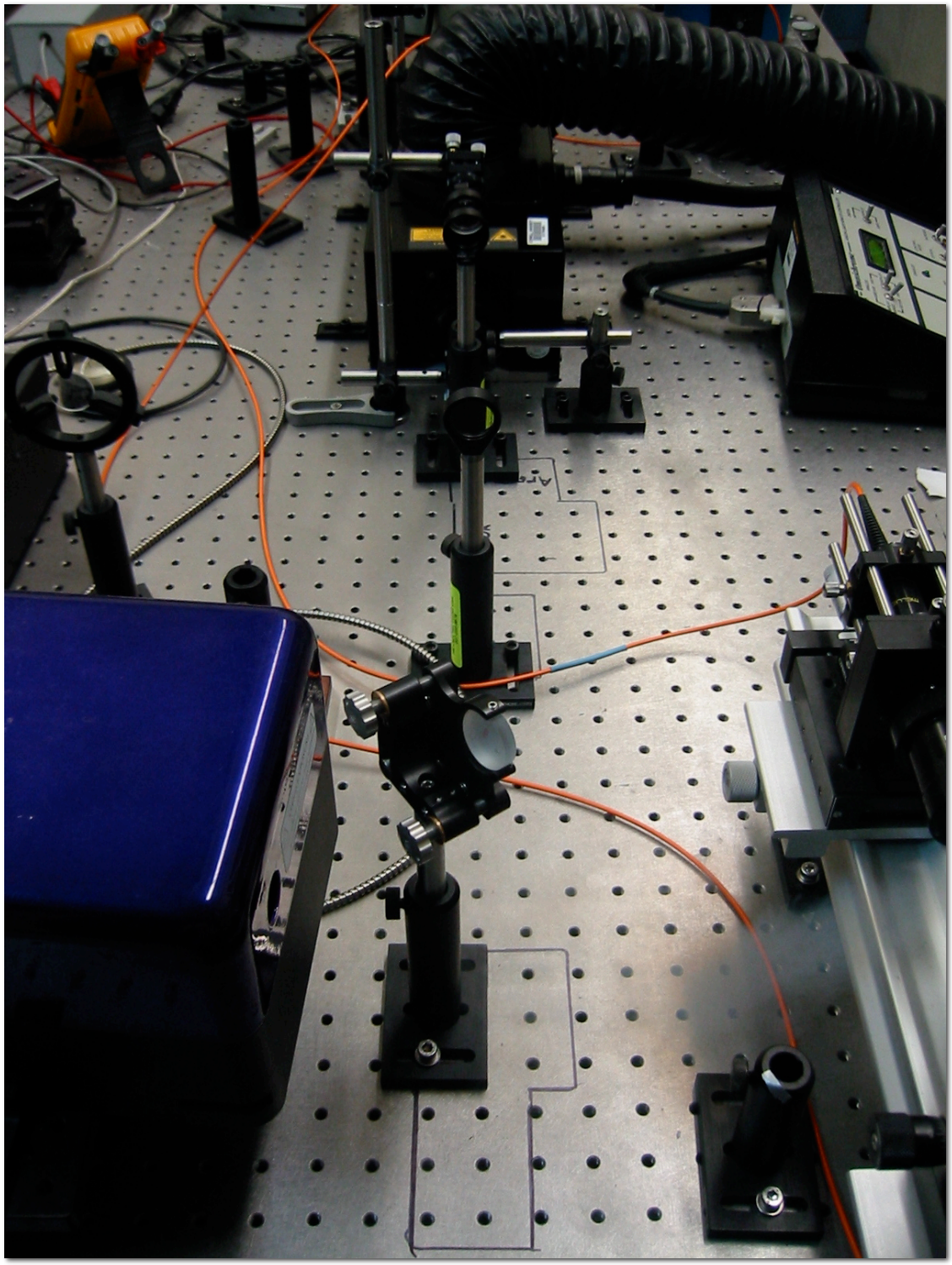


Figure 4.15: Close-up of the sample stage, cryostream and fluorescence collection optics (at the bottom of the photograph) in the second generation micro-spectrophotometer.

propagated through free space, and directed onto the sample by a mirror. This mirror can be seen in the top right of figure 4.14. Two screw actuators on the mirror mount allowed adjustment of the excitation beam position. The path of the excitation laser beam is shown in figure 4.16.

This apparatus allowed considerably more detailed investigation of the spectra of high-pressure cryocooled Citrine samples. Most importantly, experiments with this device suggested that small size variations in Citrine crystals could cause apparent shifts in the fluorescence spectrum of the sample.

Figure 4.16: Pathway of the fluorescence excitation light from the krypton-argon laser at the top of the photograph, to the excitation beam positioning mirror, where it is deflected by ≈ 90˚ toward the sample at the right of the bottom right of the photograph.



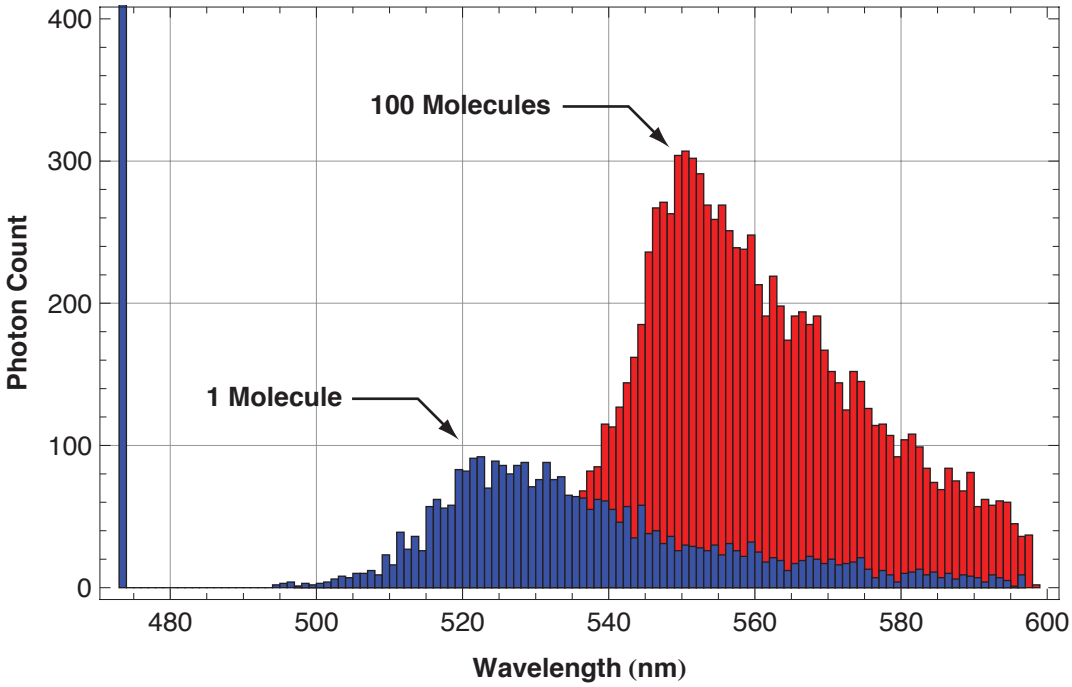


Figure 4.17: Simulated distribution of photon wavelengths after 1 (blue) and 100 (red) molecules.

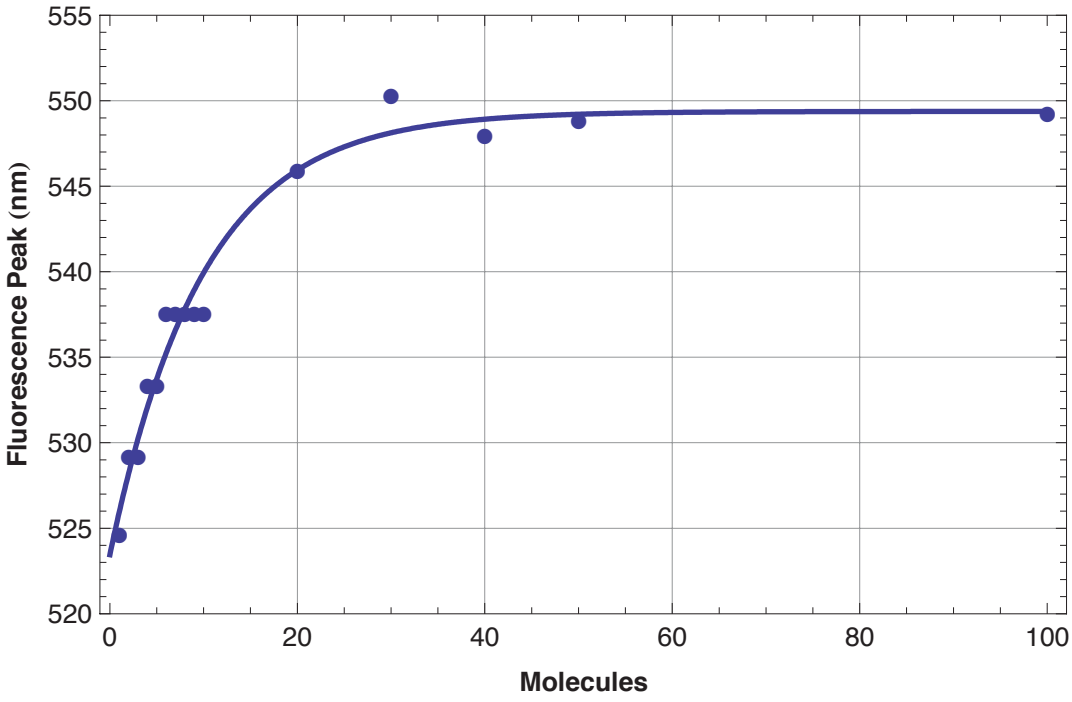


Figure 4.18: Simulated peak shift with increasing number of simulated molecules.

We believe this apparent peak shift is due to repeated absorption and re-emission of fluorescence photons. As the fluorescence and absorption spectrum of Citrine overlap, photons emitted by fluorescence may be reabsorbed. As the quantum yield of Citrine is not 1.0, there is a small, but finite probability of the complete loss of an absorbed photon. Given a sufficiently high number of Citrine molecules, this process depletes the number of photons with wavelengths that overlap with absorption spectrum of Citrine, resulting in an apparent peak shift of the sample.

A simple Monte-Carlo computer code was implemented to assist in understanding of this effect. In this model, excitation photons are generated with a wavelength of 473 nm. The measured absorption spectrum of a dilute solution of Citrine was interpolated to define the probability of absorption of a photon of any wavelength. If the photon was absorbed, the fluorescence spectrum of a random number was generated to determine the emission wavelength based upon the fluorescence spectrum of a dilute solution of Citrine. The total probability of emission was normalized to 0.89, the quantum yield of Citrine, so there existed a small probability of non-emission. A histogram showing the distribution of photon wavelengths after re-emission of 10,000 photons by 1 molecule and absorption and re-emission by a line of 100 molecules is shown in figure 4.17. The two histograms in figure 4.17 show a clear shift in the peak of the Citrine fluorescence spectrum from the solution peak of ≈ 527 nm, to close to the very far red edge of the Citrine absorption spectrum at ≈ 550 nm.

This Monte-Carlo experiment was attempted for several different lengths of molecule chain. A plot of the approximate apparent fluorescence peak position versus molecule number is shown in figure 4.18.

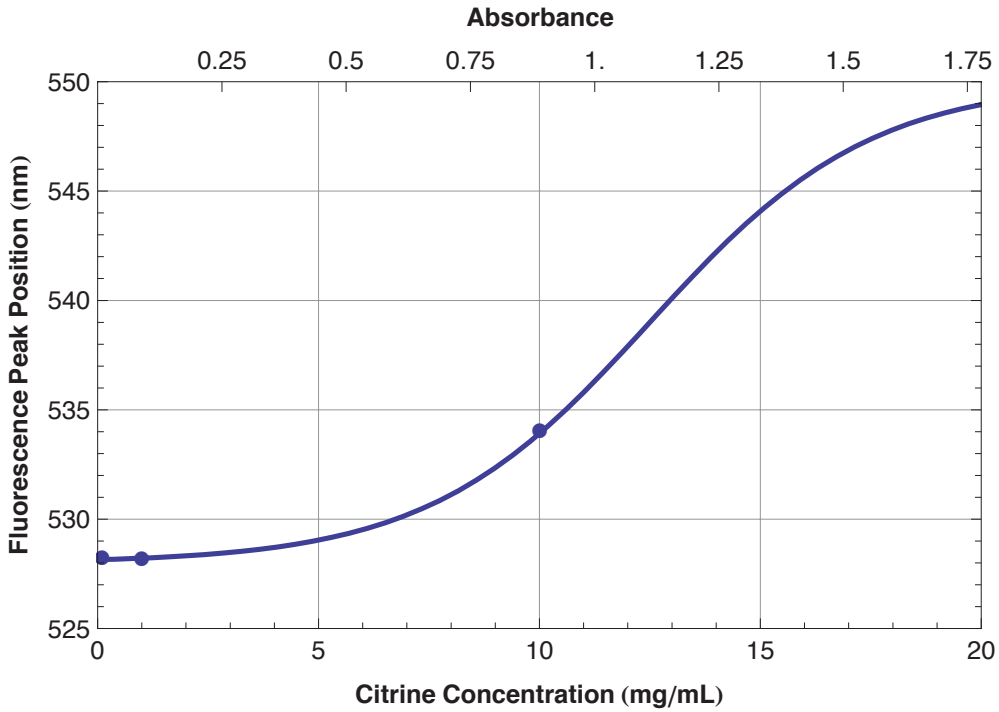


Figure 4.19: Concentration effects on the fluorescence peak of Citrine. Fitted line is a sigmoid function that asymptotes to 555 nm, the tail end of the Citrine fluorescence spectrum. The absorbance of the sample is shown along the top edge of the plot. The thickness of the sample was 0.3 mm.

A plot of the apparent fluorescence peak of Citrine solutions of varying concentration is shown in figure 4.19. At low concentrations, the Citrine fluorescence peak asymptotes to ≈ 528 nm. The Monte Carlo simulations of the increase in apparent fluorescence peak with increasing Citrine concentration indicate that the apparent peak should asymptote to ≈ 550 nm, the red edge of the Citrine fluorescence spectrum. A sigmoidal function should produce two asymptotes, one at low concentration and one at high concentration. The fluorescence peak, $\lambda_{em}(C)$, in figure 4.19 was fitted to a sigmoidal function of the protein concentration, C,

$$\lambda_{em}(C) = a + \frac{b}{1+\exp\left(-c\left(x-d\right)\right)}, \tag{4.22}$$

where, a = 528 nm, b = 22 nm, c = 0.4 and d = 12.5 mg/mL.

This function reasonably reproduces the apparent fluorescence peak shift as a function of protein solution concentration over the range 0.1 to 10 mg/mL. However, it is likely that the exact dependence of apparent fluorescence peak probably depends upon the geometry of the system under study. For instance, in the 1-dimensional case studied in the Monte-Carlo simulation, photons are limited to propagate in 1-dimension, and 1 direction only. To reach the detector, a photon must pass through all of the molecules in the chain, whereas in a 3-dimensional case, fluoresced photons are emitted over a wide range of angles and may escape the sample without interacting with all of molecules in the sample. It is interesting to note that a crystal or solution with an apparent fluorescence peak close to 550 nm (the far red edge of the Citrine absorption spectrum) was never observed, even though the optical densities of crystals are apparently very high. The density of Citrine molecule's in a crystal is one molecule per 52.5 × 61.8 × 70.7 $Å^3$, giving a concentration C = 7.2 M. Assuming that the solution extinction coefficient of Citrine ξ = 80,000 M^{-1} cm^{-1} is appropriate for a crystal, the absorbance of a Citrine crystal of length l = 50 μm is

$$A = \xi C l \tag{4.23}$$

$$A = 80000 \times 7.2 \times 50 \times 10^{4} = 2896 \tag{4.24}$$

This probably indicates that escape mechanisms serve to prevent the apparent peak shift from ever reaching the maximum implied in the Monte Carlo simulation presented in figures 4.17 and 4.18. However, these models do suggest that it is plausible to believe that small size variations in crystals may cause crystal-to-crystal

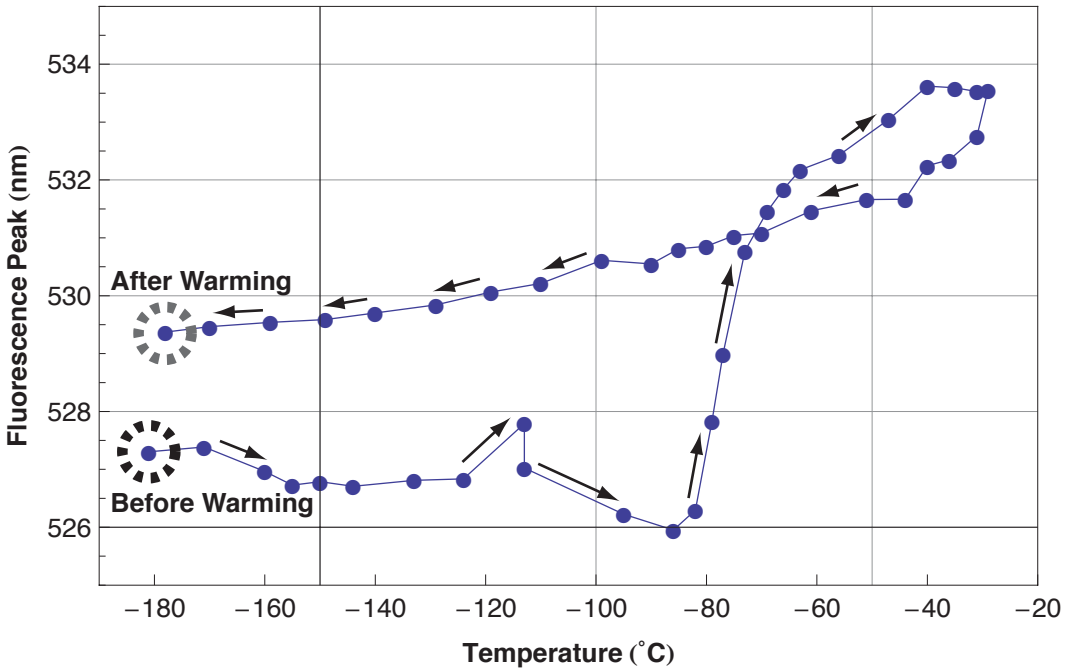
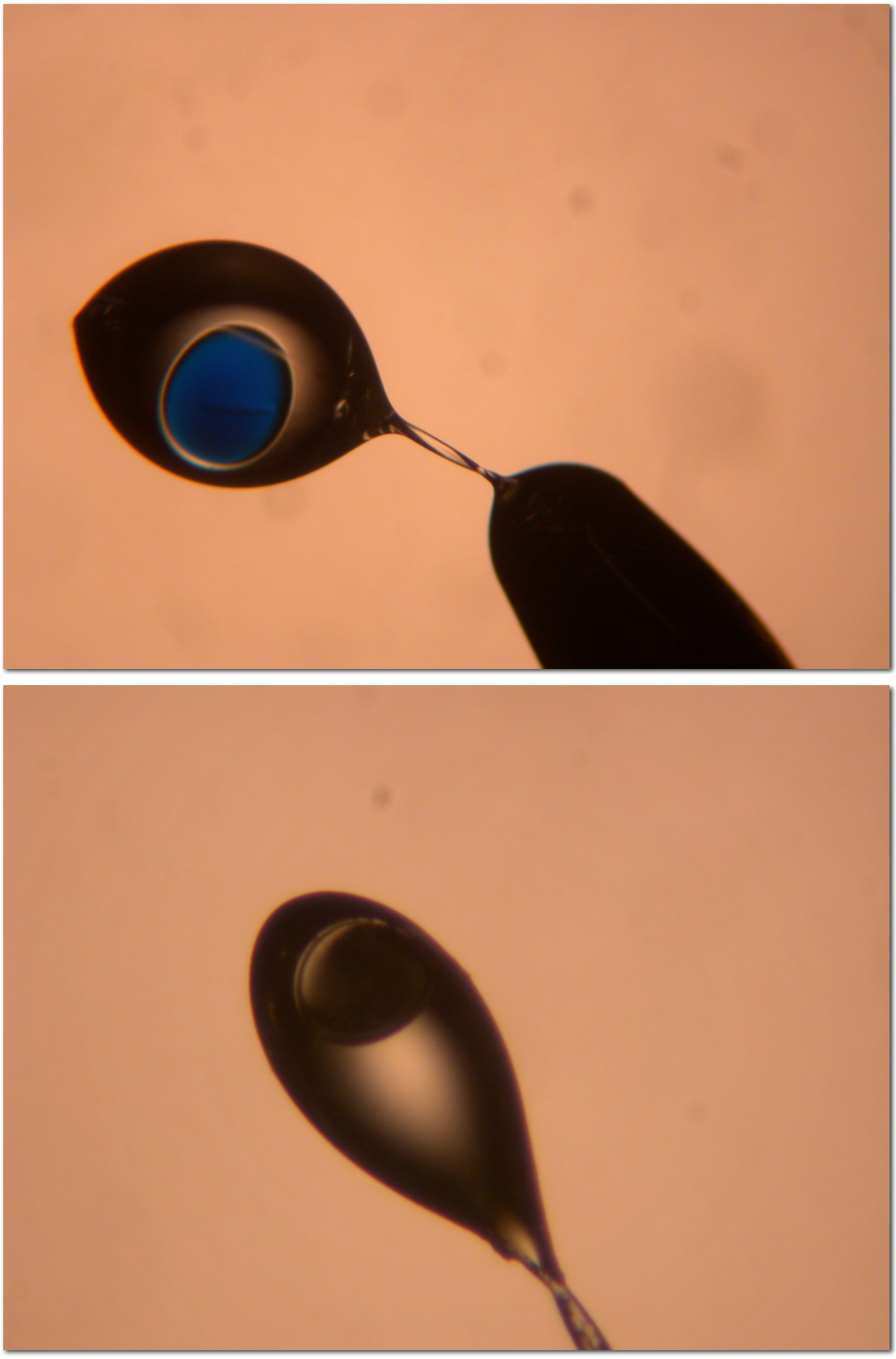


Figure 4.20: Fluorescence peak of a high-pressure cryocooled crystal of Citrine, frozen at 175 MPa, warmed above its glass transition temperature and then re-cooled.

variations in fluorescence peak that may dominate any effect due to pressure deformation of the chromophore. However, the effect of high-pressure cryocooling, and thus pressure deformation of the chromophore of on the fluorescence spectrum of Citrine was indicated by experiments in which a high-pressure cryocooled crystal of Citrine was warmed slowly, and the fluorescence peak recorded. These experiments indicated that a fluorescence peak shift is locked into the crystal by the high-pressure cryo-cooling technique. As the crystal is warmed the fluorescence peak gently shifts to the red, by a small amount, until the crystal reaches its glass transition temperature, at which point, the fluorescence shifts markedly to the red. Upon re-cooling, the fluorescence peak does not return to the blue, but remains red, indicating that a fluorescence blue-shift had been locked into the crystal by the high-pressure cryo-cooling technique. This effect is shown in figure 4.20.

Figure 4.21: Top: Droplet of food dye injected into a droplet of NVH oil in a Hampton CryoLoop. Bottom: Droplet of Citrine injected into a droplet of NVH oil.



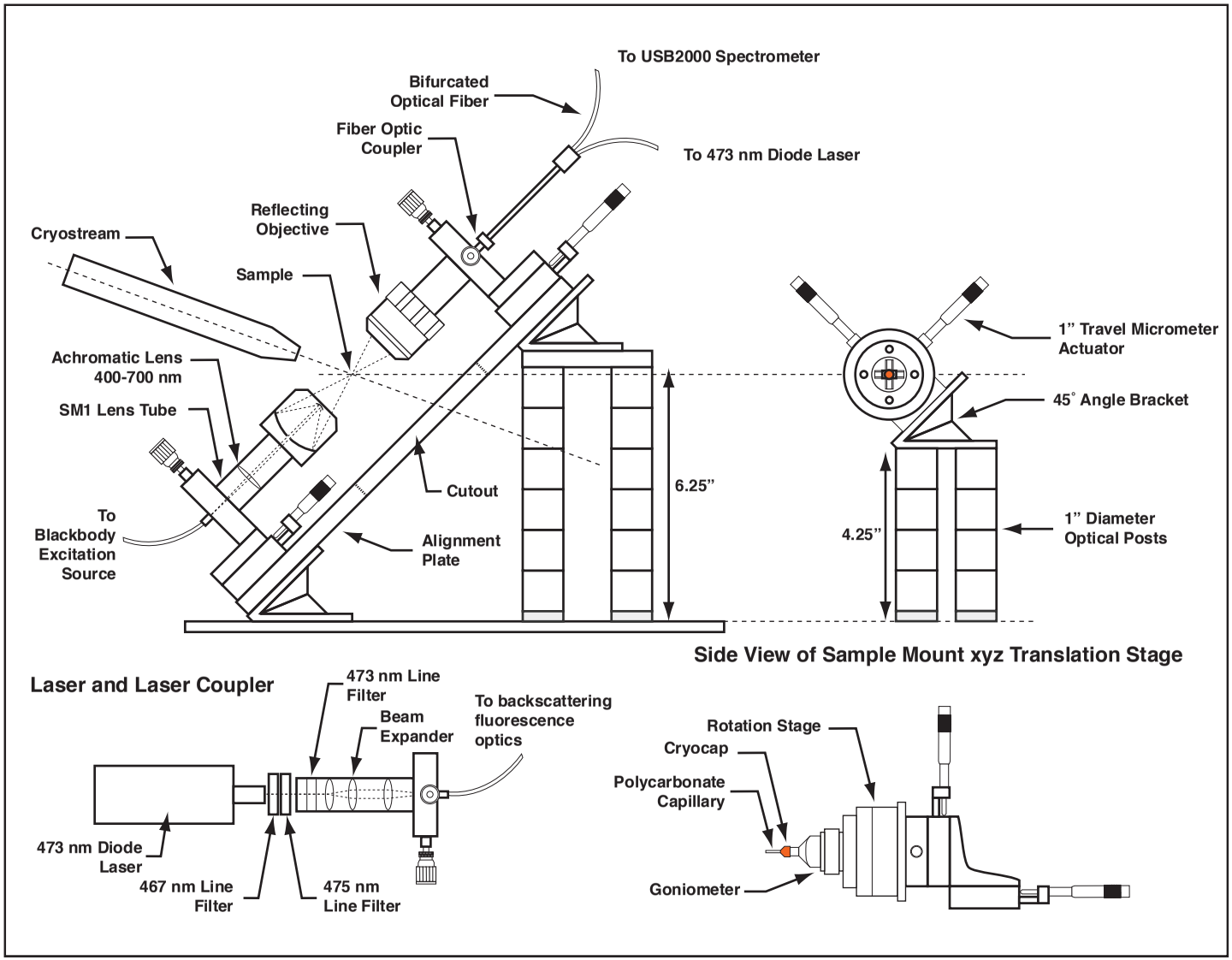
In an attempt to address the limitations of the crystals we experimented with freezing droplets of dilute Citrine solution inside droplets of NVH oil (Catalog number HR3-617, Hampton Research, Aliso Viejo, CA, USA). Droplets of NVH oil were prepared by dipping a Hampton CryoLoop into NVH oil. Dilute Citrine solutions were injected into the oil droplets with a micro-syringe (Catalog number 86257, Hamilton Company, Reno, NV, USA). This technique had a low success rate, as it was very difficult to remove air bubbles from the droplets, which caused optical scattering. Additionally, it proved difficult to produce samples with a reproducible size. Photographs of two successful droplets, one containing food dye, and the other containing Citrine are shown in figure 4.21.

4.4.3 Third Generation Micro-spectrophotometer

To further investigate the fluorescence peak shift of Citrine samples under high-pressure cryocooling conditions, we constructed a third generation micro-spectrophotometer in the Gruner lab. This gave considerable extra convenience to experiments, and with the experience gained from the previous two micro-spectrophotometers used in these experiments, were able to construct a device that was easy to use, and allowed rapid measurement of many samples.

We constructed a fluorescence and absorption spectroscopy apparatus, drawing upon the designs of Hadfield [178] and Klink *et al*. [179]. Like Hadfield's apparatus [178], the device has the ability to measure the absorption of small samples. Additionally, following the design of Klink *et al.* [179], the device features backscattering fluorescence, allowing easy measurement of the fluorescence spectrum of a sample without the need to align the collection and excitation optics, as they are both combined into a single optic. The system has been optimized for ease use, ease of

Figure 4.22: Schematic of third generation micro-spectrophotometer.


To USB2000 Spectrometer
Bifurcated Optical Fiber
Fiber Optic Coupler
To 473 nm Diode Laser
Reflecting Objective
Cryostream
Sample
Achromatic Lens 400-700 nm
SM1 Lens Tube
To Blackbody Excitation Source
Cutout
Alignment Plate
6.25”
1” Travel Micrometer Actuator
45° Angle Bracket
4.25”
1” Diameter Optical Posts
Side View of Sample Mount xyz Translation Stage
Laser and Laser Coupler
473 nm Line Filter
Beam Expander
To backscattering fluorescence optics
473 nm Diode Laser
467 nm Line Filter
475 nm Line Filter
Rotation Stage
Cryocap
Polycarbonate Capillary
Goniometer

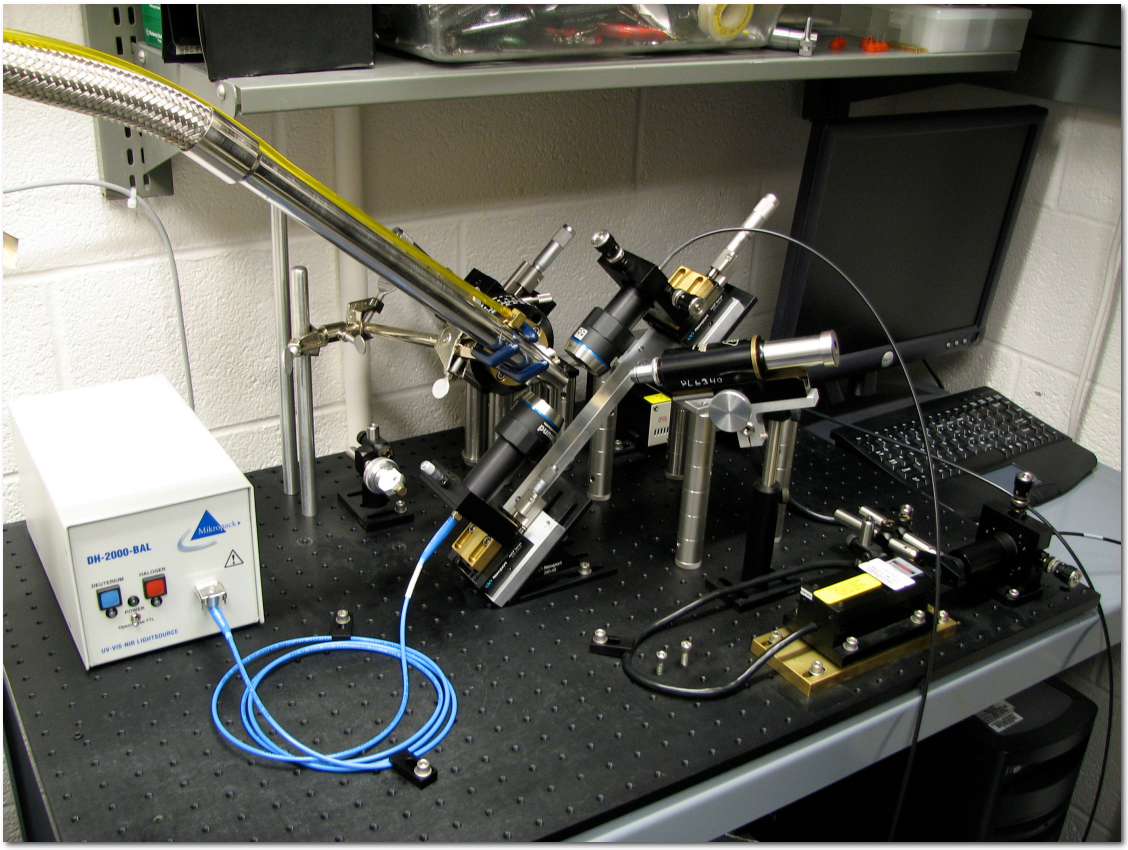


Figure 4.23: Photograph of the third generation micro-spectrophotometer showing, from the left, the excitation lamp, the cryostream, the excitation optic, the fluorescence optic, the sample alignment telescope, the fluorescence excitation laser, and the data acquisition computer.

sample handling, use at low temperatures, reproducibility of measurements and for operator safety and comfort. These design considerations dictated the physical layout of the optical and cooling elements of the system. A schematic of the micro-spectrophotometer is shown in figure 4.22. A photograph of the entire micro-spectrophotometer is shown in figure 4.23. A closer view of the optics of the micro-spectrophotometer is shown in figure 4.24, and a close-up of the sample mounting area is shown in figure 4.25.

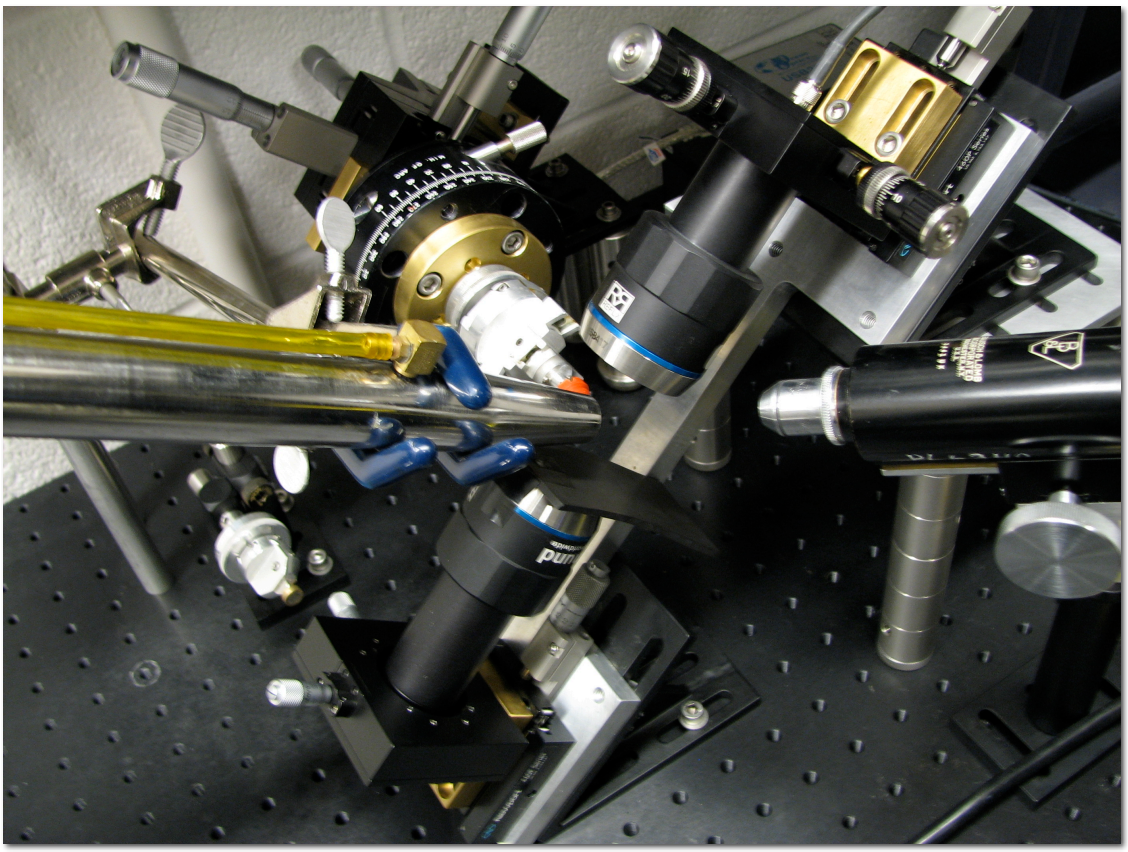


Figure 4.24: Photograph of the third generation micro-spectrophotometer showing, from the left, the cryostream, the excitation optic, the fluorescence optic and the sample alignment telescope.

4.4.4 Design Elements of Third Generation Micro-spectrophotometer

4.4.4.1 General Layout

Use of the micro-spectrophotometer with high-pressure cryocooled samples necessitated several features of the layout of the apparatus. Firstly, the layout of the apparatus needed to accommodate a low temperature nitrogen cryostream (Molecular Structure Corporation, The Woodlands, TX, USA) to allow for the use of samples at low temperature. Secondly, rapid and convenient sample loading were desirable, to prevent inadvertent warming of high-pressure cryocooled samples and to allow the inspection of many samples without user fatigue. To prevent accidental exposure of

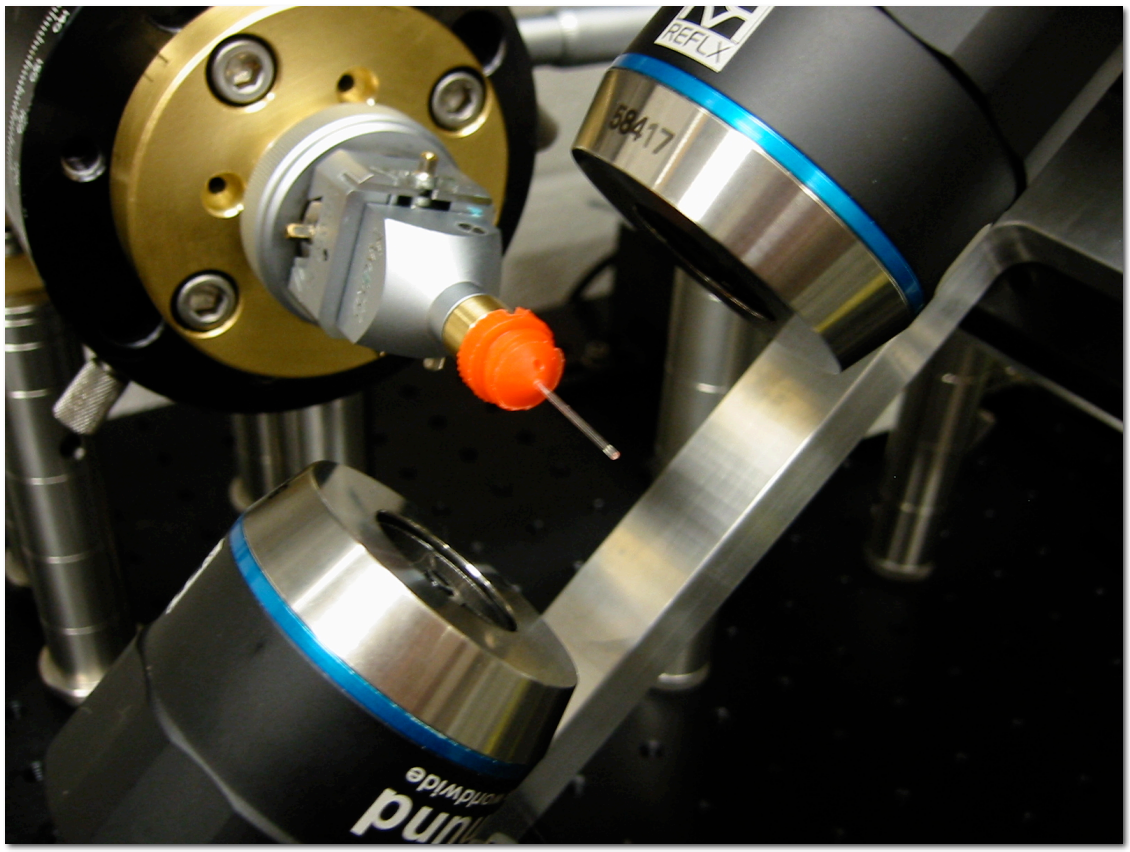


Figure 4.25: Photograph of the micro-spectrophotometer fluorescence and excitation optics, sample mount and alignment plate. A polycarbonate capillary is mounted to an orange Hampton CryoCap.

the user to the laser beam the excitation optics were required to not point towards the user. Finally, the sample stage, light sources and optics had to be rigidly mounted, to permit exact positioning of the excitation light.

All components were mounted on a 2 ft × 3 ft threaded optical breadboard (Catalog number MB2436, Thor Labs, Newton, NJ, USA) that was itself mounted on Isomode pads (Catalog number 60015K21, McMaster-Carr, Santa Fe Springs, CA, USA) to isolate it from vibration. The breadboard and Isomode pads were mounted on a steel workbench (Modern Equipment Company, Omaha, NE, USA) with vibration

damping, leveling feet (Catalog number 6167K13, McMaster-Carr, New Brunswick, NJ, USA). As moisture accumulation was likely on many surfaces due to the presence of a low temperature cryo-stream, all screws used in construction were coated with bearing grease (Phil Waterproof Grease, Phil Wood Company, San Jose, CA, USA) to prevent galvanic fusion.

The sample stage was mounted ≈ 6 inches (approximately half of the length of the author's forearm) above the optical breadboard, to allow the user to easily transfer a frozen sample from a low form dewar (Catalog number HR5-102, Hampton Research, Aliso Viejo, CA, USA) on the breadboard to the sample stage using a pair of CryoTongs (Hampton Research, Aliso Viejo, CA, USA). The sample stage is mounted on four 1-inch thick posts (Pedestal Posts, Newport Corporation, Irvine, CA, USA) to ensure its stability under the heavy weight of the sample stage. The sample is mounted on a two-arc goniometer (Charles Supper Company, Natick, MA, USA) with a Hampton Research magnetic base (Either HR4-627, HR4-943 or HR4-629, Hampton Research, Aliso Viejo, CA, USA). The goniometer is attached through a custom mating piece (figure 4.26) to a 360˚ micrometer driven rotation stage (Newport Corporation, Irvine, CA, USA. Although the item used in this design is discontinued, the 481-AS is extremely similar). The rotation stage is affixed through a custom mating piece (figure 4.27) to an *x-y-z* micrometer translation stage (460P-XYZ, Newport Corporation) with 1 inch of travel in each dimension, actuated by micrometer drives (SM-25, Newport Corporation). The goniometer is suspended several inches forward from the footprint of the translation stage pillars on the optical breadboard, and faces towards the user, allowing the user to view the sample mounting area and ease sample loading.

Figure 4.26: Custom mating piece to attach goniometer to Newport Rotation stage.

Adaptor to Couple Newport PR50CC Rotation Stage to Huber Goniometer

Buz Barstow
1st December 2006
Material: Brass or Aluminum Rod

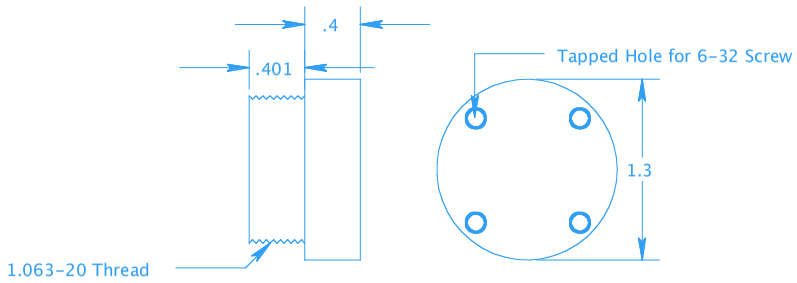
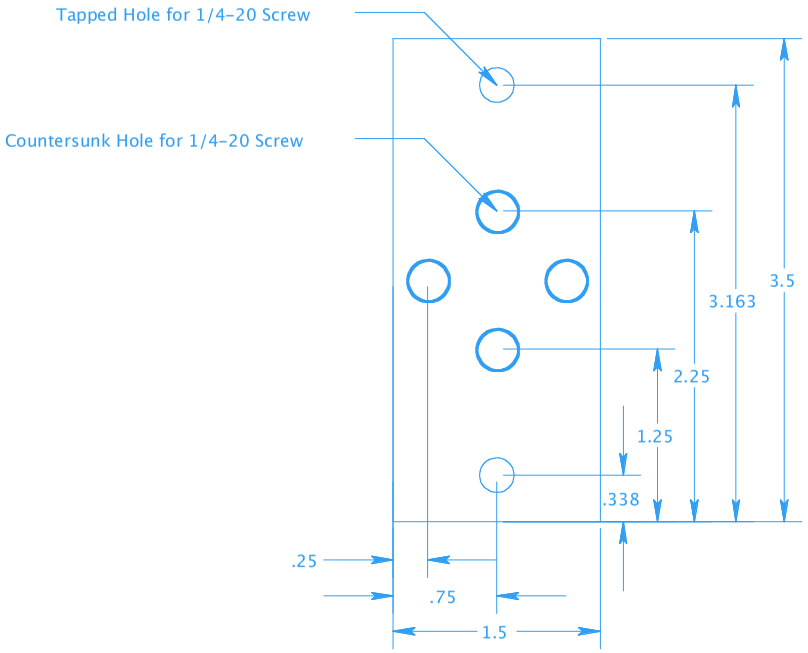


Figure 4.27: Adaptor to couple manual rotation stage to *x*-*y*-*z* translation stage.

Adaptor to Couple Newport xyz Translation Stage to Manual Rotation Stage

Buz Barstow
1st December 2006
Material: 3/8" Thick Aluminum or Brass



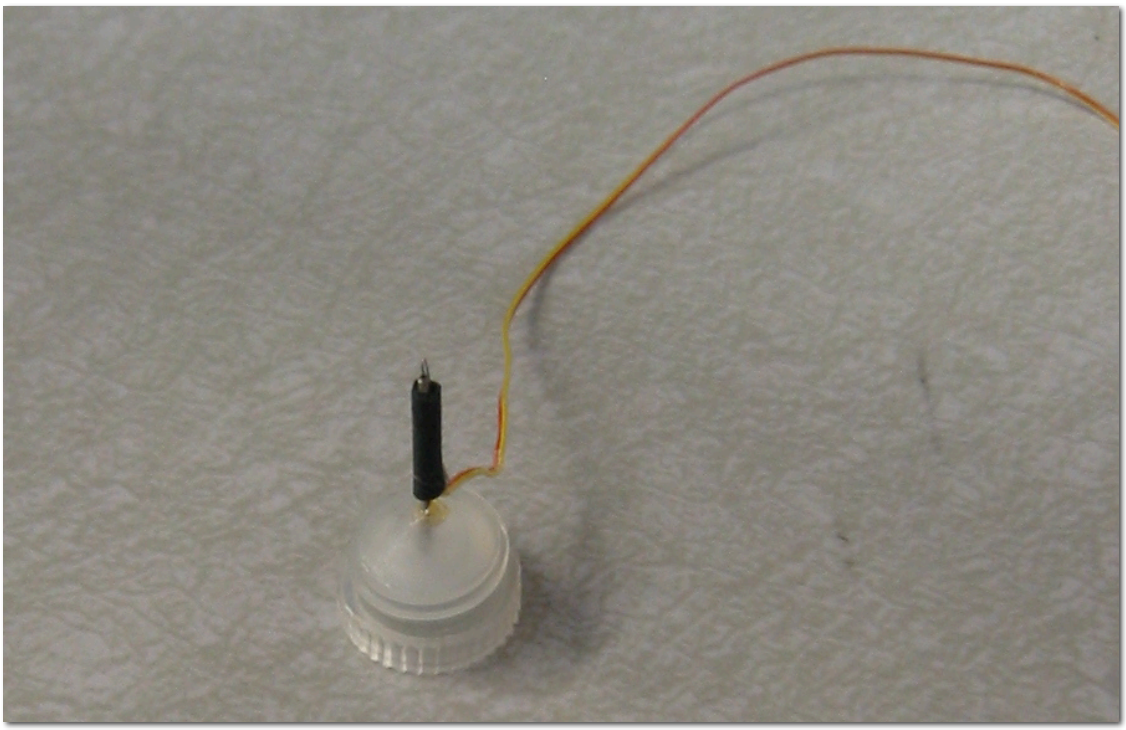


Figure 4.28: N-type thermocouple mounted on a Hampton Research CryoLoop and CryoBase.

The cryostream nozzle is directed at right angles to the rotation axis of the sample stage, and points down at a shallow angle, of approximately 30˚ to the horizontal. The temperature of the stream at the sample position was verified with a custom made sensor. An N-type thermocouple was attached to a CryoLoop (Hampton Research, Aliso Viejo, CA, USA) with heat-shrink material and mounted on a CryoBase (Hampton Research). A photograph of this sensor is shown in figure 4.28.

The fluorescence excitation/collection optics are opposite the cryostream, and are also mounted above the optical breadboard using four 1-inch thick posts (Pedestal Posts, Newport Corporation, Irvine, CA, USA), above the level of the sample stage, and perpendicular to the rotation axis of the sample stage. The optic is directed down towards the sample stage by a 45˚ angle bracket (Catalog number 360-45, Newport

Corporation). The face of the excitation/collection reflecting objective is 1 inch away from the sample, the working distance of this objective. This arrangement limits the user's exposure to the excitation light as it is directed at right angles to the line of sight of the user, and also limits the possibility of the excitation light leaving the immediate area of the crystal fluorescence monitor. This arrangement also minimizes exposure of the excitation/collection optics to the cryostream, as it is mounted above the level of the cryostream nozzle. A photograph of the fluorescence excitation/collection optics can be seen in figures 4.24 and 4.25. The widely separated posts that form the base of the excitation/collection optics also give the cryostream a large distance over which to dissipate past the sample, allowing the cryostream flow to remain laminar. The legs of the mount for the fluorescence excitation/collection optics can be seen in figure 4.24. The absorption excitation optic, shown on the left in figures 4.24 and 4.25 are mounted underneath the cryostream, and could be accidentally exposed to liquid oxygen that may have pooled in the cryostream and may be ejected when the cryostream flow rate is turned to high. The absorption excitation optic is normally protected from liquid emissions from the cryostream by a neoprene sheet that is taped over the face of the optic. This protective sheet can be seen in figure 4.24.

The sample is aligned to the excitation/collection optics by translations of the crystal station goniometer. Sample alignment is viewed with a 40 × telescope with a working distance of 1 inch (2.54 cm), using a grid reticule (Catalog number NT39-446, Edmund Optics, Barrington, NJ, USA) as a reference mark. This telescope can be seen on the right of figures 4.23 and 4.24.

The optics room in which the crystal fluorescence monitor is located is darkened during data collection as the spectrum of the room lights overlaps with the Citrine

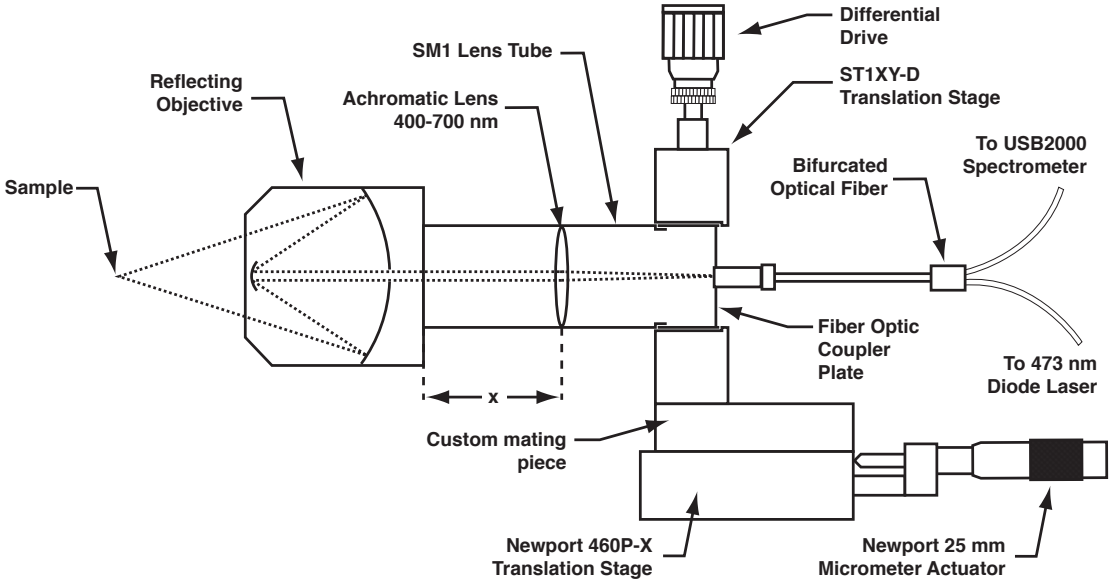


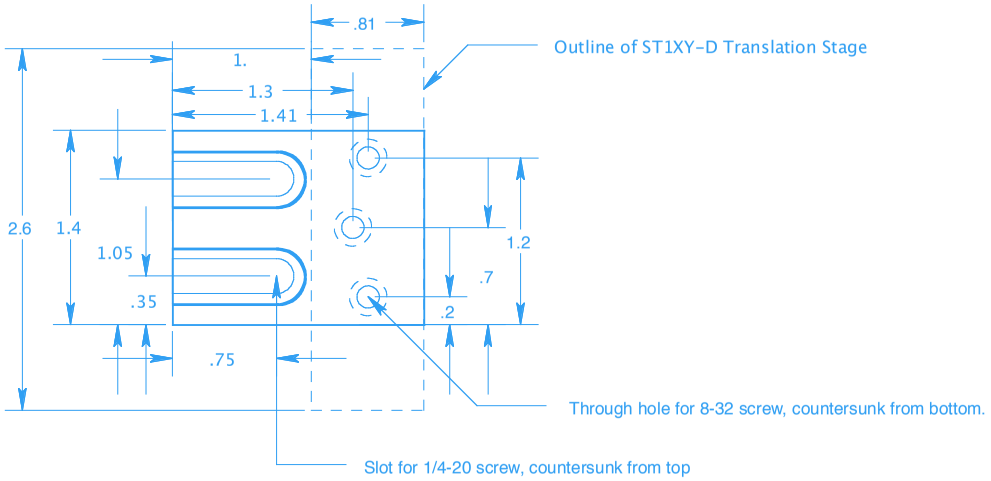
Figure 4.29: Fluorescence backscattering/excitation optics for third generation micro-spectrophotometer. The absorption excitation optics have an almost identical layout. Note that the distance *x*, is variable as the light from the reflecting objective is collimated. The distance x is slightly longer in the absorption excitation optics to permit more convenient placement of the ST1XY-D lens translation stage. Also, a 12.5 mm micrometer actuator is used in the absorption excitation optics. A schematic of the custom mating piece is shown in figure 4.30.

excitation and fluorescence spectra. The LCD monitor of the data acquisition computer monitor is turned to the lowest brightness and contrast setting. All other light sources, typically LED lights on electronics, are taped over with black electrical tape to further reduce the background light signal.

Figure 4.30: Custom mating piece to couple ST1XY-D lens translation stage to Newport 460P-X linear translation stage in fluorescence excitation/ collection optics and absorption excitation optics.

Base Plate for ST1XY-D Translation Stage

Buz Barstow
6th December 2006
Material: 0.5” Brass Plate reduced in height by 8.2 mm (0.323”)



4.4.4.2 Absorption Excitation and Fluorescence Optics

As in the design by to Klink *et al*. [179] the front end of the fluorescence excitation/ collection optic is a large numerical aperture, infinite conjugate, reflecting objective (Catalog number NT58-417, Edmund Optics, Barrington, NJ, USA) that is achromatic over a wide range of UV, optical and infrared wavelengths. In order to couple these lenses to an optical fiber, a focusing lens is required. An achromatic lens is used to accomplish this task. Currently, this optic is a 400 to 700 nm achromat (Catalog number AC254-030-A1, Thor Labs, North Newton, NJ, USA).

However, if operation in another wavelength range is desired, the lens tube may be easily unscrewed and replaced with another achromatic lens. The achromat focuses collimated light from the reflecting objective onto a bifurcated optical fiber. One bundle of the fiber leads to a USB2000 spectrometer (Ocean Optics, Dunedin, FL, USA), while the other bundle comes from a diode laser that provides fluorescence excitation light. The interior surfaces of the lens tube are covered with black flocked paper to absorb scattered light and reduce backgrounds due to scattering in the system.

The reflecting objective and lens tube is mounted on a Thor Labs ST1-XY translation mount. Motion of the fluorescence optics is actuated by differential drives, while the excitation optics have micrometer drive actuators. The micrometer drives were found to perform better than the differential drives. This permits fine positioning of the optics relative to one another over a short range of distances.

A schematic of the fluorescence excitation/collection optic is shown in figure 4.29. A photograph of the rear lens tube is shown in figure 4.31. The absorption excitation optic is essentially identical to the fluorescence excitation/collection optic, except that

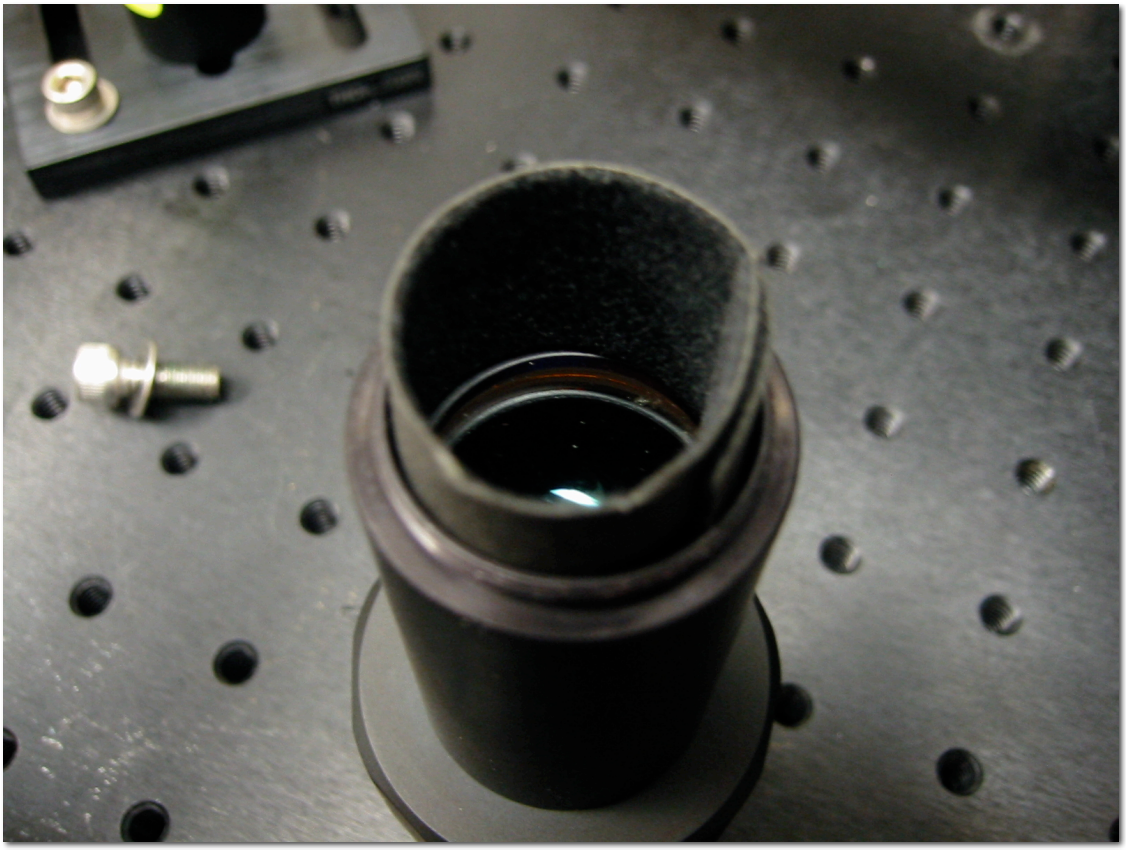


Figure 4.31: Photograph of the fluorescence optics showing the achromatic focusing lens, and black flocked paper lining of the lens tube.

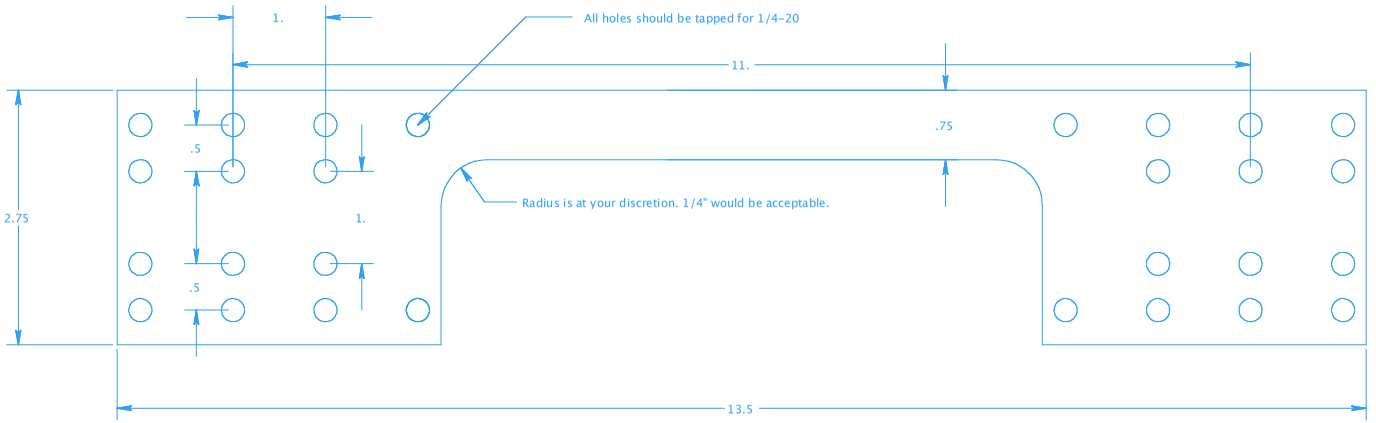
the lens tube connecting the reflecting objective to the optic translation stage is slightly longer, so as to permit more convenient placing of the optic translation stage.

For absorption measurements, the fluorescence collection/excitation optic is used purely as an absorption light collection optic. The absorption excitation and fluorescence collection/excitation optics are aligned with respect to one another with a custom machined alignment plate. This alignment plate is shown in the photographs in figures 4.23, 4.24 and 4.25. A schematic of the alignment plate is shown in figure 4.32.

Figure 4.32: Plate to align fluorescence excitation/collection optics and absorption excitation optics for micro-spectrophotometer.

Alignment Plate for Absorption Upgrade to Crystal Fluorescence Monitor

Buz Barstow
16th March 2007
Material: 3/8" Aluminum Plate



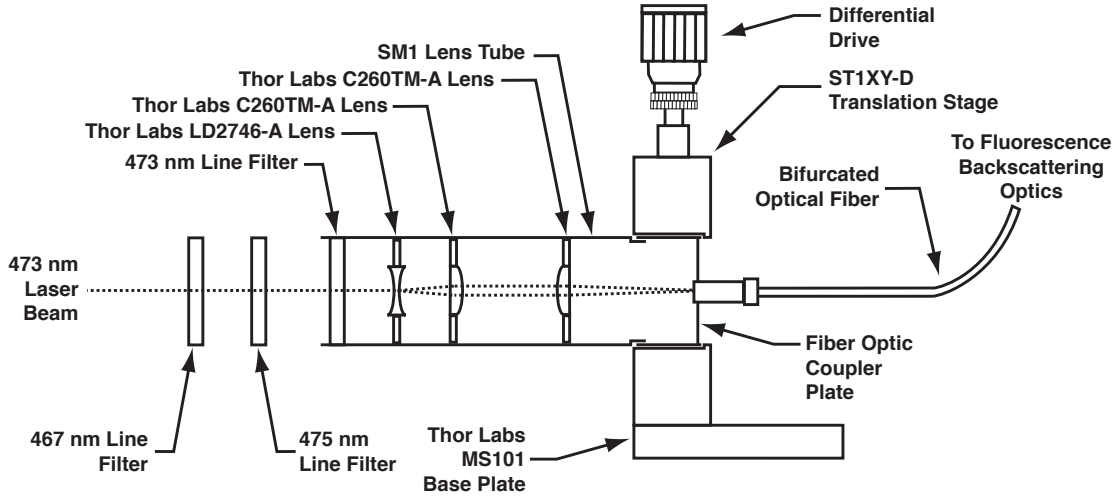


Figure 4.33: Schematic of laser to optical fiber coupler.

Fine alignment of the absorption excitation and fluorescence excitation/collection optics is performed by placing a target, usually a piece of white paper, at the focal spot of the excitation optic. The focus appears as a bright dot on the paper. The fluorescence excitation/collection optic may then be aligned to this spot by turning the fluorescence excitation laser on, and aligning the laser spot to the focal spot of the absorption excitation optic.

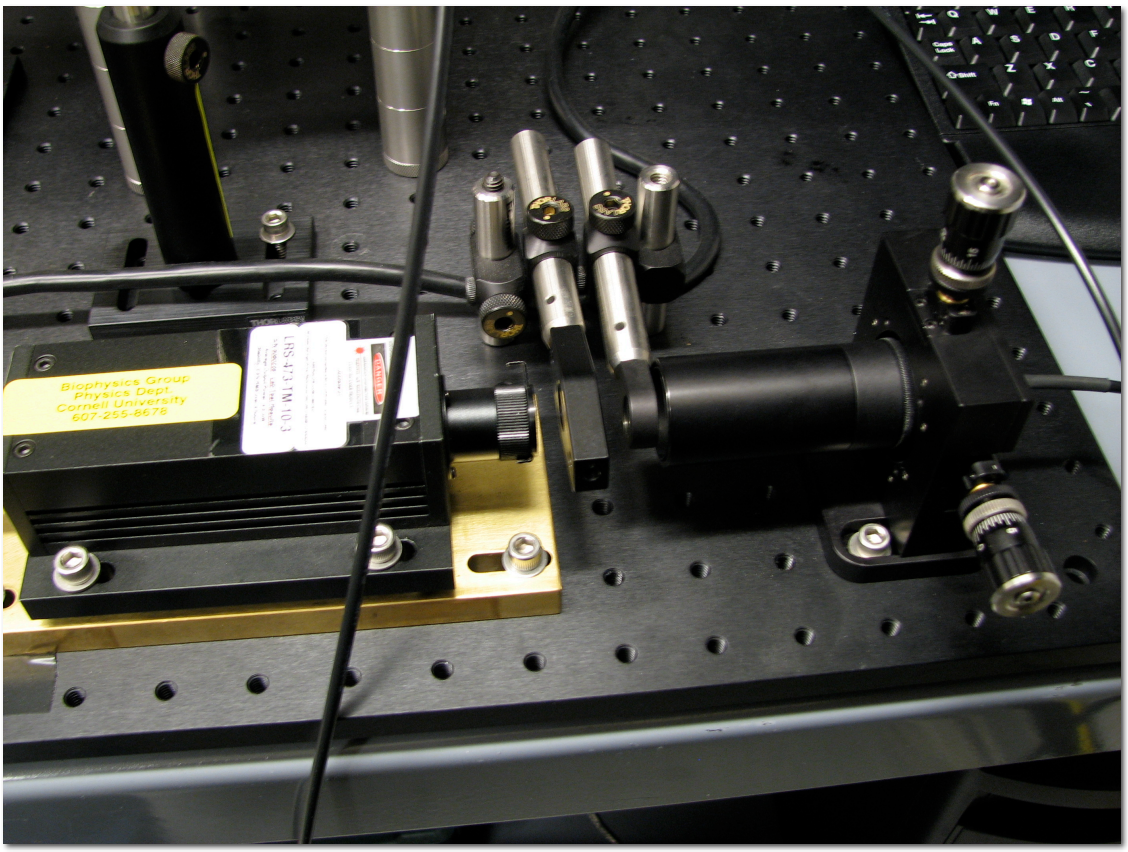


Figure 4.34: Photograph of the excitation laser and fiber coupler optics. In normal use this assembly is covered by a box with a black flocked paper interior.

4.4.4.3 Excitation Sources

Fluorescence excitation was provided by a high-stability 473 nm diode laser (Catalog number LRS-473-TM-10, LaserGlow Inc, Toronto, Canada). This laser was specifically chosen to provide excitation at the blue-edge of the Yellow Fluorescent Protein absorption spectrum [52], so as to minimally overlap with the fluorescence spectrum. The power output of the device was limited upon our request to 5 mW to ensure classification as a class IIIa laser), coupled by a bifurcated fiber optic bundle (Ocean Optics, Dunedin, FL, USA) to a long working length, high numerical aperture reflecting objective (Catalog number NT58-417, Edmund Optics, Barrington, NJ,

USA), used as both excitation and collection optics in the fluorescence backscattering mode.

The spectral width of the fluorescence excitation laser beam was narrowed by a train of bandpass optical interference filters; 1 high-transmission 473 nm filter (Catalog number LD01-473/10-12.5, Semrock, Rochester, NY, USA), and two lower transmission filters, at 470 (Catalog number NT43-062, Edmund Optics, Barrington, NJ, USA) and 467 nm (Catalog number NT43-061, Edmund Optics).

A schematic of the laser to optical fiber coupler is shown in figure 4.33. A photograph of this optical train is shown in figure 4.34. The diode laser is raised slightly above the optical breadboard by a 0.5" thick brass plate that also serves as a heat sink. This heat sink plate may be seen in figure 4.34. Absorption excitation light was provided by a high-intensity blackbody lamp (DH-2000, Ocean Optics, Dunedin, FL).

4.4.4.4 Spectrometer

The backscattered fluorescence light was analyzed with an Ocean Optics USB2000 spectrophotometer (Ocean Optics, Dunedin FL, USA) with a 200 μm slit and a #2 grating. The spectrophotometer was calibrated prior to use by fiber connection to a blackbody calibration source (DH-2000 Calibration lamp, Ocean Optics, Dunedin, FL). It should be noted that Ocean Optics sells a blackbody lamp, and blackbody calibration source both called the DH-2000. We used the calibration version that has a carefully calibrated spectrum, but much lower intensity for calibration purposes, and also used the high intensity version of the DH-2000 for absorption excitation.

4.4.4.5 Preparation of High-Pressure Cryocooled Solutions

As discussed in section 4.3.2, due to the high optical density and overlapping absorption and fluorescence spectra of Citrine crystals, we were unable to repeatably measure the position of the peak of the fluorescence spectra of pressure cooled Citrine crystals due to small size variations in the crystals. Thus, we were forced to seek a substitute: high-pressure cryocooled solutions in polycarbonate capillaries (Plastic part number 8-000-1000, length 5 inches (127 mm), inner diameter 0.012 inches (304.8 μm), outer diameter 0.036 inches (914.4 μm), Drummond Scientific Company, Broomall, PA, USA). High-pressure cryocooling of Citrine in flat glass capillaries with an outer diameter of 0.1 mm (VitroCells, Catalog number 8505, VitroCom Inc., Mountain Lakes, NJ, USA) was also attempted, but the capillaries were found to be too mechanically fragile for use in the high-pressure cryocooling apparatus.

Citrine stock solution (20 mg/mL Citrine in 50 mM HEPES, pH 7.5) was slowly dialyzed into crystallization precipitant solution (5% PEG 3350, 50 mM sodium acetate, 50 mM ammonium acetate, pH 5.0). The curve of fluorescence peak versus concentration shown in figure 4.19 indicates that a Citrine solution with a

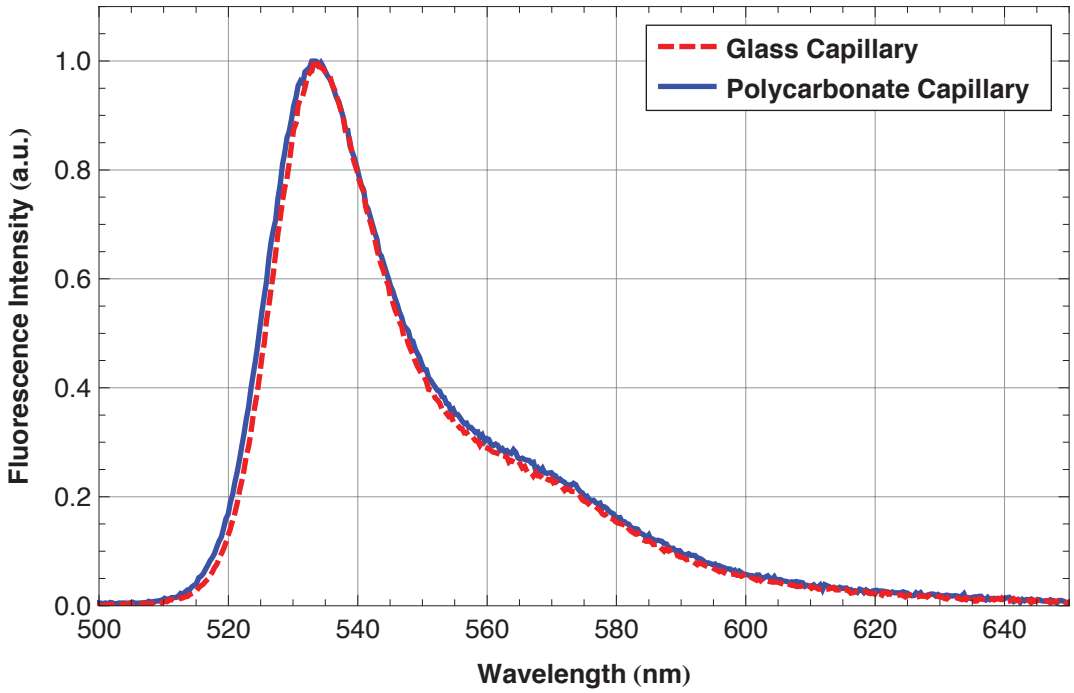


Figure 4.35: Comparison of the fluorescence spectrum of 10 mg/mL Citrine sample in 50 mM Tris, pH 7.5, flash frozen at room pressure, at -170˚ C, measured in a polycarbonate capillary and measured in a flat glass capillary.

concentration of 1 mg/mL in a 300 µm diameter capillary should be free of concentration-induced fluorescence spectrum artifacts.

The solution was loaded into optically clear polycarbonate capillaries that were cut to approximately 18 mm in length, so as to match the length of a mounted cryo-loop, and were frozen in the high pressure cryo-cooling apparatus [117] (chapter 2). The frozen capillaries were transferred in a low form liquid nitrogen dewar to the micro-spectrophotometer and transferred by hand to the station's goniometer mount.

Fluorescence spectra from flash frozen Citrine samples in polycarbonate capillaries and glass capillaries are shown for comparison in figure 4.35.

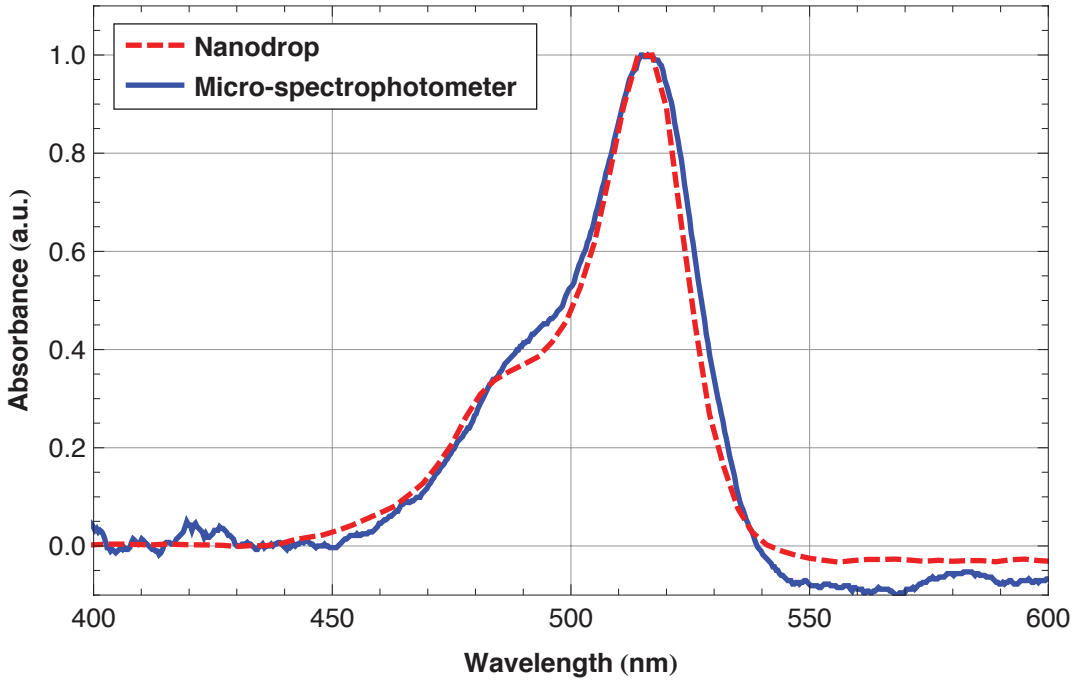


Figure 4.36: Absorbance of Citrine solution sample at room temperature measured in Nanodrop ND1000 and microspectrophotometer using SpectraSuite (Ocean Optics, Dunedin, FL, UDA) software for absorbance calculation. 10 mg/mL Citrine in 50 mM Tris pH 7.5.

4.3.4.6 Absorption

To demonstrate the absorption measurement capabilities of the micro-spectrophotometer, the room temperature absorbance of a Citrine sample measured with the micro-spectrophotometer is shown in figure 4.36. For comparison, the absorption spectrum of the same sample measured with a Nanodrop ND-1000 absorption spectrophotometer is plotted alongside. At first examination, the two spectra are very similar. It is worth noting that the absorbance spectra do differ noticeably at wavelengths longer than ≈ 540 nm. The negative value of the absorbance is due to fluorescence emission from the sample. We believe the difference in negative magnitude of the absorbance to be due to differences in light collection by the Nanodrop and the micro-spectrophotometer. The Nanodrop collects light directly

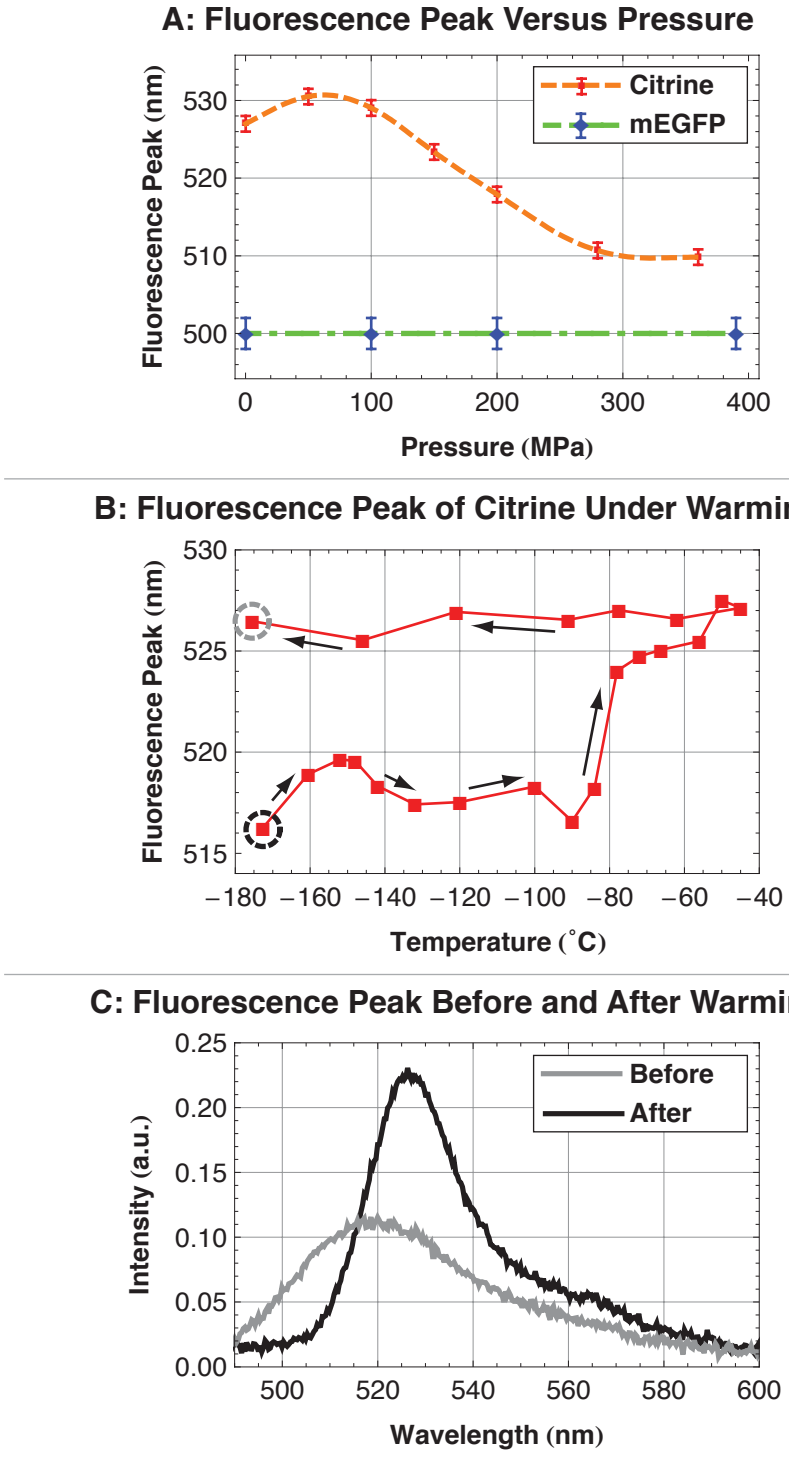
through an optical fiber, while the micro-spectrophotometer uses a large-area reflecting objective that will collect a large amount of fluorescence light, resulting in a more negative value of the absorbance at wavelengths exceeding 540 nm, where the absorbance of Citrine is very low.

4.5 The Fluorescence Spectrum of Citrine Under High Pressure Cryo-Cooling Conditions

Due the difficulties in collecting room temperature X-ray diffraction data from crystals of Citrine, and thus solving the room temperature high-pressure structure of Citrine, it was important to establish a direct link between the high-pressure cryo-cooled X-ray structures of Citrine and the fluorescence spectra of high-pressure cryo-cooled Citrine samples. The fluorescence spectra of high-pressure cryo-cooled Citrine samples were measured with a micro-spectrophotometer based upon the designs by Hadfield and Hajdu [178] and Klink *et al.* [179]. The construction of this device and the preparation of high-pressure cryocooled Citrine solution samples in polycarbonate capillaries was described in section 4.3.

The fluorescence peak of high-pressure cryo-cooled Citrine solutions shifts from 527 nm when frozen at ambient pressure, to 530 nm at 50 MPa, and to 510 nm at 360 MPa with most of the shift occurring by 250 MPa. A plot of the peak shift of the Citrine samples versus freezing pressure is shown in figure 4.37A. As a control, the peak of high-pressure cryo-cooled monomeric Enhanced Green Fluorescent Protein (mEGFP) samples are also shown in figure 4.37A. mEGFP contains a similar main chromophore to Citrine, but lacks the perturbing tyrosine 203 ring. In addition, mEGFP, and all other monomeric *Aequorea* fluorescent proteins, contain the surface mutation A206K that strongly discourages dimerization [180]. The fluorescence peak of mEGFP does

Figure 4.37: A: (■) Citrine's fluorescence peak under high-pressure cryocooling. Connecting line drawn to aid eye. (◆) mEGFP's fluorescence peak under identical conditions. Fit shown as green dot-dashed line. B: Fluorescence peak of Citrine solution in capillary pressurized to 200 MPa as it is warmed. Note dashed circles surrounding start (gray) and end (black) points of curve, corresponding to spectra shown in C: spectra of high pressure cryo-cooled Citrine solution before and after warming.


A: Fluorescence Peak Versus Pressure
Fluorescence Peak (nm)
530
520
510
500
0
100
200
300
400
Pressure (MPa)
Citrine
mEGFP
B: Fluorescence Peak of Citrine Under Warming
Fluorescence Peak (nm)
530
525
520
515
−180
−160
−140
−120
−100
−80
−60
−40
Temperature (˚C)
C: Fluorescence Peak Before and After Warming
Intensity (a.u.)
0.25
0.20
0.15
0.10
0.05
0.00
500
520
540
560
580
600
Wavelength (nm)
Before
After

not shift as function of the freezing pressure. The fluorescence peak of conventional EGFP also does not shift as a function of pressure at room temperature [53].

The fluorescence peak shift of high-pressure cryocooled Citrine from yellow to green, and the insensitivity of the fluorescence peak of EGFP to pressure, suggests that the perturbing interaction of tyrosine 203 is slowly removed by the application of high pressure.

It is interesting to note that the fluorescence peak of Citrine approaches ≈ 510 nm at 350 MPa (it should be noted that we have no evidence that the peak asymptotes to this value). This peak fluorescence wavelength is approximately the same wavelength as that of mEGFP at room temperature (511 nm), suggesting that the orbitals of the Citrine main chromophore are returning to an unperturbed EGFP-like state as the perturbation from the tyrosine 203 phenol ring is removed.

As is the case with high-pressure cryocooled crystals of Citrine (figure 4.20), the fluorescence peak of high-pressure cryocooled Citrine relaxes to yellow upon warming of the sample above a critical temperature (≈ 180 K), and does not return to its original position upon re-cooling, suggesting that the original fluorescence shift to the green had been locked in by the high-pressure cryocooling procedure.

The fluorescence peak of a high-pressure cryo-cooled solution that was slowly warmed and re-cooled is shown as function of temperature in figure 4.37B. The fluorescence peak remains constant until a critical temperature of ≈ -80 ˚C, and then suddenly shifts to the blue. This critical temperature at which the fluorescence spectrum suddenly relaxes is the same as that observed in high-pressure cryocooled

Citrine crystals (figure 4.20). The fluorescence spectra of a high-pressure cryo-cooled sample before and after warming above the critical temperature are shown in figure 4.37C.

The shift of the fluorescence peak of Citrine to the blue was anticipated from early structures of the Citrine molecule under high pressure and the simple quantum chemical model of the Citrine fluorescence spectrum that was presented in section 3.6.4. Early structures of Citrine under high-pressure cryocooling at 192 MPa indicated that the main chromophore and the perturbing tyrosine 203 ring separated at high pressure. The quantum chemical model of the Citrine fluorescence spectrum indicates that as the perturbing tyrosine 203 ring separates from the main chromophore, the fluorescence band-gap should increase, causing the fluorescence emission from the chromophore to shift to the blue.

The fluorescence spectra of high-pressure cryocooled Citrine solutions shows that the fluorescence peak of Citrine first shifts slightly to the red with increasing freezing pressure, and then shifts to the blue at freezing pressures above 500 MPa. The fluorescence peak of Citrine shifts to 510 nm by a freezing pressure of 350 MPa, approximately the same fluorescence peak as mEGFP at room temperature, suggesting that the perturbing interaction of tyrosine 203 has been largely removed by this pressure. This shift of the fluorescence spectrum to the blue is consistent with computer models of the Citrine fluorescence spectrum in which the main chromophore and tyrosine 203 are spatially separated.

4.6 Crystallization of Citrine

4.6.1 Hanging Drop Crystallization of Citrine

Citrine was first crystallized using a batch of Citrine that was expressed from the pRSETB vector provided by Professor Roger Tsien. Initially, attempts to crystallize Citrine used a close approximation of the crystallization conditions reported by Griesbeck *et al*. [1]: 50 mM sodium acetate, 50 mM ammonium acetate, 7% w/v PEG 3400, pH 5.0 at 4 ˚C. PEG 3400 appears to be currently commercially unavailable, and PEG 3350 (Lot number 259134, Catalog number HR2-591, Hampton Research, Aliso Viejo, CA, USA was used as a substitute. Unfortunately, early attempts to reproduce these conditions failed to yield any crystals of Citrine.

Although alternative Yellow Fluorescent Protein conditions do exist [52], the crystallization conditions reported by Griesbeck *et al*. [1] produced crystals which displayed the highest diffraction resolution. Attempts to reproduce the Yellow Fluorescent Protein crystallization experiments by Wachter *et al*. [52] also failed, and it was decided that persistence should by applied to reproducing the crystallization experiments by Griesbeck *et al*. [1].

A large crystallization trial, sampling a wide range of crystallization conditions surrounding those reported by Griesbeck *et al*. [1] was performed. This crystallization trial sampled precipitant (PEG 3350) concentrations from 3% to 15%, and from pH 3.0 to pH 9.0. The sodium acetate and ammonium acetate concentrations were both fixed at 50 mM. In total, approximately 70 individually hand-made crystallization solutions were tried. Citrine concentrations of 10 mg/mL and 20 mg/mL were used. Hanging drops of 1 μL protein, and 1 μL well solution were prepared.

For safekeeping, and a certain degree of psychological comfort on the part of the experimenter, it was decided to store these crystals in a 4 ˚C cold room rather than the refrigerator normally used to store protein crystals in the Gruner lab. Prior to storing these crystallization trays in the cold room, the crystallization trays were stored in two large Styrofoam boxes. The crystallization trays were left undisturbed for 5 days, the duration of time reported for crystal nucleation by Griesbeck *et al*. [1]. After 7 days, a single Citrine crystal was found at 11% PEG 3350, pH 4.5. This crystal can be seen in figure 2.6A. This crystal served as a seed for all following Citrine crystals. I speculate the difference in crystallization experience between myself and Griesbeck *et al*. [1] may be due to differences in the precipitant used in our respective experiments. It seems unlikely, although not entirely implausible, that the small difference in polymer molecular weight (3350 versus 3400 Da) would greatly decrease the precipitating power of PEG 3350 when compared to PEG 3400. However, the composition of PEGs is known to vary greatly from manufacturer to manufacturer and even from batch to batch [95]. For this reason, all of our Citrine crystallization experiments were performed with a single batch of PEG 3350 (Lot number 259134, Catalog number HR2-591, Hampton Research, Aliso Viejo, CA, USA). I did attempt crystallization with another batch of PEG 3350 (Lot number 259147, Hampton Research), and found no difference in crystallization experience. It is possible that the PEG 3400 used by Griesbeck *et al.* [1] contained a contaminant not present in the Hampton PEG 3350 that promoted nucleation of Citrine crystals. I asked Robert Campbell who performed the crystallization experiments presented in the report on Citrine by Griesbeck *et al*. [1] (and one of the co-authors) where he found the PEG 3400 used to crystallize Citrine, and he replied “off the shelf”, literally meaning off a shelf.

Later attempts to reproduce the successful Citrine crystallization experiment without the Styrofoam outer box failed. I believe that the box is crucial for success of the crystallization trial, as it prevents the exterior temperature of the cover-slide from dropping too rapidly when the crystallization tray is placed in the refrigerator and causing excess condensation of vapor from the well onto the cover-slide, causing dilution of the protein droplet, and preventing precipitation of the protein.

Streak seeding was used to produce a second generation of Citrine crystals. In streak seeding, the side of a crystal is tapped with a fine needle (Hampton Research, Aliso Viejo, CA, USA; sells a product called the Crystal Probe (Catalog number HR4-217) for this purpose), to dislodge crystal seeds from its surface. Some of these seeds will attach to the needle. The needle is then run through a hanging droplet containing a mixture of protein and well solution. Some of the crystal seeds that were attached to the needle will be deposited along the trace or streak of the needle in the new droplet. Crystals typically form in a line that marks the trace of the needle through the droplet. These second generation crystals were produced at 7% w/v PEG 3350, pH 4.5, 50 mM sodium acetate, 50 mM ammonium acetate. A third generation of crystals was produced by micro-seeding with a Seed Bead (Catalog number HR2-320, Hampton Research, Aliso Viejo, CA, USA) from these crystals. This third generation of crystals was the best batch of Citrine crystals that was produced in this thesis work, showing diffraction out to at least 1.3 Å.

At this time, the Citrine supply from pRSETB was depleted, and we moved to a new Citrine purification batch grown from a pET-24 (Novagen, a division of Merck KGaA, Darmstadt, Germany) vector. This batch of protein showed no problems with nucleation; however, it produced a plate like form of the Citrine crystal. These plate-

like crystals were mechanically fragile, and were later shown to diffract extremely poorly. To discourage the formation of these plate-like crystals, seeds were derived from the original generation of brick-like crystals. A photograph of the plate-like Citrine crystals is shown in figure 4.1. Although the formation of the plate-like crystals could not be completely eliminated, the crystals could be partially eliminated by leaving a sealed hanging droplet crystallization well at room temperature for several hours. The plate-like Citrine crystals appear to be much less stable at room temperature than the brick-like crystals, and largely dissolve over the course of several hours. At this time, the crystallization well may be placed back into the 4 ˚C refrigerator. Although some plate-like crystals will reappear, they will be far less numerous than before the exposure to room temperature.

Further crystals of Citrine were grown by micro-seeding, using a 5% PEG 3350, 50 mM sodium acetate, 50 mM ammonium acetate, pH 5.0 precipitant solution, using a Seed Bead (Catalog number HR2-320, Hampton Research, Aliso Viejo, CA, USA). All crystallization solutions were stored at 4 ˚C, but were allowed to warm to room temperature before proceeding. No more than 12 crystal wells were prepared at a single time. 60 µL of precipitant, and 60 µL of 10 mg/mL Citrine in 50 mM HEPES, pH 7.5, were mixed in a Seed Bead, and then 10 µL of this mixture was dispensed onto the cap of the Seed Bead. A brick-like crystal of Citrine was transferred to the cap of the centrifuge tube, and the Seed Bead was then centrifuged in a centrifuge pre-cooled to 4 ˚C for ≈ 10 seconds while accelerating to 14,000 × *g*, to remove the droplet from the cap. The Seed Bead was then vortexed for approximately $^1/_4$ of a second, making sure that the bead moved vigorously around inside the Seed Bead. The Seed Bead was then re-centrifuged to pellet the solution, and immediately transferred to ice. The seeding solution was then dispensed onto cover-slides in 10 µL droplets, and each

slide sealed onto a well before proceeding to the next well. For these crystallization experiments, Nextal Biotech (Montreal, Quebec, Canada. Nextal Biotech is now a subsidiary of Qiagen) single screw cap crystallization wells were used (unfortunately, these single wells are no longer available from Nextal Biotech, however, replacement screw caps are still available under part number 132073, and the individual wells may be cleaned and reused). The Seed Bead was chilled on ice between dispensations, and the whole procedure lasted no longer than 5 minutes.

The sealed wells were immediately transferred to a Styrofoam box and refrigerated at 4 ˚C. Crystals appeared overnight, and grew to approximately 300 x 50 x 50 µm within a few days.

At least three forms of Citrine crystals will grow under these conditions; bricks, with the most equal dimensions, plates with two long dimensions, and a very thin third dimension, and rods with one long dimension and two very short dimensions. The plate form of the Citrine crystal does not appear to diffract well, and has imperfect symmetry, only indexing well in a P2 or P1 space-group. The rod crystals have not been tested.

4.6.2 Crystallization of Citrine Crystals in Polycarbonate Capillaries

While experimenting with solving the structure of Citrine under high pressure at room temperature, it was found that Citrine crystals did not react well to immobilization in Sephadex, the method commonly used to immobilize crystals inside the beryllium cell, described by Kundrot and Richards [23], Urayama [70] and Collins [181]. Thus, a different method of immobilizing the crystals inside the high-pressure beryllium cell was sought. One method that was explored was to grow Citrine crystals inside

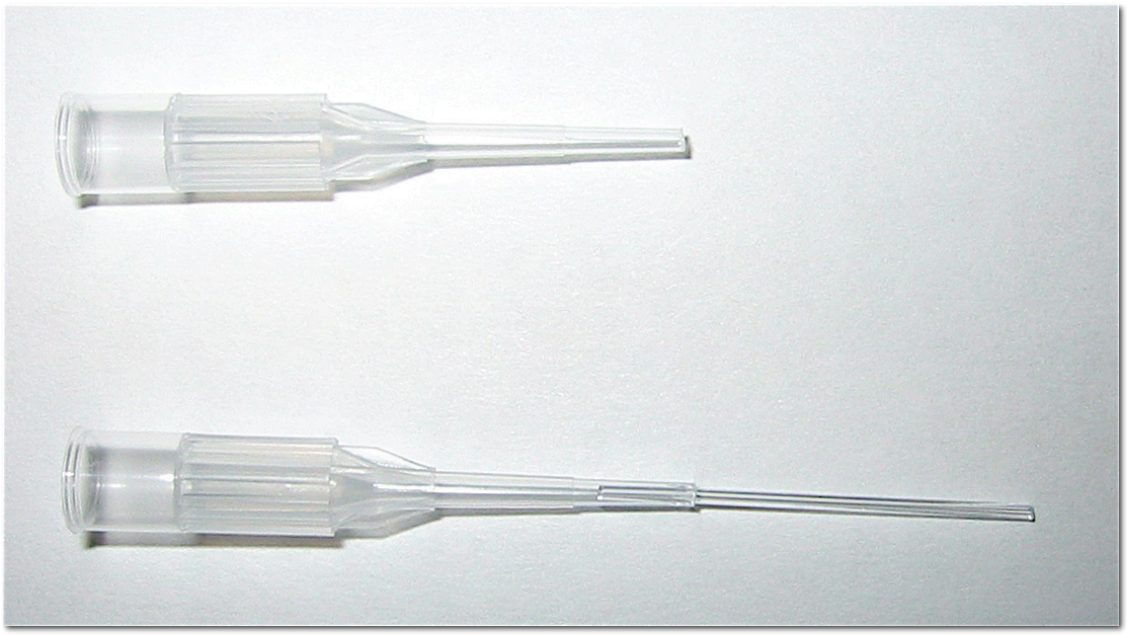


Figure 4.38: Modified 10 µL Rainin LTS pipet tips for use with polycarbonate capillaries.

polycarbonate capillaries, as suggested by recent experiments by Chae Un Kim, who was working on freezing Thaumatin crystals grown in capillaries for the NIH Protein Structure Initiative. The Citrine crystallization procedure was modified slightly in order to accommodate the different vapor exchange rates found in a capillary crystallization trial compared with those found in a hanging drop crystallization trial.

The most successful capillary crystallization procedure initially followed the hanging drop crystallization procedure described in section 4.5.1. Micro-seeded hanging drops were prepared in Nextal screw cap wells as described in section 4.5.1. These droplets were left to equilibrate for between 16 and 40 hours, and were examined for signs of crystal formation every 8 hours. When crystal formation was noticed, characterized by a slight shimmer on the surface of the hanging droplet, the hanging droplets were drawn up into polycarbonate capillaries (Plastic part number 8-000-1000, length 5 inches (127 mm), inner diameter 0.012 inches (304.8 µm), outer diameter 0.036 inches

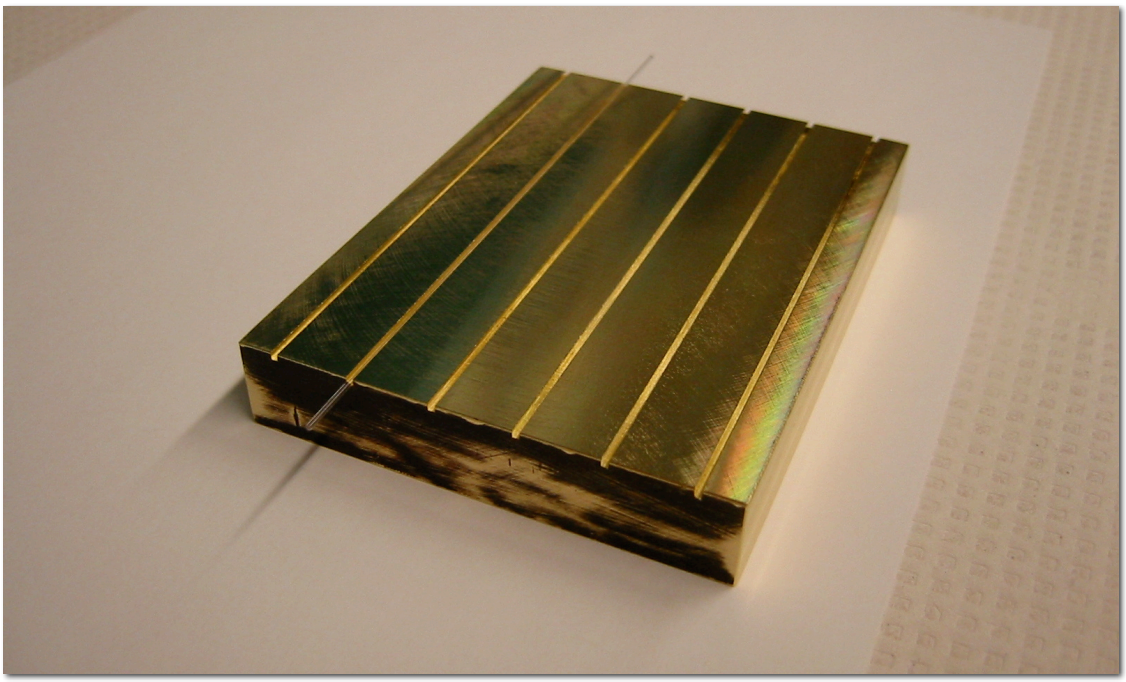


Figure 4.39: Polycarbonate capillary cooling block. The block was fabricated by Martin Novak.

Figure 4.40: 10 mL pipette tip used for storing Citrine crystals grown in polycarbonate capillaries.

(914.4 µm), Drummond Scientific Company, Broomall, PA, USA) using a modified 10 µL Rainin LTS pipette tip (RT-L10, Rainin Instruments, Woburn MA, USA) (figure 4.38) attached to a 10 µL Rainin LTS pipettor. Immediately after a hanging drop was drawn up into a capillary, the pipette tip holding the capillary was gently pierced with the point of a sharp X-acto blade to allow air pressure equilibration between the inside

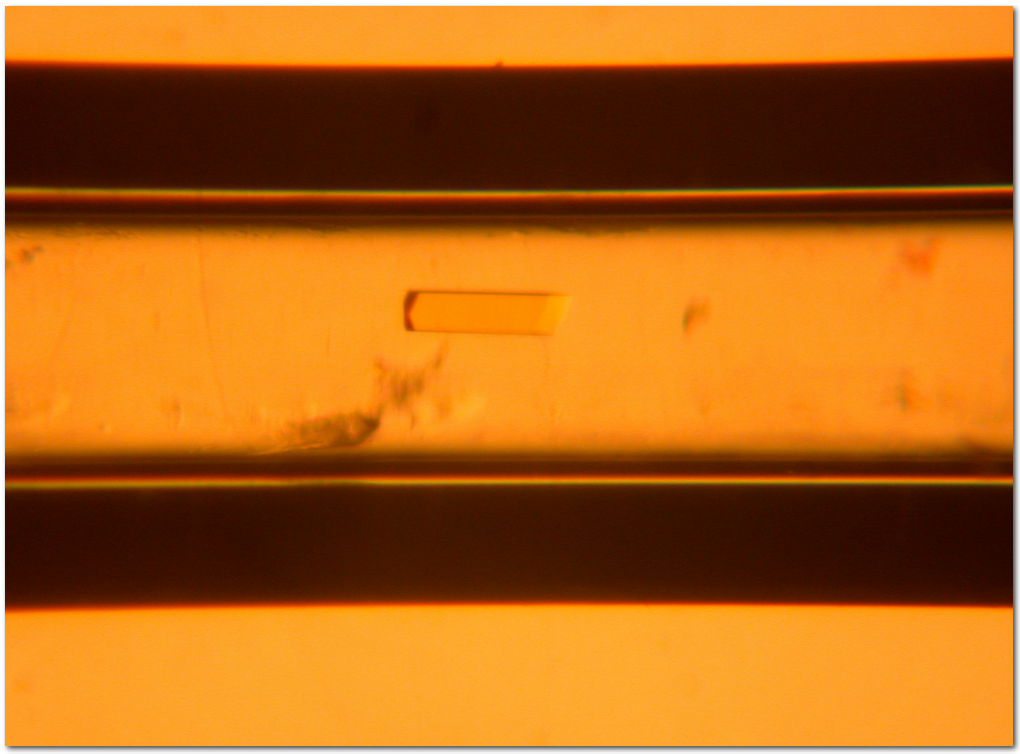


Figure 4.41: Citrine crystal grown in a polycarbonate capillary.

and outside of the tip (pulling the tip off the pipettor would result in discharge of the capillary contents), and the pipette tip was gently removed. Finally, the capillary was gently removed from the pipette tip with curved micro-tweezers, and immediately placed on a pre-cooled brass block with fine channels for the capillaries (shown in figure 4.39). No more than 6 capillaries were prepared at once.

Immediately after finishing the last capillary, all 6 capillaries were transferred to a 10 mL pipette tip (Catalog number RC-L10ML, Rainin Instruments, Woburn, MA, USA) (figure 4.40). A cotton wool ball soaked in mother liquor (5% PEG 3350, 50 mM sodium acetate, 50 mM ammonium acetate, pH 5.0) was placed into the capillary, and both ends were sealed with Parafilm. The pipette tip was placed inside a Styrofoam box and stored at 4 ˚C. The procedure yielded sizable crystals within several days.

The crystals that grew by this procedure attached to the wall of the capillary, and appeared to remain attached to the wall even under pressures of up to 200 MPa. A photograph of a brick-like crystal in shown in figure 4.41. Unfortunately, these crystals suffered cracking and dissolution under high pressure, and radiation damage. A photograph of a capillary-grown crystal at CHESS station F1, showing the signs of radiation damage can be see in figure 4.42.

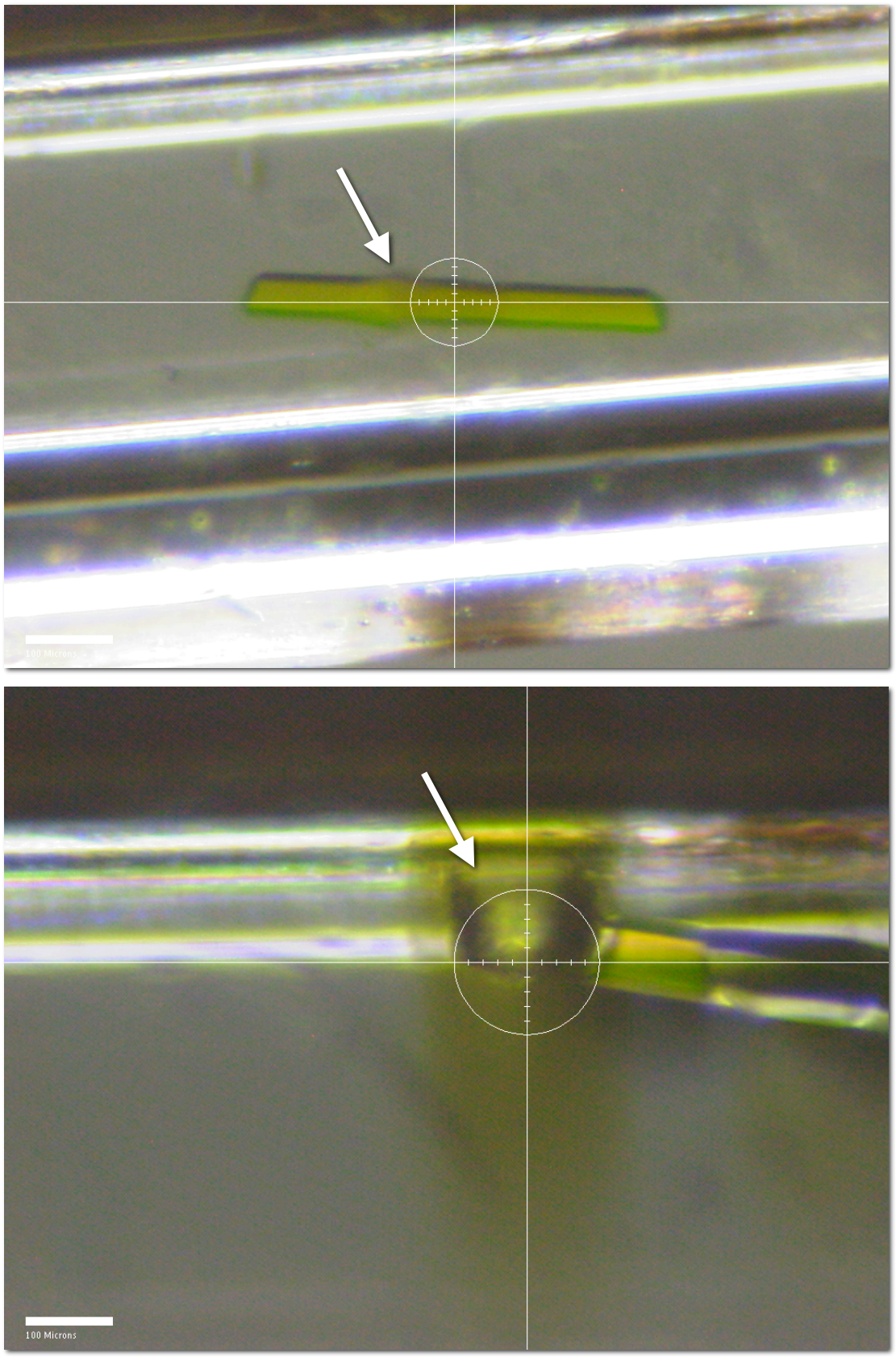
4.7 X-ray Diffraction Data Collection

To investigate the structural origin of Citrine's fluorescence shift, the high-pressure cryo-cooling technique developed by Kim *et al*. [3], and described in chapter 2, was used to prepare Citrine crystals, at cooling pressures ranging from 50 MPa to 500 MPa. A consequence of this preparation method is that each structure at each pressure is derived from a different crystal.

X-ray diffraction data were collected at CHESS (Cornell High Energy Synchrotron Source, Ithaca, NY) macromolecular crystallography station F2 at an X-ray wavelength of 0.9795 Å (the selenium edge), with a 0.1 mm X-ray collimator using a Quantum 210 CCD detector (Area Detector Systems Corporation, Poway, CA, USA). For data-sets taken in April and November of 2007, the detector used an upgraded computer for image processing. The datasets were collected between May 2005 and November 2007. In total, approximately 50 diffraction datasets were collected, of which approximately 30 were of suitable quality.

Crystals were transferred by hand from a liquid nitrogen storage dewar to a cryogenic nitrogen stream at 100 K (Oxford Cryo-systems, Oxford, UK). Care should be taken while performing the transfer as even slight warming of the sample will allow the

Figure 4.42: Citrine crystal grown in a polycarbonate capillary at CHESS beamline F1. The upper image shows the early signs of radiation damage, indicated by the white arrow. The lower image shows the crystal after significant radiation damage has accumulated. The area of maximum damage is indicated by a white arrow in the lower image. The circle diameter in the top image is 100 μm. In the lower image the circle diameter is 160 μm.


100 Microns

pressure-induced deformations in the sample to relax, and release the helium gas trapped in the crystal.

Typically, between 120 and 180 oscillation images were taken from each crystal, with an angular separation of 1˚, and a 1˚ oscillation angle, with exposure times ranging from 30 to 60 seconds. Full data collection details are shown in table 4.3.

The X-ray diffraction data was indexed with Rossmann and van Beek's [99] Data Processing Suite (DPS) algorithm, integrated with MOSFLM [100], scaled with SCALA [102] and truncated with TRUNCATE [103]. Indexing quality indicators for each dataset are shown in tables 4.4 and 4.5. The indexing quality indicators shown in table 4.4 are for indexing out to the resolution ring where $\langle I \rangle / \langle \sigma \rangle = 3.0$, and those in table 4.4 are for indexing out to the maximum observable resolution ring.

4.8 The Structure of the Citrine Chromophore Under High Pressure Cryo-Cooling

The crystal structures of Citrine under high-pressure cryocooling were examined to find evidence for deformations of the chromophore in response to pressure that would explain the blue-shift of the fluorescence spectrum that was discussed in section 4.4.

The earliest structures of Citrine under high-pressure cryocooling conditions showed a separation of the main chromophore and perturbing tyrosine 203 phenol ring. The main chromophore translated by approximately 0.6 Å in a plane parallel to the plane of the tyrosine 203 phenol under a pressure of 192 MPa. This sliding deformation motion was supported by the quantum chemical model of the Citrine fluorescence peak that was discussed in section 3.6.4 that suggested that the fluorescence peak of

Table 4.3: Crystallographic data collection information. All datasets were taken at CHESS (Cornell High Energy Synchrotron Source) macromolecular crystallography station F2 with a Quantum 210 CCD detector (ADSC, Poway, CA, USA). All datasets were taken using a 0.1 mm X-ray collimator with an X-ray wavelength of 0.9795 Å. Datasets Citrine0001_10 and Citrine2500_1 to Citrine5000_3 were taken following an upgrade to this detector. All datasets were taken with an oscillation angle of 1˚, and a step of 1˚, except CitrineFF1 and CitrineFF2 which were taken with a 3˚ oscillation angle and step.

Crystal	Pressure (MPa)	Run	Distance (mm)	Exposure Time (s)	Frames	Mosaicity (°)
CitrineFF1	0.1	1	200.67	90	30	0.496
CitrineFF2	0.1	1	201.17	90	30	0.74
CitrineOFF2	0.1	1	200.42	25	120	0.538
Citrine0001_10	0.1	1	170.43	30	190	0.402
Citrine0500_1	50	1	199.35	60	120	0.527
Citrine0500_2	50	1	169.37	60	120	0.668
Citrine0500_3	50	1	194.7	30	180	0.546
Citrine0750_2	75	1	169.49	60	120	0.593
Citrine0750_4	75	1	194.71	60	120	1.038
Citrine1000_1	100	1	169.35	60	120	0.61
Citrine1000_2	100	1	169.3	60	120	0.551
Citrine1000_3	100	1	169.28	60	180	0.872
Citrine1000_5	100	1	194.89	30	180	0.912
Citrine1000_7	100	1	194.83	30	180	0.365
Citrine1000_8	100	1	195.35	30	105	0.524
Citrine1250_1	125	1	169.57	60	120	0.402
Citrine1250_2	125	1	169.09	60	120	0.912
Citrine1250_3	125	1	169.4	60	120	0.719
Citrine1250_4	125	1	179.89	25	200	0.672
Citrine1500_1	150	1	169.37	60	120	0.598
Citrine1500_2	150	1	169.3	60	120	0.843
CitrineOPF2	192	1	200.82	60	10	0.078
	192	2	160.84	60	7	0.113
	192	3	140.77	60	120	0.159
	192	4	140.71	60	120	0.125
CitrineOPF3	192	1	200.62	60	90	0.162
	192	2	199.66	60	90	0.086
Citrine1960_2	196	1	179.34	30	120	0.685
Citrine2000_2	200	1	195.05	30	120	0.566
	200	2	194.91	30	120	0.481
Citrine2000_3	200	1	194.95	60	151	0.44
	200	2	194.89	30	120	0.412
Citrine2500_1	250	1	197.71	20	120	0.291
Citrine4000_1	400	1	177.49	45	180	0.362
Citrine4000_2	400	1	177.67	30	180	0.904
Citrine4000_3	400	1	177.54	30	180	0.632
Citrine5000_3	500	1	167.75	30	180	0.554

Table 4.4: Crystallographic indexing quality indicators for the X-ray diffraction datasets presented in table 4.3 when indexed out to resolution ring with $\langle I \rangle / \langle \sigma \rangle = 3.0$. Indicators are given for the overall (OA) dataset and for the final resolution shell (FS) of the dataset. All datasets were indexed using Rossmann and van Beek's Data Processing Suite [27] algorithm and scaled using the CCP4 program Scala [29].

Dataset	P (MPa)	R_{Merge}		R_{Meas}		Total Obs.		Unique Obs.		$\langle I/\sigma \rangle$		Comp.		Mult.		Res. (Å)
		OA	FS	OA	FS	OA	FS	OA	FS	OA	FS	OA	FS	OA	FS	
CitrineFF1	0.1	0.078	0.247	0.091	0.291	49882	6947	14127	2015	13.8	5.2	99.6	99.9	3.5	3.4	2.07
CitrineOFF2	0.1	0.104	0.266	0.118	0.3	38506	5656	8286	1190	13.2	6.6	96.8	98.1	4.6	4.8	2.46
Citrine0001_10	0.1	0.072	0.253	0.078	0.276	156833	19104	21406	2994	21	6.1	98.8	96.2	7.3	6.4	1.8
Citrine0500_1	50	0.078	0.291	0.088	0.327	58588	8522	12169	1786	16.4	6.5	96.4	98.4	4.8	4.8	2.17
Citrine0500_2	50	0.062	0.339	0.07	0.38	92045	13407	19193	2737	20.1	4.8	97.9	96.6	4.8	4.9	1.87
Citrine0500_3	50	0.062	0.301	0.067	0.329	143182	18197	20518	2932	22.7	5.4	99.9	99.7	7	6.2	1.84
Citrine0750_2	75	0.056	0.3	0.063	0.337	99597	14165	20832	3034	19.3	4.3	98.6	99.7	4.8	4.7	1.82
Citrine0750_4	75	0.071	0.293	0.08	0.331	69765	10118	15261	2165	16.7	4.9	100	100	4.6	4.7	2.04
Citrine1000_1	100	0.072	0.287	0.081	0.322	79351	11489	16780	2412	17.8	5	99.3	99.9	4.7	4.8	1.97
Citrine1000_2	100	0.079	0.325	0.089	0.364	85903	12501	18149	2581	17.1	4.6	99.5	99.3	4.7	4.8	1.92
Citrine1000_3	100	0.087	0.273	0.094	0.295	107660	15551	15568	2219	20.9	6.1	100	100	6.9	7	2.03
Citrine1000_5	100	0.062	0.292	0.067	0.315	112129	16377	16109	2310	24.5	6.4	100	100	7	7.1	2.01
Citrine1000_7	100	0.058	0.271	0.063	0.307	165929	11813	25494	2831	22.8	4	94.9	73.9	6.5	4.2	1.67
Citrine1000_8	100	0.071	0.26	0.081	0.303	79760	10431	19594	2807	15.6	4.4	99.9	99.8	4.1	3.7	1.86
Citrine1250_1	125	0.061	0.306	0.069	0.343	97386	14128	20246	2933	18.4	4.7	98.9	99.9	4.8	4.8	1.84
Citrine1250_2	125	0.064	0.303	0.072	0.343	70155	10212	15203	2169	18.9	4.6	99.9	100	4.6	4.7	2.04
Citrine1250_3	125	0.066	0.287	0.074	0.32	72130	10386	15091	2178	18.3	5.1	98.3	99.5	4.8	4.8	2.04
Citrine1250_4	125	0.06	0.342	0.064	0.366	159534	23049	20337	2907	26.5	5.9	100	100	7.8	7.9	1.86
Citrine1500_1	150	0.05	0.124	0.056	0.139	71483	10441	14686	2145	23.3	11.1	97	986	4.9	4.9	1.78
Citrine1500_2	150	0.068	0.307	0.075	0.341	70320	10237	13736	2027	19.6	5.4	91.5	93.8	5.1	5.1	2.04
CitrineOPF2	192	0.09	0.236	0.097	0.255	228033	32119	30481	4414	23.9	5.4	99.9	100	7.5	7.3	1.58
CitrineOPF3	192	0.085	0.239	0.092	0.28	126611	7074	20560	2050	19.9	3.3	92.4	64.7	6.2	3.5	1.75
Citrine1960_2	196	0.051	0.294	0.058	0.334	93837	12557	20204	2875	21.2	5.2	99.9	99.4	4.6	4.4	1.86
Citrine2000_2	200	0.07	0.303	0.074	0.325	194213	23902	21016	2987	27.6	5.3	100	99.8	9.2	8	1.83
Citrine2000_3	200	0.094	0.3	0.099	0.315	186809	27088	17459	2488	25.4	6.9	100	100	10.7	10.9	1.95
Citrine2500_1	250	0.057	0.263	0.064	0.313	89750	5932	20485	2212	17.7	3.1	91.9	69.5	4.4	2.7	1.78
Citrine4000_1	400	0.068	0.276	0.074	0.302	144633	18401	20822	2963	21.4	5.7	99.8	99.4	6.9	6.2	1.82
Citrine4000_2	400	0.089	0.274	0.097	0.296	70739	10304	10412	1472	20.1	7	100	100	6.8	7	2.25
Citrine4000_3	400	0.069	0.259	0.076	0.28	104036	15283	15772	2254	22.2	7.3	100	99.9	6.6	6.8	2.02
Citrine5000_3	500	0.07	0.269	0.075	0.29	144252	20256	20121	2843	23.7	7.6	98	96.8	7.2	7.1	1.8

Table 4.5: Crystallographic indexing quality indicators for the datasets first shown in table 4.3 when indexed out to maximum detectable resolution ring. Indicators are given for the overall (OA) dataset and for the final resolution shell (FS) of the dataset. All datasets were indexed using Rossmann and van Beek's Data Processing Suite [99] algorithm and scaled using the CCP4 program SCALA [102].

Dataset	P (MPa)	R_{Merge}		R_{Meas}		Total Obs.		Unique Obs.		$\langle I/\sigma \rangle$		Comp.		Mult.		Res. (Å)
		OA	FS	OA	FS	OA	FS	OA	FS	OA	FS	OA	FS	OA	FS	
CitrineFF1	0.1	0.078	0.247	0.091	0.291	49882	6947	14127	2015	13.8	5.2	99.6	99.9	3.5	3.4	2.07
CitrineOFF2	0.1	0.137	3.347	0.156	4.689	91145	944	25427	838	6.8	0.1	76.1	17.8	3.6	1.1	1.55
Citrine0001_10	0.1	0.074	0.797	0.087	0.985	98198	4314	28646	1830	12	1	77.9	35	3.4	2.4	1.5
Citrine0500_1	50	0.078	0.291	0.088	0.327	58588	8522	12169	1786	16.4	6.5	96.4	98.4	4.8	4.8	2.17
Citrine0500_2	50	0.062	0.339	0.07	0.38	92045	13407	19193	2737	20.1	4.8	97.9	96.6	4.8	4.9	1.87
Citrine0500_3	75	0.062	0.301	0.067	0.329	143182	18197	20518	2932	22.7	5.4	99.9	99.7	7	6.2	1.84
Citrine0750_2	75	0.056	0.3	0.063	0.337	99597	14165	20832	3034	19.3	4.3	98.6	99.7	4.8	4.7	1.82
Citrine0750_4	75	0.071	0.293	0.08	0.331	69765	10118	15261	2165	16.7	4.9	100	100	4.6	4.7	2.04
Citrine1000_1	100	0.072	0.287	0.081	0.322	79351	11489	16780	2412	17.8	5	99.3	99.9	4.7	4.8	1.97
Citrine1000_2	100	0.079	0.325	0.089	0.364	85903	12501	18149	2581	17.1	4.6	99.5	99.3	4.7	4.8	1.92
Citrine1000_3	100	0.087	0.273	0.094	0.295	107660	15551	15568	2219	20.9	6.1	100	100	6.9	7	2.03
Citrine1000_5	100	0.062	0.292	0.067	0.315	112129	16377	16109	2310	24.5	6.4	100	100	7	7.1	2.01
Citrine1000_7	100	0.058	0.271	0.063	0.307	165929	11813	25494	2831	22.8	4	94.9	73.9	6.5	4.2	1.67
Citrine1000_8	100	0.071	0.26	0.081	0.303	79760	10431	19594	2807	15.6	4.4	99.9	99.8	4.1	3.7	1.86
Citrine1250_1	100	0.061	0.306	0.069	0.343	97386	14128	20246	2933	18.4	4.7	98.9	99.9	4.8	4.8	1.84
Citrine1250_2	125	0.064	0.303	0.072	0.343	70155	10212	15203	2169	18.9	4.6	99.9	100	4.6	4.7	2.04
Citrine1250_3	125	0.066	0.287	0.074	0.32	72130	10386	15091	2178	18.3	5.1	98.3	99.5	4.8	4.8	2.04
Citrine1250_4	125	0.06	0.342	0.064	0.366	159534	23049	20337	2907	26.5	5.9	100	100	7.8	7.9	1.86
Citrine1500_1	150	0.05	0.124	0.056	0.139	71483	10441	14686	2145	23.3	11.1	97	98.6	4.9	4.9	2.04
Citrine1500_2	150	0.068	0.307	0.075	0.341	70320	10237	13736	2027	19.6	5.4	91.5	93.8	5.1	5.1	2.04
CitrineOPF2	192	0.09	0.236	0.097	0.255	228033	32119	30481	4414	23.9	5.4	99.9	100	7.5	7.3	1.58
CitrineOPF3	192	0.085	0.239	0.092	0.28	126611	7074	20560	2050	19.9	3.3	92.4	64.7	6.2	3.5	1.75
Citrine1960_2	196	0.051	0.294	0.058	0.334	93837	12557	20204	2875	21.2	5.2	99.9	99.4	4.6	4.4	1.86
Citrine2000_2	200	0.07	0.303	0.074	0.325	194213	23902	21016	2987	27.6	5.3	100	99.8	9.2	8	1.83
Citrine2000_3	200	0.173	0.618	0.187	0.678	107483	11718	15656	1998	13.9	3.6	99.6	98.7	6.9	5.9	1.99
Citrine2500_1	250	0.057	0.263	0.064	0.313	89750	5932	20485	2212	17.7	3.1	91.9	69.5	4.4	2.7	1.78
Citrine4000_1	400	0.068	0.276	0.074	0.302	144633	18401	20822	2963	21.4	5.7	99.8	99.4	6.9	6.2	1.82
Citrine4000_2	400	0.089	0.274	0.097	0.296	70739	10304	10412	1472	20.1	7	100	100	6.8	7	2.25
Citrine4000_3	400	0.069	0.259	0.076	0.28	104036	15283	15772	2254	22.2	7.3	100	99.9	6.6	6.8	2.02
Citrine5000_3	500	0.07	0.269	0.075	0.29	144252	20256	20121	2843	23.7	7.6	98	96.8	7.2	7.1	1.8

Citrine should shift to the blue under this deformation motion, and the observed shift of the Citrine fluorescence peak to the blue under high-pressure cryocooling conditions. However, it was felt that due to the small magnitude of this deformation (≈ 0.6 Å), careful refinement of a series of structure was required to definitively identify this deformation motion. Additionally, we desired multiple structures at each pressure in order to estimate the positional error at each pressure level.

Although in retrospect, the precision estimate from Cruickshank's formula, discussed in section 2.8, indicates that a 0.6 Å deformation motion should be readily identifiable by X-ray crystallography, we remained skeptical of this result, and spent considerable time analyzing the structure of Citrine under high-pressure for evidence of systematic error.

The results of two crystallographic refinement procedures are presented in this thesis, and can be compared to estimate some of the uncertainty on atomic coordinates that result from different refinement procedures. One of these methods was used to produce the set of high-pressure atomic models of Citrine that were presented in our article on the deformation of the Citrine chromophore and the accompanying fluorescence peak shift [5]. The X-ray diffraction datasets used to derive each Citrine atomic model are listed in table 4.6. The unit cell axes for each model are shown in table 4.7. The second refinement procedure was used to produce the set of atomic models that were deposited in the Protein Data Bank. The PDB accession codes of the models are listed in table 4.8.

Table 4.6: X-ray diffraction datasets used to derive each Citrine atomic model.

Model Name	Datasets Used
Citrine0001_2	CitrineOFF2
Citrine0001_18	Citrine0001_10
Citrine0500_9	Citrine0500_1, Citrine0500_2
Citrine0750_16	Citrine0750_2, Citrine0750_4
Citrine1000_1	Citrine1000_1
Citrine1000_2	Citrine1000_2
Citrine1000_3	Citrine1000_3
Citrine1000_5	Citrine1000_5
Citrine1000_7	Citrine1000_7
Citrine1000_8	Citrine1000_8
Citrine1250_1	Citrine1250_1
Citrine1250_2	Citrine1250_2
Citrine1250_3	Citrine1250_3
Citrine1250_4	Citrine1250_4
Citrine1500_1	Citrine1500_1
Citrine1500_2	Citrine1500_2
Citrine1920_2	CitrineOPF2
Citrine1920_3	CitrineOPF3
Citrine1960_2	Citrine1960_2
Citrine2000_2	Citrine2000_2
Citrine2000_3	Citrine2000_3
Citrine2500_1	Citrine2500_1
Citrine4000_1	Citrine4000_1
Citrine4000_2	Citrine4000_2
Citrine4000_3	Citrine4000_3
Citrine5000_3	Citrine5000_3

4.8.1 Refinement Procedure 1

The first crystallographic refinement process (*refinement procedure 1*) was developed by performing many refinement trials in which the refinement steps and parameters were varied, in order to find a globally good procedure that minimized the R_{Free} factor for the maximum number of structures. This large search was conducted using the Feynman parallel computer cluster at CHESS.

Table 4.7: Unit cell axes for Citrine models. All models were indexed in space group $P2_12_12_1$ with unit cell angles $\alpha = \beta = \gamma = 90^\circ$.

Model	*a* (Å)	*b* (Å)	*c* (Å)
citrine0001_2	51.385	62.624	69.994
citrine0001_18	51.376	62.375	70.639
citrine0500_9	51.383	63.056	70.804
citrine0750_16	51.336	62.886	71.1
citrine1000_1	51.408	62.99	71.315
citrine1000_2	51.362	62.928	71.297
citrine1000_3	51.413	63.122	71.225
citrine1000_5	51.475	62.935	71.664
citrine1000_7	51.401	61.976	70.561
citrine1000_8	51.45	62.272	70.366
citrine1250_1	51.36	62.8	71
citrine1250_2	51.193	62.977	71.128
citrine1250_3	51.44	63.356	71.575
citrine1250_4	51.462	63.34	71.672
citrine1500_1	51.459	62.784	71.163
citrine1500_2	51.343	62.982	71.084
citrine1920_2	51.447	63.223	66.653
citrine1920_3	51.271	62.77	66.11
citrine1960_2	51.549	63.127	71.464
citrine2000_2	51.428	62.944	71.205
citrine2000_3	51.418	63.207	70.904
citrine2500_1	51.363	62.74	70.185
citrine4000_1	51.366	69.993	62.499
citrine4000_2	51.126	60.56	67.209
citrine4000_3	51.531	62.653	71.471
citrine5000_3	51.621	61.06	68.132

The procedure that was found to be most satisfactory was similar to that suggested by Kleywegt and Jones [182]. Each dataset was truncated at a resolution where the average signal-to-noise calculated by SCALA ($\langle I \rangle / \langle \sigma \rangle$) was 3.0, minimizing the R_{Free} factor and minimizing the Log Likelihood Gain (LLG) [107] of the resulting model. For each structure, molecular replacement was performed by MOLREP [85] in the space group $P2_12_12_1$, using the 1HUY structure of Griesbeck *et al*. [1], stripped of solvent, as a search model.

Table 4.8: Protein Data Bank accession codes of high-pressure Citrine atomic structures. The second refinement procedure was used for the deposited structures.

PDB Accession Code	Label	Pressure (MPa)
3DPW	Citrine0001_2	0.1
3DQO	Citrine0001_18	0.1
3DQN	Citrine0500_9	50
3DQM	Citrine0750_16	750
3DQL	Citrine1000_1	100
3DQK	Citrine1000_2	100
3DQJ	Citrine1000_3	100
3DQI	Citrine1000_5	100
3DQH	Citrine1000_7	100
3DQF	Citrine1000_8	100
3DQE	Citrine1250_1	125
3DQD	Citrine1250_2	125
3DQC	Citrine1250_3	125
3DQA	Citrine1250_4	125
3DQ9	Citrine1500_1	150
3DQ8	Citrine1500_2	150
3DQ7	Citrine1920_2	192
3DQ6	Citrine1920_3	192
3DQ5	Citrine1960_2	196
3DQU	Citrine2000_2	200
3DQ4	Citrine2000_3	200
3DQ3	Citrine2500_1	250
3DQ2	Citrine4000_1	400
3DQ1	Citrine4000_2	400
3DPZ	Citrine4000_3	400
3DPX	Citrine5000_3	500

Molecular replacement was followed by 10 cycles of rigid body refinement using REFMAC5 [107]. Next, 10 cycles of restrained refinement with overall temperature factor refinement were performed by REFMAC5 [107], followed by 10 cycles of restrained refinement with isotropic B-factor refinement for each atom. For all steps using REFMAC5, the maximum likelihood residual was used [107].

Table 4.9: Quality indicators and estimated standard uncertainties for Citrine atomic models derived by refinement procedure 1. ESU R_{Free} is the estimated standard uncertainty derived from Cruickshank's R_{Free} positional uncertainty formula (equation 2.31). ESU R is the estimated standard uncertainty derived from Cruickshank's R positional uncertainty formula (equation 2.30). C is the completeness of the dataset used to derive the model out the maximum resolution indicated in the table. n is the number of observations used to derive the model. Cro Mean B is the mean B-factor of all atoms in the main chromophore. Y203 Mean B is the mean B-factor of all atoms in Tyr203. ESU B is the estimated standard uncertainty on the average B factor for the whole molecule.

Label	*P* (MPa)	*R*	R_{Free}	*n*	Atoms	*C*	Hi Res (Å)	Low Res (Å)	Mean B (Å^2)	Cro Mean B (Å^2)	Y203 Mean B (Å^2)	ESU B (Å^2)	Bond Lengths RMSD (Å)	Angles RMSD (°)	ESU R_{Free} (Å)	ESU *R* (Å)
Citrine0001_2	0.1	0.21	0.31	7872	1860	1	2.46	46.68	17.8	16.3	14.1	12.2	0.047	3.43	0.38	1.06
Citrine0001_18	0.1	0.195	0.248	18968	2138	1	1.81	16.14	11.1	11.9	15.4	2.9	0.017	1.805	0.15	0.16
Citrine0500_9	50	0.183	0.235	17735	2080	1	1.9	41.59	15.3	15.1	19.6	3.4	0.021	1.951	0.16	0.16
Citrine0750_16	75	0.185	0.26	10425	1944	1	2.28	41.63	15.2	12.2	18.3	6.7	0.034	2.591	0.26	0.36
Citrine1000_1	100	0.191	0.251	15895	2012	1	1.97	39.84	16.9	14.6	19.3	4.4	0.025	2.112	0.18	0.19
Citrine1000_2	100	0.192	0.243	17189	2045	1	1.92	34.75	16.9	14.7	19.6	3.8	0.023	2.039	0.16	0.18
Citrine1000_3	100	0.194	0.261	14712	1984	1	2.03	39.87	17.8	15.1	22.3	5	0.029	2.224	0.2	0.21
Citrine1000_5	100	0.185	0.237	15242	2021	1	2.01	41.81	19.3	19.4	23.2	4.4	0.026	2.119	0.18	0.2
Citrine1000_7	100	0.184	0.232	24160	2181	1	1.67	41.56	12.7	12.8	15.6	2	0.013	1.738	0.12	0.12
Citrine1000_8	100	0.179	0.227	18573	2146	1	1.86	28.48	14.5	13.8	17.4	3	0.019	1.919	0.15	0.15
Citrine1250_1	125	0.192	0.262	19179	2055	1	1.84	35.51	16	15.2	20	3.1	0.019	1.921	0.16	0.15
Citrine1250_2	125	0.19	0.246	14358	1991	1	2.04	34.69	19.5	15.9	22.8	4.9	0.027	2.28	0.19	0.22
Citrine1250_3	125	0.193	0.265	14306	1967	1	2.04	39.94	20.4	16.8	22.9	5.4	0.029	2.334	0.2	0.22
Citrine1250_4	125	0.194	0.237	19260	2058	1	1.86	31.67	18.3	19	24.9	3.1	0.02	1.952	0.15	0.16
Citrine1500_1	150	0.198	0.23	20926	2058	1	1.78	39.81	17.6	16	19	2.7	0.02	1.92	0.13	0.14
Citrine1500_2	150	0.197	0.262	13028	1945	1	2.04	34.73	19.4	14.7	19.9	5	0.028	2.333	0.21	0.24
Citrine1920_2	192	0.2	0.236	28938	2105	1	1.58	34.24	14.2	11.3	14.3	1.7	0.014	1.673	0.1	0.1
Citrine1920_3	192	0.194	0.253	19419	2072	1	1.75	34.04	15.4	13.2	15.8	2.7	0.017	1.822	0.15	0.15
Citrine1960_2	196	0.2	0.249	19156	2004	1	1.86	41.81	19.9	15.6	18.4	3.9	0.023	2.119	0.15	0.16
Citrine2000_2	200	0.196	0.239	19916	2051	1	1.83	41.7	17.3	15.3	17.6	3.1	0.021	1.99	0.14	0.15
Citrine2000_3	200	0.185	0.239	16553	2035	1	1.95	41.63	16.5	15.8	17.9	3.6	0.021	1.945	0.17	0.18
Citrine2500_1	250	0.193	0.232	19412	2055	1	1.78	24.12	15.9	14.5	16.8	2.8	0.018	1.855	0.14	0.15
Citrine4000_1	400	0.217	0.266	19704	2054	1	1.82	41.42	13.7	11.1	13.3	2.9	0.018	1.862	0.16	0.17
Citrine4000_2	400	0.201	0.302	9832	1916	1	2.25	22.5	13.8	10	12.2	8.5	0.041	2.95	0.31	0.42
Citrine4000_3	400	0.2	0.264	14911	2012	1	2.02	20.44	16.7	12.6	16.1	4.3	0.026	2.13	0.2	0.22
Citrine5000_3	500	0.18	0.242	19030	2157	1	1.8	16.42	10.5	8.3	9.5	2.7	0.018	1.835	0.15	0.15

Table 4.10: Average positional estimated standard uncertainties (ESU) at each pressure level compared with standard deviations of main chromophore phenol ring and imidazolinone ring center of mass positions for Citrine atomic models derived by refinement procedure 1.

Pressure (MPa)	Average ESU (Å)	Std Dev Phenol CoM			Std Dev Imidazolinone CoM		
		X (10^{-2} Å)	*Y* (10^{-2} Å)	*Z* (10^{-2} Å)	*X* (10^{-2} Å)	*Y* (10^{-2} Å)	*Z* (10^{-2} Å)
0.1	0.265	1.23	0.03	0.392	6.31E-01	0.02	0.001
50	0.156	-	-	-	-	-	-
75	0.260	-	-	-	-	-	-
100	0.164	6.49	27.14	0.190	3.35E+00	21.36	19.390
125	0.176	8.85	7.29	1.164	8.34E-01	3.08	1.273
150	0.171	0.12	0.08	0.001	3.05E-02	1.91	0.085
196	0.142	1.96	2.64	0.616	2.06E+00	2.02	1.108
250	0.137	-	-	-	-	-	-
400	0.221	4.02	2.66	0.595	1.12E+01	2.82	3.461
500	0.149	-	-	-	-	-	-

Finally, 5 cycles of solvent addition followed by 5 cycles of restrained refinement with isotropic B-factor refinement were performed using the ARP/wARP program by Perrakis *et al*. [111, 183] and REFMAC5 [107], until the R_{Free} [138] factor and log-likelihood gain (LLG) function [107] were simultaneously minimized. The number of restrained refinement cycles in the final solvent addition-restrained refinement cycle was chosen to minimize R_{Free} and LLG. Finally, the structures were validated with COOT [184] and PROCHECK [114]. The maximum resolutions of the data sets used in the first refinement procedure ranged from 2.46 Å to 1.5 Å. Refinement quality indicators for each structure derived by refinement procedure 1 are shown in table 4.9. In addition to the model quality indicators, R and R_{Free}, table 4.9 shows the quantities necessary to compute the positional estimated standard uncertainty for each model (equations 2.30 and 2.31). Table 4.9 also shows overall B-factor for each model, and the mean B-factor for the main chromophore and tyrosine 203.

It should be noted that the overall B-factor for the structure is similar to, often within error of, the mean B-factors of the main chromophore and tyrosine 203, suggesting that the average uncertainty formulas (equations 2.30 and 2.31) are appropriate for the main chromophore and tyrosine 203.

Table 4.10 shows the average positional estimated standard uncertainty at each pressure level, and the standard deviations of the main chromophore and tyrosine 203 phenol ring center of mass positions for the atomic models derived by refinement procedure 1.

4.8.2 Refinement Procedure 2

The second refinement procedure was similar to the first, except that the lengths of the peptide bonds from residues 64 and 68 to the main chromophore were tightly constrained to satisfy the stereochemical constraints required for accession into the Protein Data Bank. Additionally, the maximum resolution used in refinement was raised to the maximum found in the X-ray diffraction dataset. Although raising the maximum resolution used in the refinement reduces the *R*-factor of the resulting model as the signal to noise ratio of higher-resolution spots is lower, the precision of the model should increase. Additionally, the completeness of the diffraction dataset out to the maximum resolution used in the refinement, *C*, will be reduced due to an inability to collect all reflections in the maximum resolution ring as the detector has a square face. Cruickshank's estimates for the coordinate uncertainty in a protein atomic model (equations 2.30 and 2.31) indicate that while the precision of the atomic model may suffer from a reduced *R*-factor and reduced completeness, *C*, this will be more than compensated for by an increase in the number of observations, *n*, relative to model parameters, *p*.

Table 4.11: Quality indicators and estimated standard uncertainties for Citrine atomic models derived by refinement procedure 2. ESU R_{Free} is the estimated standard uncertainty derived from Cruickshank's R_{Free} positional uncertainty formula (equation 2.31). ESU R is the estimated standard uncertainty derived from Cruickshank's R positional uncertainty formula (equation 2.30). C is the completeness of the dataset used to derive the model out the maximum resolution indicated in the table. n is the number of observations used to derive the model. Cro Mean B is the mean B-factor of all atoms in the main chromophore. Y203 Mean B is the mean B-factor of all atoms in Tyr203. ESU B is the estimated standard uncertainty on the average B factor for the whole molecule.

Label	*P* (MPa)	*R*	R_{Free}	*n*	Atoms	*C*	Hi Res (Å)	Low Res (Å)	Mean B (Å^2)	Cro Mean B (Å^2)	Y203 Mean B (Å^2)	ESU B (Å^2)	Bond Lengths RMSD (Å)	Angles RMSD (°)	ESU R_{Free} (Å)	ESU *R* (Å)
Citrine0001_2	0.1	0.221	0.284	19819	2024	0.675	1.59	20	27.4	27.2	23.7	2.9	0.025	2.168	0.17	0.17
Citrine0001_18	0.1	0.218	0.259	27148	2094	0.772	1.5	16.14	16.3	13	16.2	1.7	0.028	2.388	0.12	0.12
Citrine0500_9	50	0.201	0.257	29494	2164	0.737	1.44	20	22.4	17.9	21.2	1.8	0.028	2.41	0.11	0.1
Citrine0750_16	75	0.206	0.26	30416	2149	0.751	1.44	20	23.7	19.7	21.7	2.2	0.027	2.295	0.11	0.1
Citrine1000_1	100	0.216	0.271	29074	2100	0.769	1.47	20	24.2	18.3	23	2.5	0.029	2.331	0.12	0.11
Citrine1000_2	100	0.216	0.265	28699	2120	0.656	1.4	20	23.8	17.8	21.4	2	0.027	2.263	0.12	0.11
Citrine1000_3	100	0.225	0.298	27928	2092	0.79	1.51	20	25	19.8	24	2.6	0.027	2.24	0.14	0.12
Citrine1000_5	100	0.212	0.266	27526	2120	0.653	1.42	20	27.3	22	26.5	2.7	0.027	2.248	0.12	0.12
Citrine1000_7	100	0.185	0.237	29566	2243	0.766	1.45	20	19.6	13.6	15.8	1.5	0.03	2.552	0.11	0.1
Citrine1000_8	100	0.198	0.244	27490	2165	0.721	1.46	20	21.3	14.8	18.8	1.8	0.029	2.388	0.11	0.11
Citrine1250_1	125	0.207	0.263	32312	2162	0.783	1.43	20	23.2	16.6	22	2.1	0.028	2.4	0.11	0.1
Citrine1250_2	125	0.216	0.28	25884	2110	0.593	1.4	20	27.5	20.9	24.4	2.7	0.025	2.251	0.14	0.13
Citrine1250_3	125	0.222	0.279	26526	2083	0.715	1.49	20	27.4	22.5	25.6	3	0.026	2.238	0.13	0.13
Citrine1250_4	125	0.205	0.245	32126	2153	0.777	1.44	20	26.2	20.4	23.9	2.3	0.031	2.459	0.1	0.1
Citrine1500_1	150	0.214	0.271	32507	2117	0.743	1.4	20	24.8	17	19.9	2.2	0.028	2.411	0.11	0.1
Citrine1500_2	150	0.225	0.283	25014	2071	0.7111	1.51	20	26.5	20.6	23.8	2.7	0.028	2.359	0.14	0.13
Citrine1920_2	192	0.212	0.25	45708	2123	0.761	1.23	20	20.5	11.7	14.7	1	0.033	2.722	0.07	0.07
Citrine1920_3	192	0.199	0.255	21658	2062	0.791	1.6	20	20.8	13.5	15.7	2.1	0.026	2.346	0.14	0.14
Citrine1960_2	196	0.213	0.255	28145	2079	0.778	1.5	20	26.2	17.3	20.4	2.6	0.028	2.37	0.12	0.11
Citrine2000_2	200	0.207	0.257	28898	2130	0.684	1.42	20	24.2	16.3	18.6	2.1	0.029	2.393	0.11	0.11
Citrine2000_3	200	0.198	0.24	28543	2148	0.754	1.47	20	23.7	17.3	19.6	2	0.03	2.424	0.11	0.11
Citrine2500_1	250	0.189	0.253	20949	2136	0.864	1.7	20	21.7	14.5	16.7	2.4	0.028	2.322	0.15	0.14
Citrine4000_1	400	0.224	0.282	26250	2063	0.91	1.6	20	19.4	11.9	13.9	2.1	0.026	2.321	0.13	0.13
Citrine4000_2	400	0.227	0.296	21287	2040	0.952	1.7	20	21.8	15.9	19.8	3.6	0.024	2.181	0.16	0.16
Citrine4000_3	400	0.224	0.294	24113	2055	0.974	1.7	20	23.5	16.2	18.8	3.1	0.028	2.273	0.15	0.14
Citrine5000_3	500	0.199	0.24	27118	2176	0.813	1.5	16.42	16.3	8.7	10.1	1.6	0.028	2.474	0.11	0.11

Table 4.12: Average positional estimated standard uncertainties (ESU) at each pressure level compared with standard deviations of main chromophore phenol ring and imidazolinone ring center of mass positions for Citrine atomic models derived by refinement procedure 2.

Pressure (MPa)	Average ESU (Å)	Std Dev Phenol CoM			Std Dev Imidazolinone CoM		
		X (10^{-2} Å)	Y (10^{-2} Å)	Z (10^{-2} Å)	X (10^{-2} Å)	Y (10^{-2} Å)	Z (10^{-2} Å)
0.1	0.144	0.211	1.159	0.047	0.350	10.160	1.788
50	0.113	-	-	-	-	-	-
75	0.111	-	-	-	-	-	-
100	0.119	1.580	1.958	0.710	0.610	24.910	7.932
125	0.119	0.296	1.137	0.980	0.996	12.630	2.713
150	0.125	0.089	0.001	0.003	0.259	0.004	0.004
196	0.110	1.304	1.452	0.582	6.783	6.869	1.420
250	0.147	-	-	-	-	-	-
400	0.148	2.277	4.225	0.369	7.208	8.726	3.874
500	0.111	-	-	-	-	-	-

Refinement quality indicators for set of atomic models produced by refinement procedure 2 are shown in table 4.11. The average positional estimated standard uncertainty at each pressure level, and the standard deviations of the main chromophore and tyrosine 203 phenol ring center of mass positions for the atomic models derived by refinement procedure 2 are shown in table 4.12.

The model-set derived from refinement procedure 2 was analyzed to find the effects of high pressure on the Citrine scaffold. This analysis is described in a forthcoming article by us [6] and in chapter 5.

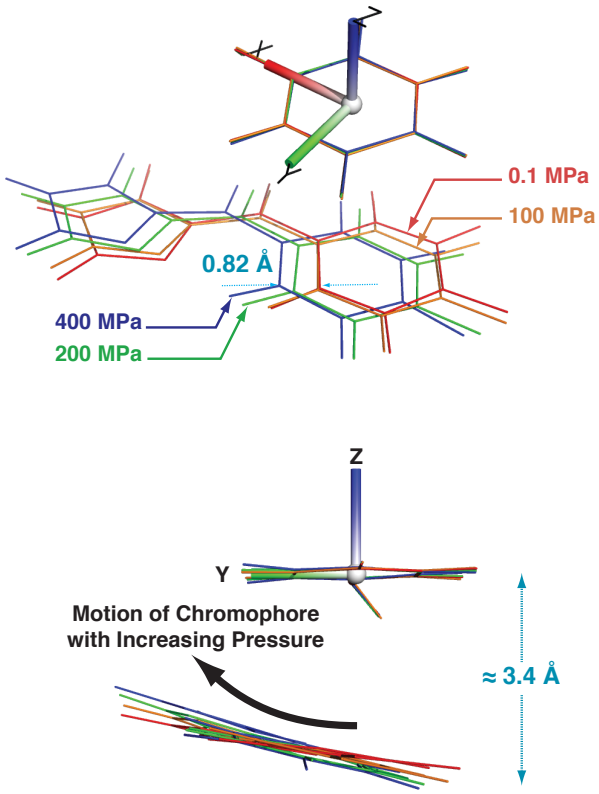


Figure 4.43: X-ray crystal structures of Citrine's chromophore at selected pressures with the lowest R_{Free} factors. Note the *x*-*y*-*z* coordinate axes defined in this figure. The origin of the coordinate system is at the center of the tyrosine 203 phenol ring. Note that the centers of mass of the main chromophore phenol and imidazolinone rings have negative *z* coordinates.

4.8.3 Analysis of High-pressure Atomic Models

Refinement procedures 1 and 2 both yielded a set of atomic models that reveal a small, but progressive reorientation of the two stacked aromatic rings that compose Citrine's chromophore. This deformation of the chromophore smoothly increases with pressure and stands out from atomic coordinate error. Structures of the main chromophore and tyrosine 203 of the highest quality structures derived by refinement procedure 1 at

representative pressures are shown in figure 4.43. The tyrosine 203 phenol rings of the high pressure cryo-cooled structures were aligned using the LSQKAB program [185] to allow inspection of the main chromophore position relative to tyrosine 203. This *aligned tyrosine 203 coordinate system* used in this discussion is defined in figure 4.43. The origin of the aligned tyrosine 203 coordinate system is at the center of the tyrosine 203 phenol ring. The *x* and *y* axes of this system are embedded in the plane of the phenol ring while the *z* axis is normal to this plane. The *x*, *y* and *z* axes are oriented so that all motions with increasing pressure have a positive sign.

Plots of the motion of the main chromophore's two rings in the aligned tyrosine 203 coordinate system are shown in figures 4.44 and 4.45. Figure 4.44 shows the motion of the main chromophore in the tyrosine 203 coordinate system for the model-set derived by refinement procedure 1. Figure 4.45 shows the same motions for the model-set derived by refinement procedure 2.

Both model-sets show highly similar chromophore deformation motions. In the model-sets derived by both refinement procedures, the center of the main chromophore's phenol ring slides underneath the stacked tyrosine 203 phenol with increasing pressure. This sliding motion is largely confined to a plane parallel to the *x*-*y* plane, and 3.4 Å below it, at $z \approx -3.4$ Å. In the model-set derived from refinement procedure 2 there is a slight downward motion of the main chromophore relative to tyrosine 203 of ≈ 0.05 Å over 500 MPa.

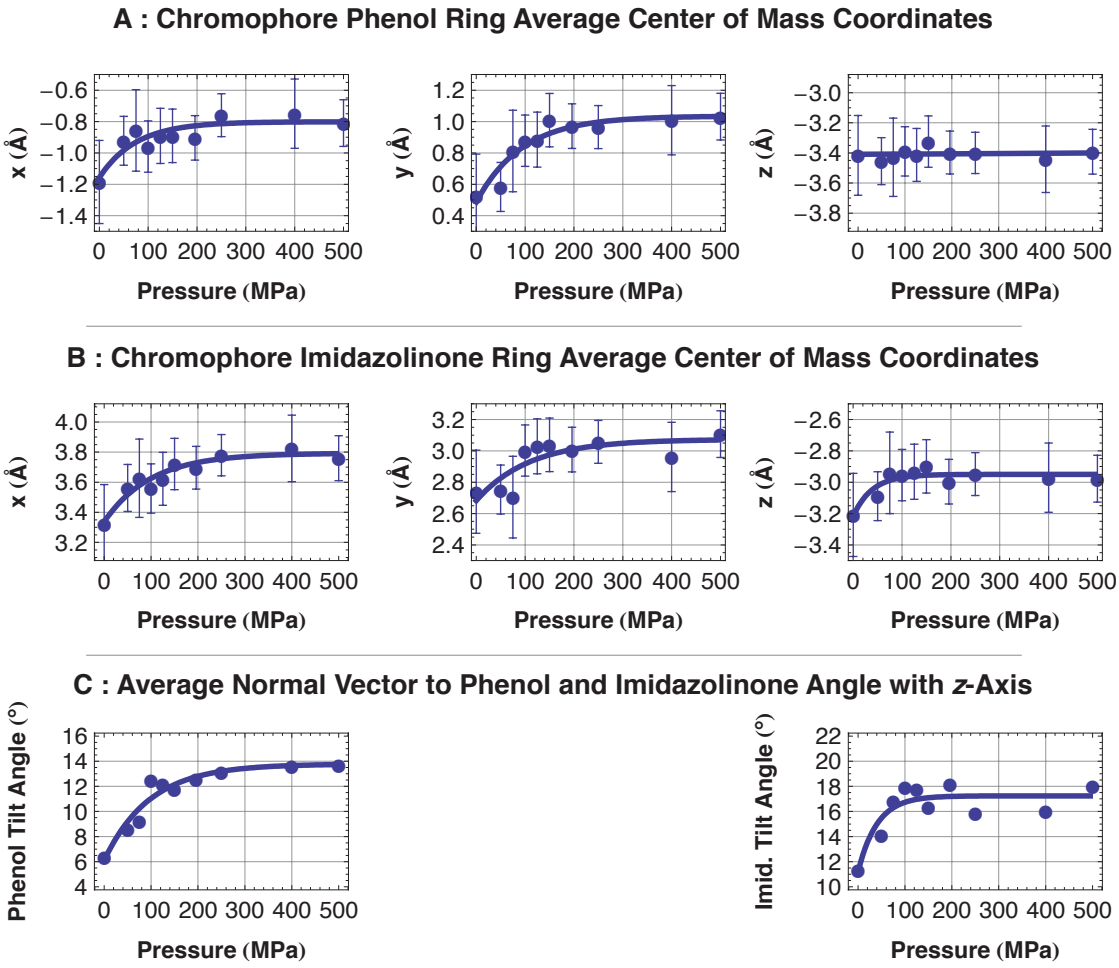


Figure 4.44: Average center of mass positions of the main chromophore phenol and imidazolinone rings in the coordinate system defined in figure 4.43. Error bars estimated by Cruickshank's formula (equation 2.31) using the data in table 4.9. Fits are shown as solid, blue lines. Models derived from refinement procedure 1.

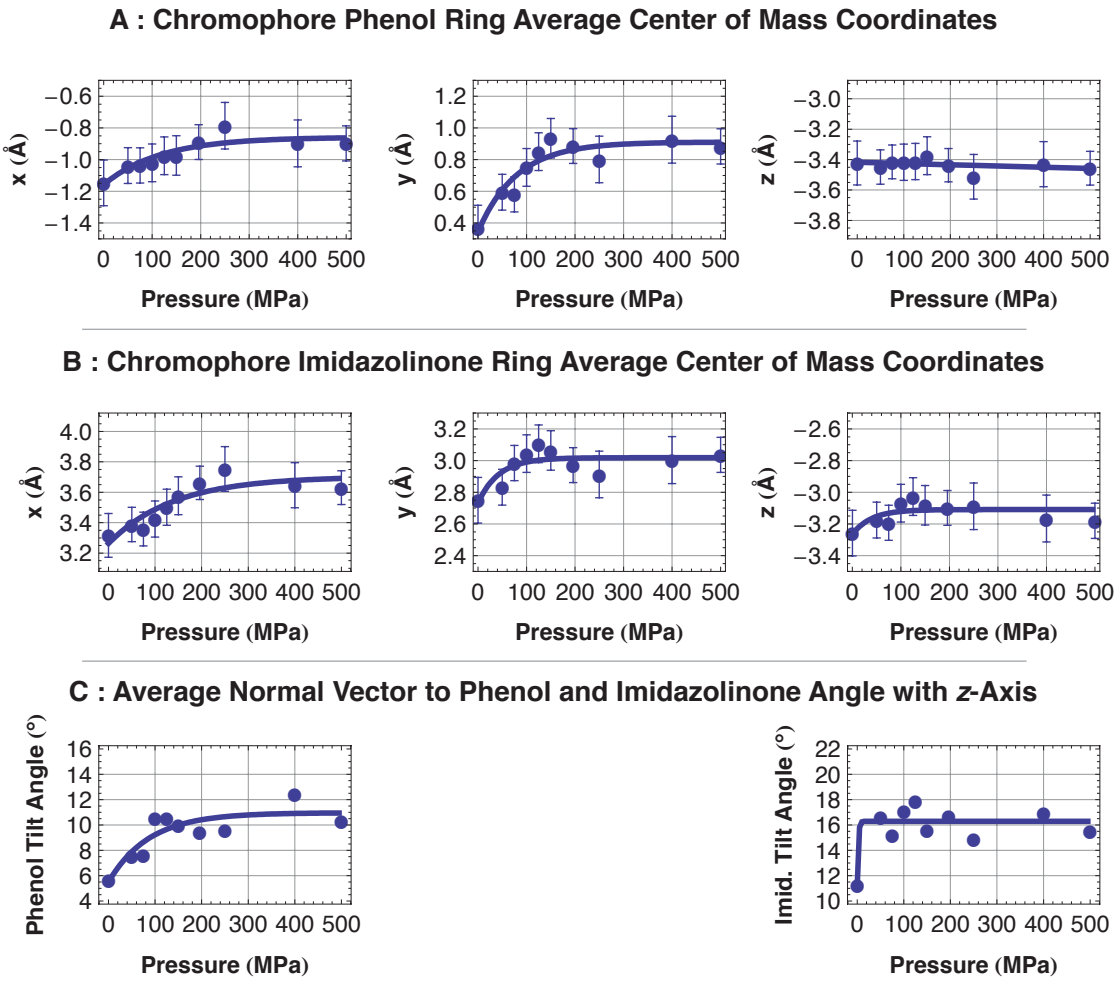


Figure 4.45: Average center of mass positions of the main chromophore phenol and imidazolinone rings in the coordinate system defined in figure 4.43. Error bars estimated by Cruickshank's formula (equation 2.31) using the data in table 4.11. Fits are shown as solid, blue lines. Models derived using refinement procedure 2.

In both model-sets the maximum extent of the swing is approximately 0.8 Å over 500 MPa. In both model-sets, the largest component of motion is approximately +0.5 Å in the *y*-direction, moving the main chromophore away from the origin. It is interesting to note that the starting position of the motion is slightly different for both model-sets, 0.5 Å for model-set 1 and 0.4 Å for model-set 2, but the extent of the motion is the same for both model-sets.

The motion in the *x*-direction brings the center of the main chromophore's phenol closer to the origin of the coordinate system. The motion is approximately 0.4 Å for model-set 1 and 0.35 Å for model-set 2.

The motion of the main chromophore imidazolinone ring shows greater differences between the two model-sets. This is not surprising, as the main chromophore is attached to the rest of the Citrine polypeptide through the imidazolinone ring, and these bonds were tightly constrained in refinement procedure 2. For refinement procedure 1, the center of mass of the main chromophore's imidazolinone ring moves by 0.4 Å in *y*, 0.4 Å in *x* and 0.2 Å in *z*. For refinement procedure 2, the center of mass of the main chromophore's imidazolinone ring moves by ≈ 0.3 Å in *y*, 0.4 Å in *x* and 0.1 Å in *z*

In addition to sliding, the main chromophore phenol ring re-orients with respect to the tyrosine 203 phenol ring, and to the main chromophore's imidazolinone ring. In refinement procedure 1, the normal vector to the main chromophore's phenol ring rotates from 6˚ with respect to the *z*-axis to 14˚ by 500 MPa, while in refinement procedure 2, this rotation is from 6˚ to 11˚. In refinement procedure 1, the normal to the main chromophore imidazolinone ring rotates by less than 5˚, from ≈ 12˚ to 17˚,

changing the orientation of the two rings composing the main chromophore. For refinement procedure 2, a rotation of ≈ 5˚, from 11˚ to 16˚ occurs from 0.1 to 50 MPa, and then stops. The re-orientation as a function of pressure is shown in figures 4.44 and 4.45.

The progressive deformation of Citrine's chromophore under high-pressure cryocooling conditions appears to be reproduced under differing refinement conditions. It is noteworthy, that just prior to publication of our article on the effects of pressure on the structure and fluorescence spectrum of the Citrine chromophore, we detected an error in the model of the Citrine structure used in refinement. Residue 80 of the Citrine molecule, a surface residue, is listed as an arginine in the original Citrine structure (PDB accession code 1HUY) by Griesbeck *et al*. [1]. However, genetic sequencing indicates that residue 80 is a glutamine. Conversion of residue 80 to a glutamine produces only a minor variation in the deformation motion seen at the chromophore, further suggesting that the motion seen is robust to refinement error.

The progressive deformation of Citrine's structure under high-pressure cryo-cooling is consistent with the mechanism suggested for the shift towards the green of Citrine's fluorescence peak: a horizontal sliding of the main chromophore and tyrosine 203, that was suggested by quantum mechanical modeling in chapter 3. The relative positions of the main chromophore and the perturbing tyrosine 203 phenol ring separate with increasing pressure, removing the perturbing influence of the tyrosine 203 phenol allowing the main chromophore to return to its unperturbed, green fluorescent state. The removal of the perturbing interaction requires a structural shift of only ≈ 0.8 Å.

The fluorescence peak shift of Citrine from yellow to green, discussed in this chapter, further suggests that the deformation motion seen here is real, not an artifact of refinement.

This chapter has reported on the deformation at the chromophore of the Citrine molecule, and implicated this structural deformation in a shift of the fluorescence peak of the molecule. However, this chapter does reveal how pressure actuates the deformation seen at the chromophore. Chapter 5 discusses an analysis of the entire structure of the Citrine molecule under high pressure and proposes a mechanism of actuation of this deformation at the chromophore.

CHAPTER 5

THE DEFORMATION OF THE CITRINE SCAFFOLD UNDER HIGH PRESSURE

5.1 Introduction

Chapter 4 discussed the deformation motion at the chromophore of the Citrine molecule that is correlated with a shift in the fluorescence peak of the molecule from yellow to green. This chapter focuses on structural changes in the scaffold of the Citrine molecule that actuate the deformation motion at the chromophore; that is, how pressure is mechanically transmitted from the surface of the Citrine molecule to the chromophore. The results of this analysis show that the actuation of the chromophore motion is due to the concerted, differential motion of two clusters of residues that compose the β-barrel and central 3-10 helix of the Citrine molecule.

It has long been appreciated that the three-dimensional structure of a protein molecule can exert a strong influence over the active site of the molecule [8]. This linkage permits allosteric binding of ligands and processing of substrates. Examples of this include the relaxed (oxygenated) to tense (de-oxygenated) transition in hemoglobin [8], the allosteric inhibition of phospho-fructo-kinase-1 (PFK-1) by phospho-enolpyruvate [186], and the gating of ion channels [187]. These, and other well-known examples of allosteric behavior occur in protein molecules that possess more than one subunit.

Small structural changes, sometimes less than 1 angstrom in magnitude, in a distant part of the molecule may be communicated through the protein matrix to an active and/or binding site, promoting a change in binding constant, catalytic rate or other functional change.

The structural flexibility of multi-subunit proteins seems to permit the ready transmission of mechanical signals between distant parts of molecule, and facilitates allosteric behavior.

However, although multiple subunit proteins most obviously display allosteric behavior, recent results indicate that the possession of multiple subunits is not a prerequisite for the manifestation of allosteric behavior. As discussed in chapter 1, Ansari *et al*. [50] demonstrated that carbon-monoxy sperm whale myoglobin displays three conformational substates associated with the infrared stretch bands of the heme-bound carbon monoxide. Frauenfelder *et al*. [51] speculated that these conformational substates may possess different reactive properties, giving myoglobin ligand binding and reactive properties not traditionally associated with single subunit, non-allosteric enzymes, suggesting that even monomeric proteins display behavior that might be regarded as allosteric.

Single molecule experiments also indicate that many catalytic, monomeric proteins display a variety of catalytic rates or functional states. Blum *et al*. [7] observed that isolated molecules of Yellow Fluorescent Protein, a close relative of Citrine, transiently convert to a blue-shifted form, with a fluorescence peak at ≈ 514 nm. This blue-shifted fluorescence peak is close the fluorescence peak value that is asymptotically approached by Citrine under increasing freezing pressure.

Coureux *et al*. [61] suggested that this functional heterogeneity seen in single molecule experiments is the result of these molecules assuming a variety of conformations that differ only on the sub-angstrom level. As the active site of the

enzyme is intricately interwoven into the structure of the scaffold, these different structural configurations may result in differing active site behavior.

This interplay of the scaffold of the protein and its function suggests that deliberate perturbation of the scaffold on the 0.1 Å level could greatly affect the function of the molecule. Chapter 1 discussed the effects of high pressures on the structure and function of protein molecules. Protein atomic structures solved at pressures up to a few hundred MPa [3-5, 23, 24, 26-29, 74] indicate that atoms in protein molecules are typically displaced by approximately 0.1 to 1 Å from their ambient pressure positions. Pressures in the same range also significantly modify protein function . For example, the flash decay rate of firefly luciferase is reduced [13], the oxygen binding affinity of human hemoglobin is doubled [188], and oxidation rates by morphinone reductase are substantially increased [18]. Urayama *et al*. [4] demonstrated that sub-angstrom structural changes in the scaffold of the sperm-whale myoglobin may correspond to the A_0 sub-state of myoglobin.

While it is reasonable to assume that the functional changes follow from changes in protein structure with pressure, there is a lack of specific experimental examples that explicitly explore the mechanism of interaction between the matrix of a protein molecule, the structure of its active site and its function. Chapter 4 discussed a direct correlation between a sub-angstrom progressive deformation of the Citrine chromophore and a large progressive change in the fluorescence spectrum of the molecule [5]. This chapter attempts to understand the interaction of structural changes in the matrix of the Citrine molecule and structural deformations in the fluorescent site of the molecule.

5.2 Analysis Procedures

5.2.1 Clustering Analysis

Clusters of residues that move in concert with increasing pressure in the series of Citrine atomic structures were identified using the heuristic clustering algorithm RIGIMOL (DeLano Scientific LLC, Palo Alto, CA, USA). Input coordinate files to RIGIMOL were prepared with PYMOL (DeLano Scientific LLC, Palo Alto, CA, USA). The results of the clustering analysis and attempts to understand sources of computational artifacts in the procedure are discussed in section 5.3.4.

5.2.2 Computation of Structural Properties

Structural properties of Citrine atomic structures such as inertia tensors, centers of mass, and principal axes were computed using custom software implemented in PYTHON with use of the NUMPY [189] numerical library. Fitting of structural parameters was performed using the IPYTHON interactive interpreter (http://ipython.scipy.org) and the NUMPY AND SCIPY numerical libraries. NUMPY is an open-source numerical library for the PYTHON programming language, available at http://numpy.scipy.org [189]. SCIPY is an open-source scientific library for PYTHON available at http://www.scipy.org.

Cavity detection and volume calculations were performed with the molecular surface computation program MSMS [190] using a 1.2 Å radius probe. Cavity surface information from the MSMS program was visualized in PYMOL (DeLano Scientific LLC, Palo Alto, CA, USA) using a modified version of a custom PYTHON script by Warren DeLano.

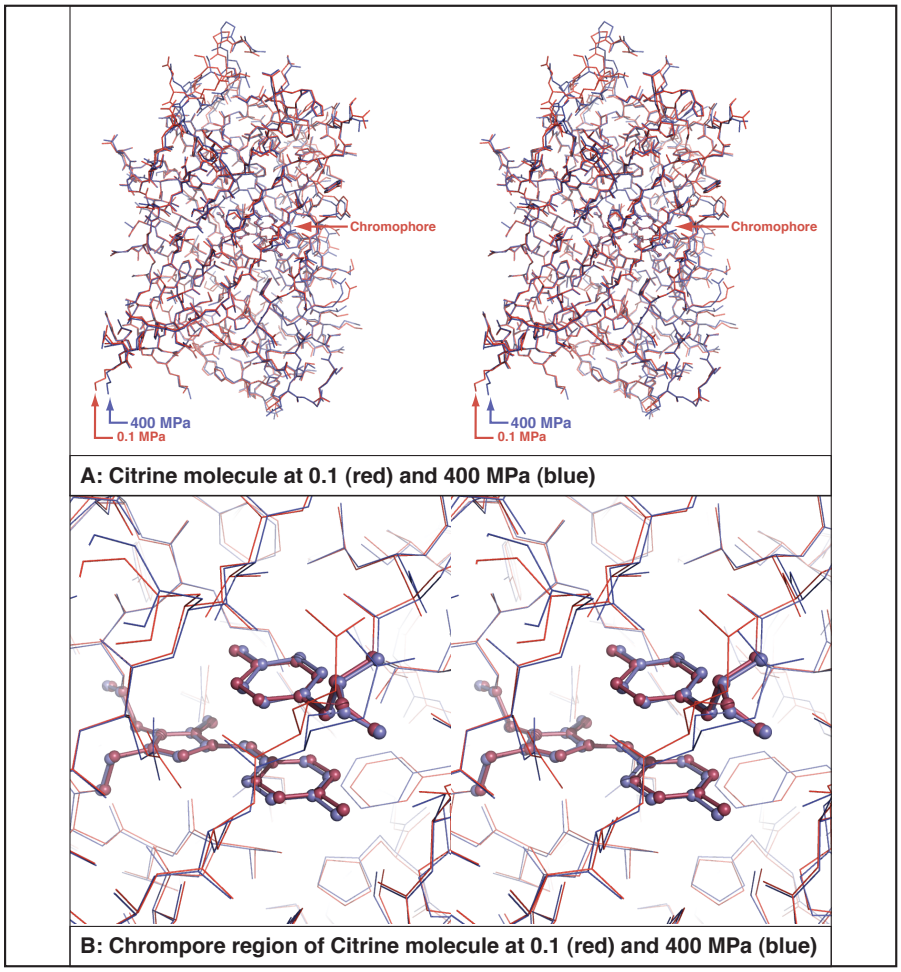


Figure 5.1: A: Stereo image of superimposed atomic structures of Citrine frozen at 0.1 MPa (red) and 400 MPa (blue). The main chromophore and tyrosine 203 are highlighted in the 0.1 MPa structure. B: Stereo image of the chromophore region of the Citrine molecule at 0.1 MPa (red) and 400 MPa (blue).

5.2.3 Assignment of Secondary Structures

Assignment of secondary structural properties to the series of Citrine structures was performed with DSSP [191].

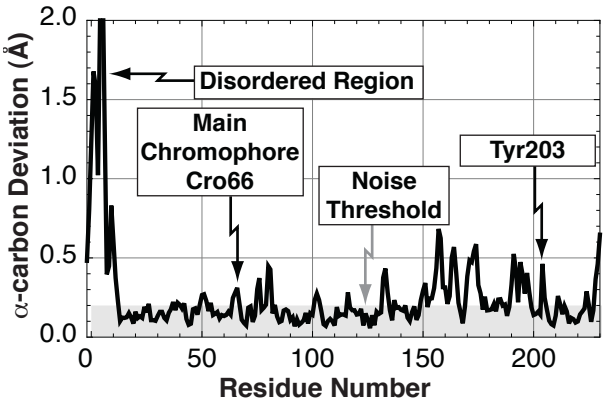


Figure 5.2: Average distance deviation between α-carbons of Citrine at 0.1 and 400 MPa versus residue number. Deviations below the grey band, the estimated noise, at 0.2 Å are at the uncertainty of the measurement.

5.2.4 Choice of Refinement Set

For the analysis presented in this chapter, and in forthcoming article on the effects of pressure on the scaffold of Citrine [6], we chose to use the atomic models of Citrine in refinement set 2 discussed in chapter 4. The reasons for this choice are two-fold. Firstly, these models were deposited into the Protein Data Bank (PDB), and are thus publicly available, so that our analysis may be repeated, critiqued and improved upon easily. Secondly, refinement of these models used the highest resolution diffraction resolution data available, and thus, these models show the lowest estimated standard uncertainties.

5.3 Results

5.3.1 Secondary and Tertiary Structure of Citrine Under High Pressure

The aligned structures of the Citrine molecule at ambient pressure and at 400 MPa are shown in figure 5.1. By visual inspection, the low and high-pressure structures of the Citrine molecule shown in figure 5.1 appear largely identical. A plot of the average displacement of the α-carbon of each residue under a pressure increase from ambient

pressure to 400 MPa is shown in figure 5.2. The displacement of the α-carbons due to pressurization indicates that the effects of pressure up to 400 MPa on the structure of the Citrine molecule are small in magnitude when compared to the overall dimension of the protein. By adding the estimated standard uncertainties on the structures of Citrine shown in table 4.10 in quadrature, we estimate that the uncertainty in the displacements between pressures is approximately 0.2 to 0.3 Å, as indicated by the "noise threshold" in figure 5.2 [4, 29]. Explicitly, we believe that the estimated standard uncertainty on a displacement of an atom, Δx, between two pressure levels, P_0 and P_1, should be estimated by the estimated standard positional uncertainties at the two pressure levels, $\sigma(x,P_0)$ and $\sigma(x,P_1)$;

$$\Sigma(\Delta x,P_0,P_1)=\sqrt{3}\left(\sigma(x,P_0)^2+\sigma(x,P_1)^2\right)^{1/2} \tag{5.1}$$

The $\sqrt{3}$ term is incorporated to account for the possibility that the error the coordinates of the atom at each pressure may not be in the same dimension. The Cruickshank uncertainty formula (equations 2.30 and 2.31) estimates the uncertainty on each component of coordinate of an atom, rather than in the position. To account for the uncertainty in a position, the coordinate uncertainty must be multiplied by $\sqrt{3}$ (equation 2.32). Thus, for figure 5.2

$$\begin{aligned}\Sigma(\Delta x,0.1\text{ MPa},400\text{ MPa}) &= \sqrt{3}\left(0.114^2+0.111^2\right)^{1/2}\\ &= 1.59\sqrt{3}\text{ Å}\\ &= 0.27\text{ Å}\end{aligned} \quad . \tag{5.2}$$

It is worth noting, that the apparent noise threshold in figure 5.2 is ≈ 0.2 Å, suggesting that the error estimate of 0.27 Å may be slightly too large. This is certainly possible as the Cruickshank error estimation formula only accounts for the diffraction data contribution to the error on the structure. Geometric constraints applied in the refinement of these structures may assist in reducing the noise threshold for detection of pressure induced perturbations, at least at the maximum diffraction resolutions, 1.5 to 2.5 Å, observed in these experiments.

It is important to emphasize that although there are likely many residues in the structure of Citrine where the displacement due to pressurization is < 0.2 to 0.3 Å, the structural resolution of the diffraction experiments does not permit the definitive identification of displacements smaller than the noise threshold [4, 29].

There are numerous residues that are displaced by several times the noise threshold. The largest of the deviations in figure 5.2 are at the start (N-terminus) of the primary sequence. This may in part be due to a change in secondary structure of this region with increasing pressure. Calculation of the secondary structure of Citrine with DSSP [191] indicates that the first 10 residues (Proline -1 to Phenylalanine 8) form an α-helix at ambient pressure. However, with increasing pressure, the number of residues in the N-terminal α-helix is reduced. In structures above 125 MPa, Phe8 consistently (14 out of 16 structures) is no longer in the α-helix. Leu7 also is not in the helix in 10 of 16 structures above 125 MPa. The N-terminal residues Pro-1 and Met0 are not part of the helix in 5 of 7 structures above 200 MPa.

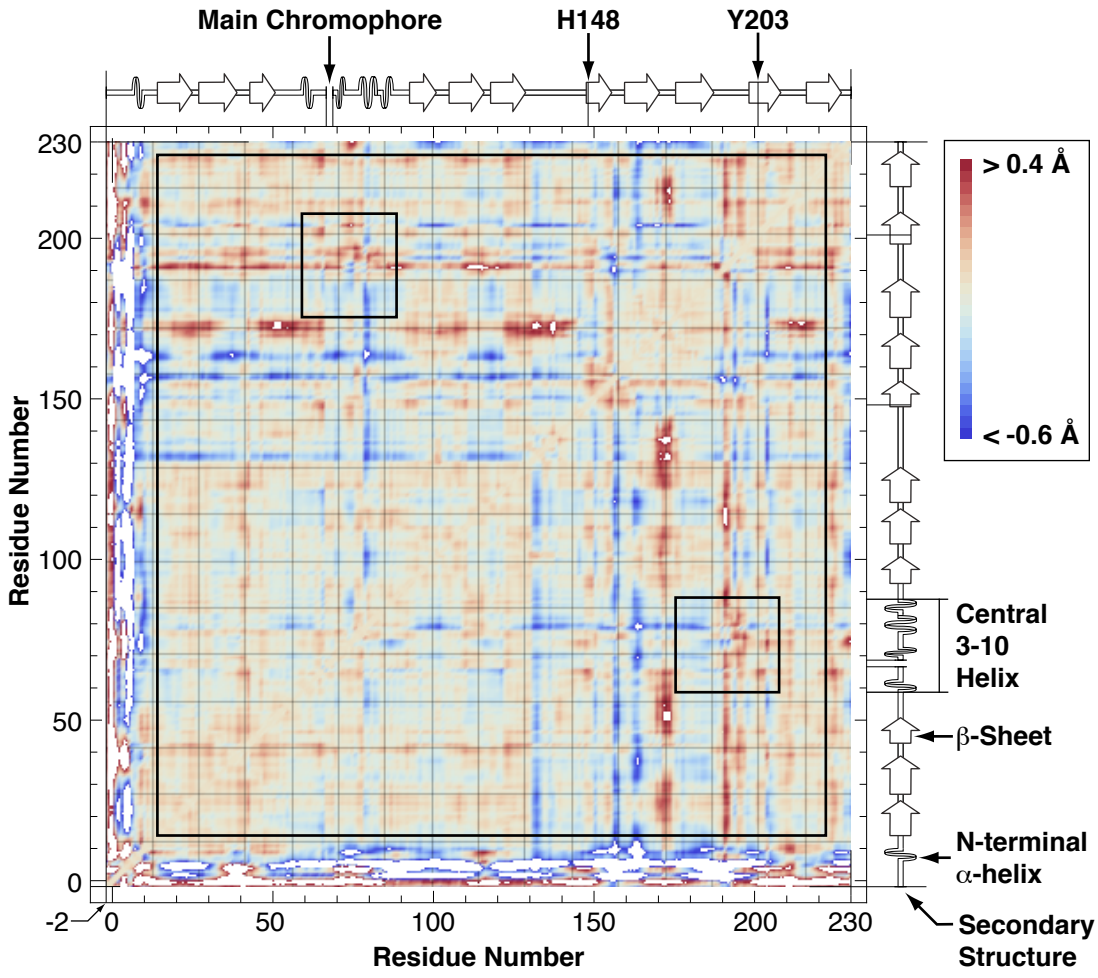
However, the large displacements at the N-terminus seen in figure 5.2 cannot be attributed solely to pressure. The N-terminus α-helix and loops of Citrine have poorly resolved electron density, indicating disorder. At high pressure, these N-terminal residues have B-factors that are considerably higher than the average B-factor for the entire Citrine molecule. Cruickshank's coordinate uncertainty formulas, discussed in section 2.8.2, predict an exponential increase in coordinate uncertainty with increasing B-factor. The variation in α-carbon position between any two structures at the same pressure in the series of high-pressure Citrine structures is of similarly large magnitude to that seen in figure 5.2. It is interesting to note that N-terminal primary sequence deletions up to Phe8, the same residues that showed changes of secondary structure with increasing pressure, have no impact on the folding or maturation of the Green Fluorescent Protein [192]. This suggests that changes of the secondary structure up to Phe8, while notable, do not affect the spectral properties of Citrine. Additionally, the N or C terminus (or both) of YFP and other *Aequorea* fluorescent proteins can be fused to other materials, including proteins, without changing the spectral properties of the molecule [2]. Therefore, we believe that these large spatial deviations seen at the N-terminus are decoupled from the pressure-induced actuation of the chromophore deformation.

The secondary structure of the 3-10 helix that fills the center of the β-barrel and the walls of the β-barrel is retained up to at least 500 MPa. These results are consistent with the spectroscopic observations by Scheyhing *et al*. [148] and Herberhold *et al*. [193] that suggest that the tertiary structures of *Aequorea* fluorescent protein molecules are stable to unfolding until pressures of at least 900 MPa and up to 1300 MPa for many mutants.

In addition to the displacements seen at the disordered N-terminus region of Citrine, there are numerous residues that also displaced by several times the noise threshold. Two notable displacements highlighted in figure 5.2 are the pressure-induced displacements of residues 66 and 203, the main chromophore and tyrosine 203, consistent with our previous report of the deformation of the chromophore [5]. It is noted that the magnitudes of the residue displacements are non-uniform, suggesting the non-isotropic nature of the compression of Citrine.

Although the displacement plot shown in figure 5.2 shows the magnitude of structural changes due to pressurization, it does not provide information on the directions of these motions. To further understand the deformation of the Citrine molecule under pressure, a distance difference matrix [24] was calcuated, showing the expansion and contraction of distances between residue pairs in the Citrine structure under a pressure increase from 0.1 to 400 MPa. The distance difference matrix of Citrine under pressure shown in figure 5.3 indicates that Citrine does not uniformly compress with pressure: some distances compress (blue regions in figure 5.3), while others expand (red regions in figure 5.3) with increasing pressure. The white regions in figure 5.3 indicate distances that either compress by more than 0.6 Å, or expand by more than 0.4 Å. The largest expansion of a distance in the whole molecule is by ≈ 2.7 Å, while the largest compression of a distance is by ≈ 1.7 Å. At least of one of the residues in the atom pairs that show these large distance compressions and expansions are often observed at the N-terminus of the Citrine molecule. The largest distance compression between residues in the β-barrel is 0.92 Å, and the largest distance expansion is 0.67 Å.

Figure 5.3: Averaged distance difference matrix for the Citrine molecule under a pressure increase from 0.1 to 400 MPa. Compressions are shown in blue and expansions are shown in red. The matrix is symmetrical about the diagonal that runs from lower left to upper right of the matrix. Note that all elements on this diagonal are zero. The white regions in the matrix, predominantly at the periphery of the protein structure, indicate distances that either compress by more than 0.6 Å, or expand by more than 0.4 Å. The largest expansion of a distance is by ≈ 2.7 Å, while the largest compression of a distance is by ≈ 1.7 Å. The largest expansion of a distance in the β-barrel and central 3-10 helix region, marked by the large black square is ≈ 0.9 Å, and the largest compression is ≈ 0.7 Å. Note the smaller rectangles marking the expansion of the distance from the central 3-10 helix to the section of the β-barrel containing Tyr203.


Main Chromophore
H148
Y203
> 0.4 Å
< -0.6 Å
Residue Number
Residue Number
230
200
150
100
50
0
-2
0
50
100
150
200
230
Central
3-10
Helix
β-Sheet
N-terminal
α-helix
Secondary
Structure

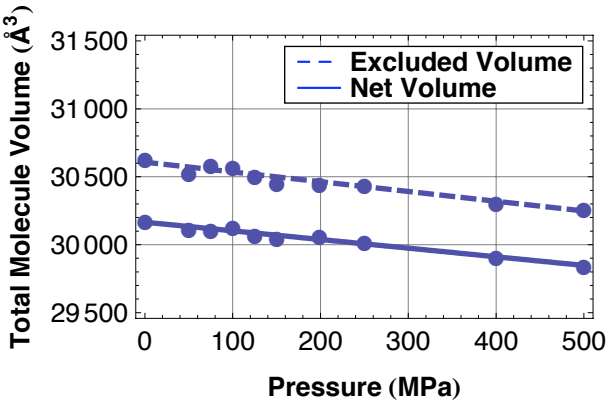


Figure 5.4: Averaged excluded volume of the Citrine molecule (dashed line), and averaged net volume of the Citrine molecule (non-dashed line). The net volume is computed by subtracting the internal cavity volume from the excluded volume of the molecule.

5.3.2 Volume Reduction of Citrine Molecule Under Pressure

The overall volume of the Citrine molecule shrinks under high pressure. The external surface and surfaces of internal cavities present in each solvent-stripped Citrine structure were identified and traced with the reduced surface computation program MSMS [190] using a 1.2 Å radius probe. The surfaces identified by MSMS were used to compute the volume enclosed by the external surface of each Citrine structure (the *excluded volume*) and the volumes of the cavities present in the interior of each Citrine structure. The *net volume* of each structure was computed by subtracting the total internal cavity volume from the excluded volume of the structure. Plots of the averaged excluded volume and averaged net volume of the solvent-stripped Citrine structures are shown as a function of pressure in figure 5.4.

Both the excluded volume and net volume of Citrine decrease by approximately 300 $Å^3$, or 1.1% over the 500 MPa pressure range investigated. The equal reduction of the excluded and net volumes of Citrine indicates that the volume reduction of the

molecule is not accounted for solely by reduction in internal cavity volume. The volume reduction of Citrine over the 500 MPa pressure range gives an isothermal compressibility of

$$\beta = \left(-\Delta V / \Delta P\right) / V_{\text{Initial}} = 2.35 \text{ MBar}^{-1} = 2.35 \times 10^{-2} \text{ GPa}^{-1} \tag{5.3}$$

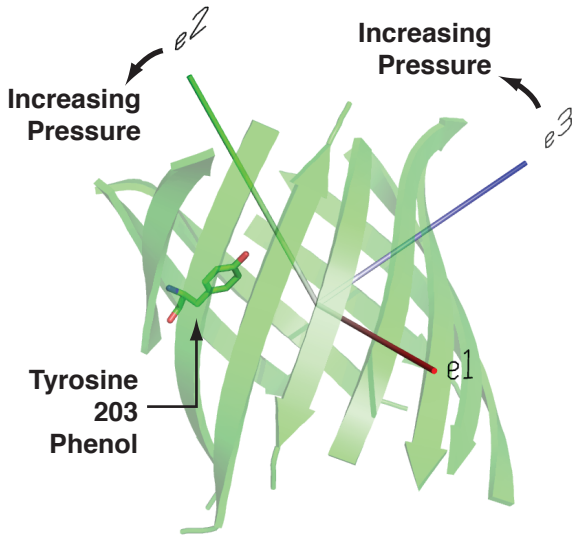
This value of the isothermal compressibility is at the low end of the protein compressibility range reported by Heremans and Smeller [194]. The inverse of the isothermal compressibility, the isotropic bulk modulus of Citrine is κ = 42.5 GPa. For comparison, the bulk modulus of steel is approximately 160 GPa and for aluminum it is 76 GPa.

5.3.3 Non-Isotropic Volume Reduction of Citrine Under Pressure

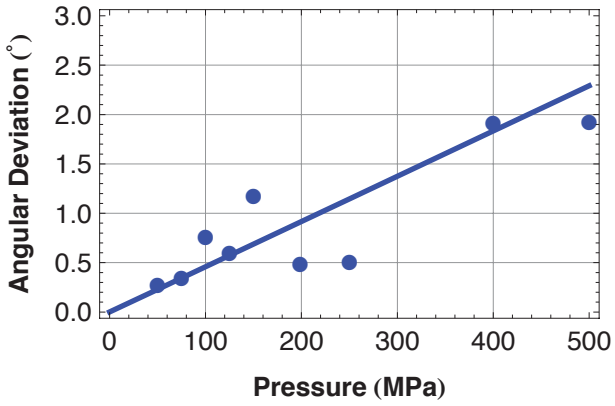
As protein molecules have non-homogeneous structures, it is reasonable to expect that the structural response to pressure will not be uniform throughout the structure. The expectation of non-uniform compressibility is supported by the observation of multiple domains of differential compressibility under pressure in the structure of hen egg white lysozyme by Kundrot and Richards [24] and Refaee *et al*. [26], in the structure of T4 lysozyme under pressure by Collins *et al*. [29], and in the differing response to pressure of the α-helices forming the structure of sperm whale myoglobin [4] that was discussed in chapter 1. If the compression of Citrine under high pressure were isotropic and homogeneous, we would expect that the relative orientation and spatial arrangement of the two elements of the chromophore would be retained as the pressure applied was increased. However, we observe that the two elements of the chromophore slide apart with very little change in their vertical separation as the pressure applied to the molecule is increased [5].

Figure 5.5: A: Orientation of the principal axes of the Citrine β-barrel walls at ambient pressure. Arrows indicate the direction of rotation of principal axes 2 and 3 with increasing pressure. B: Angular deviation of principal axis 3 with increasing pressure. C: Variation of the eigenvalues of the inertia tensor of the β-barrel walls with increasing pressure. Eigenvalues 1 (red) and 3 (blue, dotted) reduce by approximately 1% over the pressure range from 0.1 to 500 MPa. Eigenvalue 2 (green, dashed) increases by approximately 0.2% over the same pressure range.

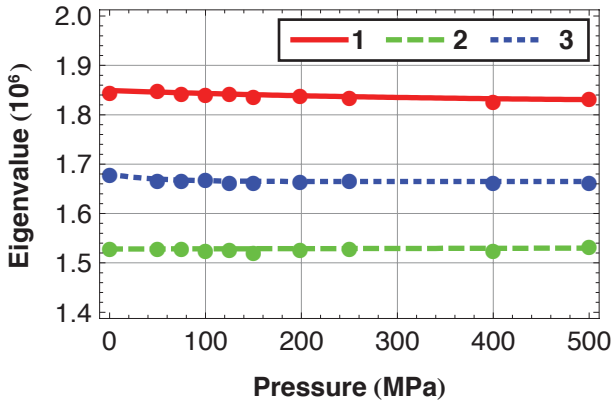
A: Orientation of β-Barrel Principal Axes



B: Angular Deviation of Principal Axis 3



C: Eigenvalues of β-Barrel Inertia Tensor



This observation of the sliding of the chromophore elements suggests, despite Citrine's overall linear volume reduction with pressure, that Citrine does not compress uniformly.

The principal axes of the Citrine β-barrel wall were computed [195] to provide more intuitive insight into the pressure-induced deformation of Citrine. The residues involved in β-sheet interactions in the β-barrel region of the Citrine molecule were extracted from the series of high-pressure Citrine structures. Analysis was limited to the β-barrel wall to avoid being dominated by the disordered regions. The α-carbons of the extracted structures were first aligned using the least squares fitting algorithm incorporated into PyMOL (DeLano Scientific LLC, Palo Alto, CA, USA) prior to computing the inertia tensor of each structure. The principal axes of each β-barrel structure were found by computing the eigenvectors of the structure's inertia tensor and associated eigenvalues using the NumPy numerical library. A diagram showing the orientation of the principal axes of the Citrine β-barrel at ambient pressure is shown in figure 5.5A. The number 2 principal axis is directed approximately along the cylindrical symmetry axis. Principal axes 1 and 3 are mutually perpendicular and lie approximately in the plane of cylindrical symmetry.

As the Citrine molecule is pressurized, the number 1 β-barrel wall principal axis remains largely fixed in direction relative to the reference frame defined by the alignment of the β-barrel structures. The other two principal axes, 2 and 3, approximately rotate about principal axis 1. A plot of the averaged angular deviation of principal axis 3 from its room pressure direction is shown in figure 5.5B. The angular deviation of principal axis 3 from its room pressure direction proceeds linearly up to the maximum pressure observed of 500 MPa at a rate of ≈ 0.5˚ per 100 MPa.

The directions of drift of principal axes 2 and 3 are shown in figure 5.5A. The slight re-orientation of the principal axes of the Citrine β-barrel wall indicates a redistribution of mass on the wall of the Citrine β-barrel. The direction of the drift of the principal axes suggests a motion of mass towards the side of Citrine β-barrel to which the perturbing tyrosine 203 ring is attached (see figure 5.5A).

The eigenvalues of the inertia tensor of the Citrine β-barrel wall are an indicator of the radius of the β-barrel wall and the height of the β-barrel. For an ideal, thin walled cylinder of mass *m*, radius *r* and height *h*, the eigenvalue associated with the principal axis aligned with the cylindrical symmetry axis is

$$I_z = mr^2/2 \tag{5.4}$$

The eigenvalues associated with the principal axes lying in the cross section of the thin walled cylinder are

$$I_x = I_y = m\left(3r^2 + h^2\right)/12 \tag{5.5}$$

Given that the amino acid mass of the cylinder is constant, the eigenvalues are sensitive indicators of the dimensions of the cylinder.

While the eigenvalues associated with principal axes 1 and 3 reduce with pressure, the eigenvalues associated with principal axis 2 remains relatively constant. A plot of the eigenvalues associated with the β-barrel principal axes is shown in figure 5.5C. The reduction of eigenvalues 1 and 3 is consistent with the overall compression of the molecule under pressure of approximately 1.1% over 500 MPa. However, the

constancy of eigenvalue 2 suggests that the side of the β-barrel wall close to principal axis 2 may not be compressing. This observation indicates that rather than isotropically compressing, part of the wall of the β-barrel in the vicinity of principal axis 2 may be retaining its original radius of curvature, appearing to bulge out as other sections of the wall reduce in radius of curvature. The Citrine molecule β-barrel appears to be buckling or bending under pressure, in a manner reminiscent of the buckling of a strut or the bending of a bimetallic strip under heating.

5.3.4 Clustering Analysis of Citrine Compression

To further quantify the non-isotropic compression of the Citrine structure under high pressure, we sought to identify groups of atoms in the Citrine structure that move in concert under the application of high pressure. The high pressure structures of hen egg white lysozyme [24, 26] and of T4 lysozyme [29] show the existence of domains of differing compressibility.

The existence of groups of residues that move in concert with increasing pressure was first suggested by manual inspection of animated distance difference matrices of the Citrine structure. Each frame in the animation corresponded to a colored, averaged distance difference matrix between the ambient structures of Citrine and the structures at an elevated pressure.

This visual analysis hinted at the presence of residues that maintained an approximately constant distance between one another with increasing pressure, yet showed varying distances between themselves and other residues not in the cluster. However, this visual analysis was tedious, and somewhat sensitive to the vagaries of the eye and mind of the inspector. The eye of the author is much more sensitive to the

motion present in these animations, and is poor at selecting these patterns in a static image. For this reason, while it was easy to gain a sense that some sort of concerted motion was occurring in the Citrine molecule with elevated pressure, it was hard for the author to definitively draw boundaries between clusters. Thus, an automated procedure was desired, that would analyze these distance difference maps, and examine them for internal correlations. The proprietary clustering algorithm RIGIMOL that is incorporated into the molecular graphics and manipulation program PyMOL (DeLano Scientific LLC, Palo Alto, CA, USA) performs this task: automated analysis of distance difference matrices between structures. RIGIMOL was originally developed to interpolate between two or more distinct structures of a protein, for instance between the open and closed states of a hinged protein, and produce a trajectory between the two states, for the purposes of producing an animation of the trajectory.

RIGIMOL rapidly identified what we believe to be the defining feature of the Citrine deformation: the presence of two groups of residues that compose the β-barrel. As protein domains tend to be visually distinguishable, while the Citrine molecule is at first inspection, cylindrically symmetrical, these groups of residues were termed *clusters*. It was very important to definitively identify the cluster assignment produced by the RIGIMOL algorithm, to reduce the possibility that the identification of these two clusters was simply an artifact of the clustering algorithm, perhaps due to random coordinate error due to coordinate uncertainty in the Citrine structure. For this reason, the clustering algorithm was re-run many times with different Citrine structures, different portions of the Citrine structure, different refinement model-sets and different values for the clustering algorithm parameters. The reproducibility of the clustering assignment was checked in each of these cases.

While these clustering-assignment reproducibility experiments consistently identified the presence of two clusters in the β-barrel of the Citrine molecule, the location, relative size, and boundary between these two clusters varied from experiment to experiment.

The RIGIMOL algorithm permits the user to select a variety of clustering parameters including the maximum allowed aspect ratio of a cluster, threshold positional and angular displacements above which atoms are no longer considered part of a cluster and the minimum size of a cluster. It was found that the cluster-assignment, at least in the case of Citrine, was largely insensitive to the clustering algorithm parameters. As an aside, the RIGIMOL algorithm was applied to a series of high-pressure cryocooled T4 lysozyme structures, and the identification of clusters of residues in this series of structures was found to be highly dependent upon clustering parameters.

The cluster-assignment reproducibility experiments, found that the clustering assignment is heavily dependent upon the choice of input structures. RIGIMOL takes a series of structures as an input. However, RIGIMOL cannot interpret sequences of structures containing multiple equivalent structures, for instance, two structures at the same pressure. For this reason, multiple sequences with different structures at each pressure were tried. The experiments were also repeated with sequences of structures derived from both model-sets, refinement set 1 and refinement set 2, discussed in chapter 4. In each of these cases, for the full Citrine molecule, the results of the clustering algorithm were sensitive to the choice of input structures.

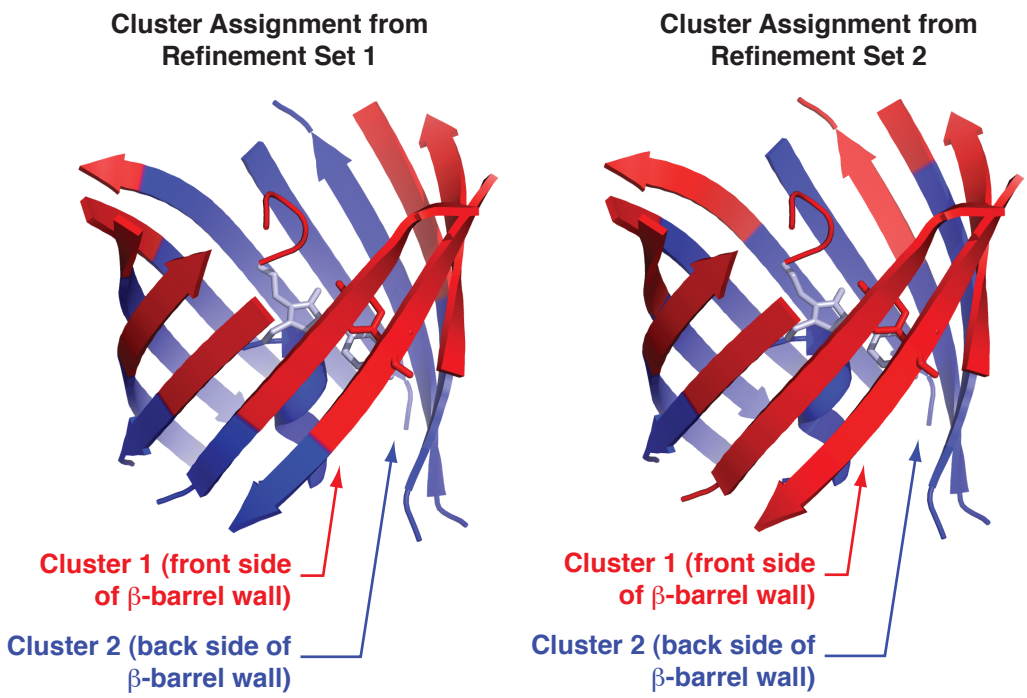


Figure 5.6: Comparison of the results of RIGIMOL cluster assignment using structures from refinement sets 1 and 2. Cluster 1 is colored red and cluster 2 is colored in blue. The main chromophore, colored in light blue, is attached to cluster 2, while tyrosine 203, colored in red, is attached to cluster 1.

The sensitivity in the results of the clustering assignment to the choice of input structures is intuitively understandable. Some regions of the Citrine structure: the N-terminal α-helix and the C-terminus show disorder and are highly variable from structure to structure, even at the same pressure. For this reason, it was suspected that random errors in the positions of atoms in these regions would bias the clustering analysis, and mask the real, but smaller pressure-induced concerted motions present in the well-ordered β-barrel scaffold. For this reason, we focused on finding a subset of the Citrine structure that would produce highly similar clustering results independent of the choice of input structures.

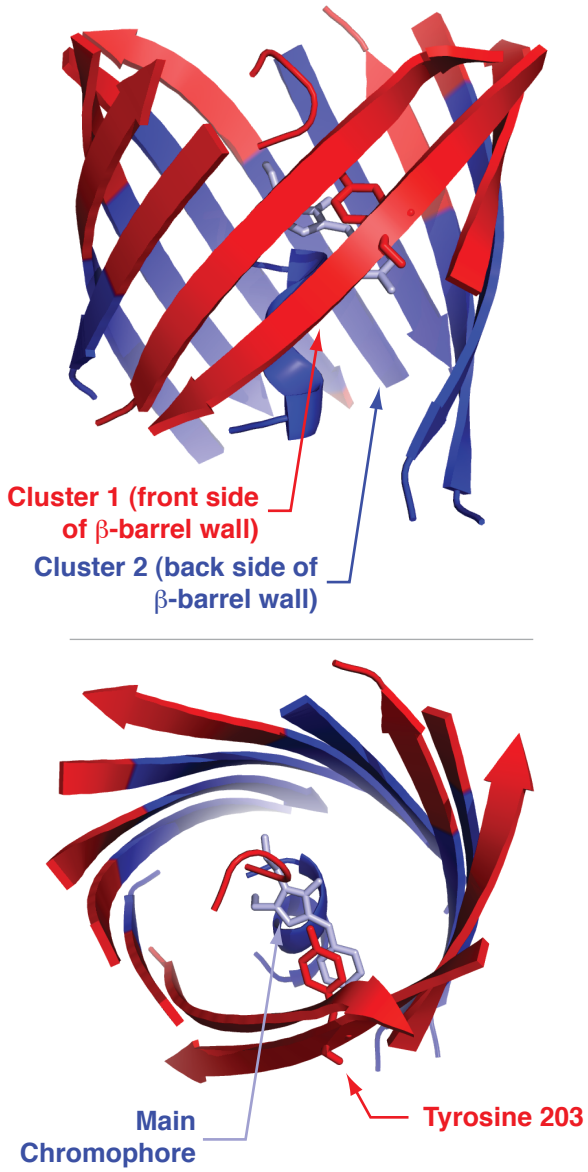


Figure 5.7: Clusters identified in the β-barrel walls and central 3-10 helix of the Citrine molecule that move in concert under high-pressure perturbation. Note that the main chromophore is attached to cluster 2, while the perturbing tyrosine 203 ring is attached to cluster 1. The main chromophore is colored in light blue for emphasis.

We found that the subset of the Citrine molecule that produced the most reproducible clustering assignment consisted of the central 3-10 helix and β-barrel walls, without the floppy β-strand composed of residues 143 to 148. Figure 5.6 shows the results of the clustering analysis for refinement sets 1 and 2. The results of the clustering assignment, using the reduced Citrine structure shown in figure 5.6, appear to be highly similar, independent of refinement set.

The RIGIMOL clustering algorithm identified two clusters of atoms in the β-barrel and central 3-10 helix of Citrine that move with respect to one another as the pressure applied to the Citrine molecule is increased. The two clusters are shown as colored regions in cartoon images of the Citrine structure in figures 5.6 and 5.7. Cluster 1 is colored in red, while cluster 2 is colored in blue. An important feature of the cluster assignment is that the perturbing tyrosine 203 is attached to cluster 1, while the main chromophore is attached to cluster 2. A relative motion of the two clusters under increasing pressure could actuate the motion of the main chromophore and tyrosine 203 that has been reported [5].

To further quantify the non-isotropic compression of the Citrine structure under pressure, the relative motion of clusters of residues was analyzed in the series of high-pressure Citrine structures. The β-barrel and the central 3-10 helix of the Citrine molecule were extracted from the series of high-pressure Citrine structures and aligned using the least squares fitting algorithm incorporated into PYMOL (DeLano Scientific LLC, Palo Alto, CA, USA). The heuristic clustering algorithm RIGIMOL (DeLano Scientific LLC, Palo Alto, CA, USA) was used to identify clusters of atoms in the extracted structures that move in concert with increased pressure.

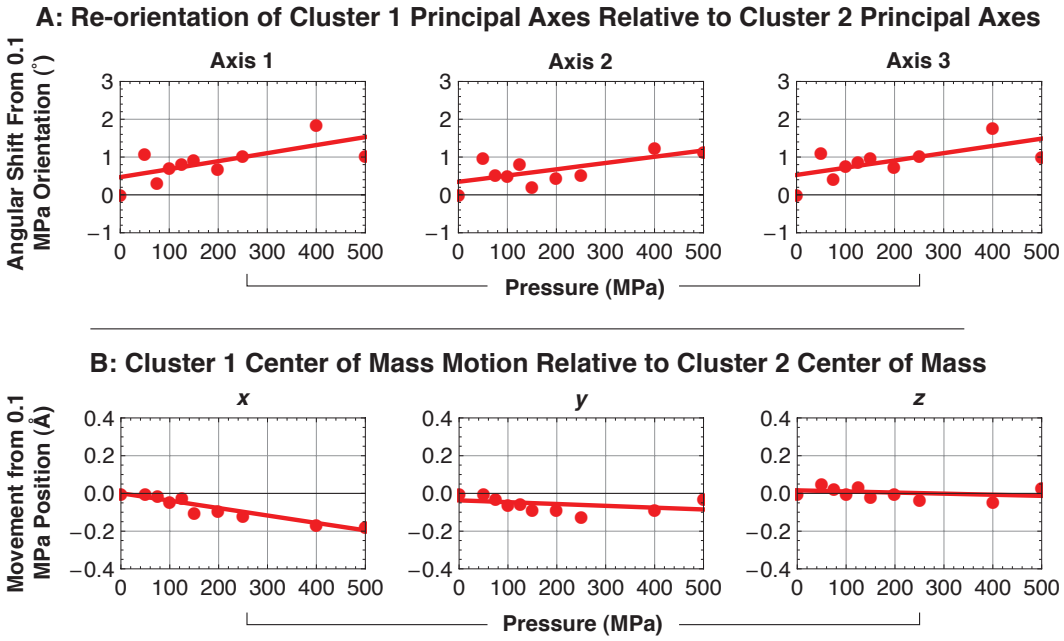
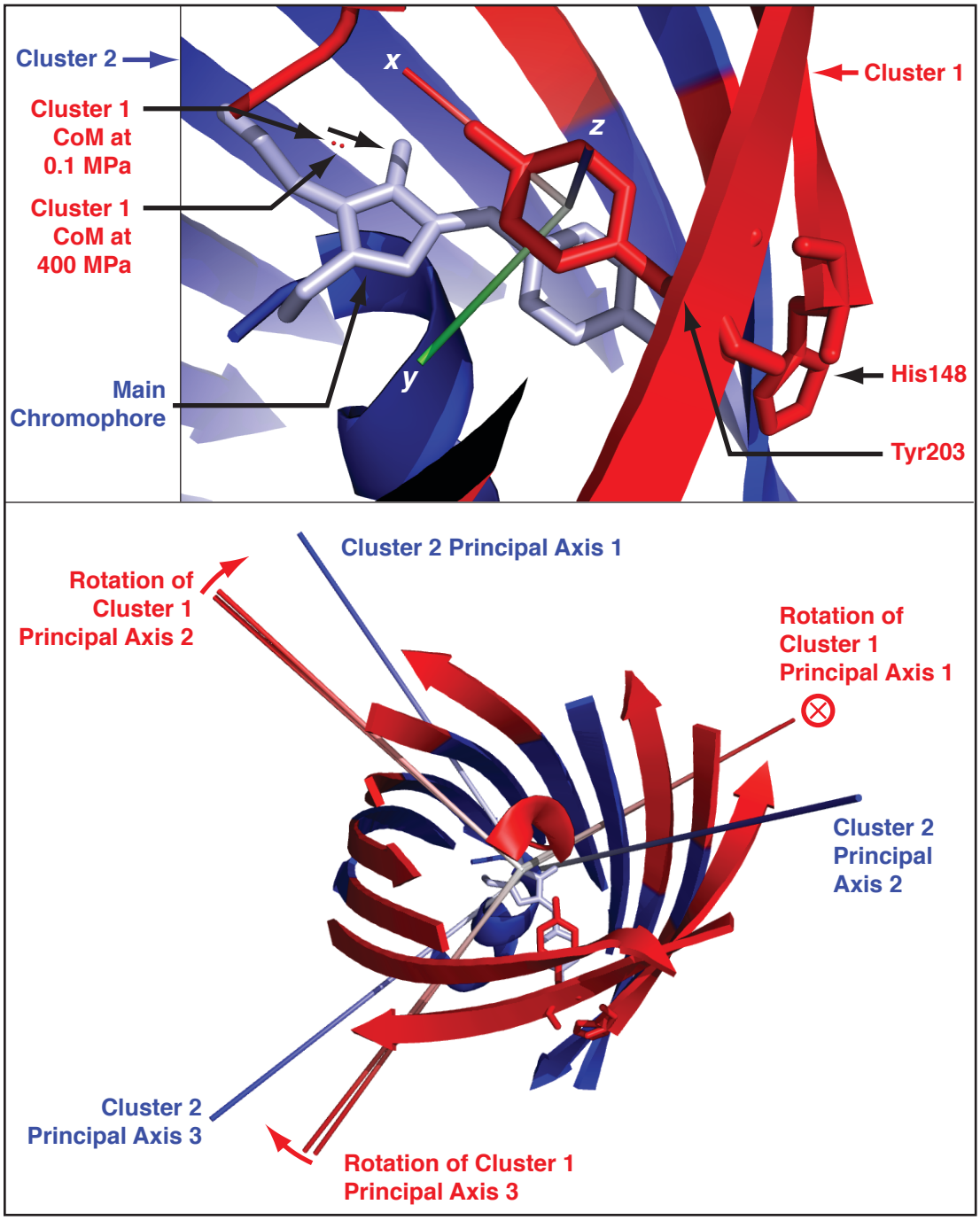


Figure 5.8: A: Rotation of the principal axes of cluster 1 in a reference frame where the principal axes of cluster 2 are fixed. B: Motion of the center of mass of cluster 1 from its ambient pressure position in a reference frame where the cluster 2 center of mass is fixed. The directions of the coordinate axes of this system are shown in figures 1 and 9 and are the same as used in chapter 4. Fit lines in all plots were determined by least squares fitting with the NUMPY (http://numpy.scipy.org) numerical library.

The centers of mass of the two clusters composing the Citrine β-barrel and central 3-10 helix move relative to one another with increasing pressure, and their principal axes rotate relative to one another. The relative motion of the centers of mass of the two clusters is small, but shows a strong trend with increasing pressure. The relative displacement motion of the center of mass of cluster 1 from its ambient pressure position in a reference frame where the center of mass of cluster 2 is fixed is shown in figure 5.8A. For ease of reference, the coordinate axes used in this discussion are the

Figure 5.9: Upper panel: Motion of the cluster 1 (shown in red) center of mass in a reference frame where the center of mass of cluster 2 (shown in blue) is fixed. Lower panel: rotation of the cluster 1 principal axes relative to the cluster 2 principal axes. The circled cross (⊗) next to cluster 1 principal axis indicates that the direction of rotation of the axis is rotating into the figure. The main chromophore is colored in light blue for emphasis.


Cluster 2
Cluster 1
CoM at
0.1 MPa
Cluster 1
CoM at
400 MPa
Main
Chromophore
x
z
y
Cluster 1
His148
Tyr203
Cluster 2 Principal Axis 1
Rotation of
Cluster 1
Principal Axis 2
Rotation of
Cluster 1
Principal Axis 1
Cluster 2
Principal
Axis 2
Cluster 2
Principal Axis 3
Rotation of Cluster 1
Principal Axis 3

same as those used in chapter 4 in the description of the deformation of the chromophore of Citrine [5]. The *x* and *y* axes of this coordinate system lie in the plane of the tyrosine 203-phenol ring, and the *z*-axis is normal to the ring. This coordinate system is shown in figures 4.43 and 5.9.

The center of mass of cluster 1, containing the perturbing tyrosine 203 ring, moves by approximately +0.2 Å in the *y*-direction, and +0.05 Å in the *x*-direction with respect to the center of mass of cluster 2. There is no discernible motion in the *z*-direction. These motions correspond well to the direction of the motions of the chromophore seen in our earlier work [5], where the perturbing tyrosine 203 ring moves by approximately +0.5 Å in the *y*-direction, +0.4 Å in the *x*-direction and relative to the main chromophore.

In addition to the relative motions of the centers of mass of clusters 1 and 2, the principal axes of the two clusters also slightly rotate with respect to one another. In a reference frame where the principal axes of cluster 2 are fixed, the principal axes of cluster 1 rotate by approximately 2˚ from their ambient pressure orientations. This rotation is shown in figure 5.8A.

The motion of the center of mass of cluster 1 relative to cluster 2, coupled with the slight rotation of the principal axes of cluster 1, appear to produce sufficient leverage to actuate the relative motion of the two elements of the Citrine chromophore: the main chromophore, attached to cluster 2, and the perturbing tyrosine 203 ring, attached to cluster 1. The rotation of the principal axes and motion of the center of mass of cluster 1 relative to cluster 2 are shown in figure 5.9.

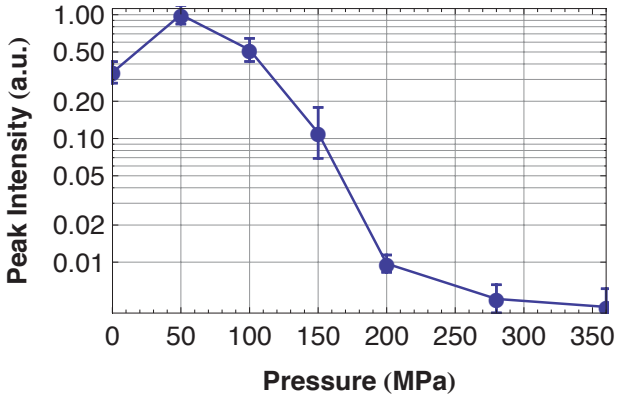


Figure 5.10: Fluorescence peak intensity of high pressure cryo-cooled Citrine solution samples.

5.3.5 Compression and Expansion of Links in Hydrogen Bonding Network in Chromophore Cavity

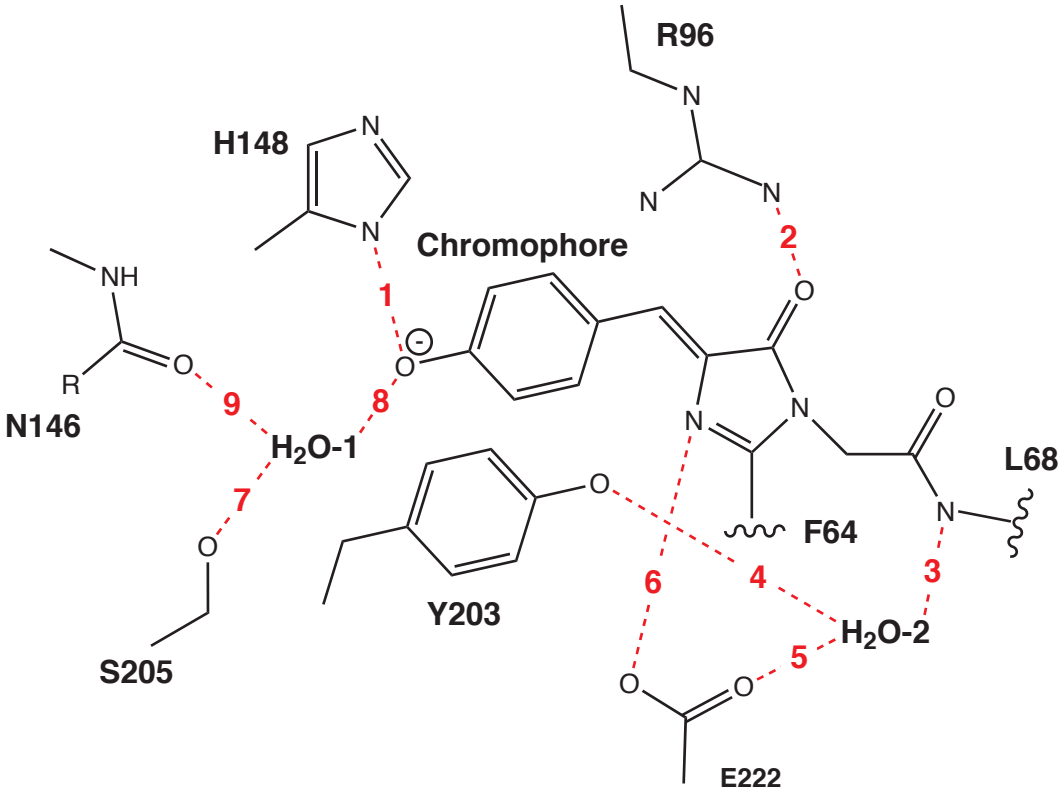
In chapter 4, we discussed the fluorescence peak shift of the Citrine molecule with increasing pressure. Accompanying this fluorescence peak shift is a considerable reduction in the fluorescence intensity of the molecule, implying a reduction in quantum yield. The fluorescence peak intensity of Citrine solution samples, high pressure cryo-cooled at a range of pressures from 50 to 360 MPa, is shown in figure 5.10. The fluorescence peak initially increases, from 0.1 MPa to 50 MPa, and then decreases with increasing pressure up the maximum observed pressure of 360 MPa. By a pressure of 200 MPa, the peak Citrine fluorescence intensity is approximately $1/100^{th}$ of its value at 50 MPa.

Experiments on other *Aequorea* fluorescent protein mutants suggest that this reduction in fluorescence intensity (see figure 5.10) may be due to pressure-induced disruption of the hydrogen bonding network [52] present in the chromophore cavity. This hydrogen bonding network serves to anchor the main chromophore, stabilizing the excited state of the molecule, preventing non-radiative decay and facilitating the high

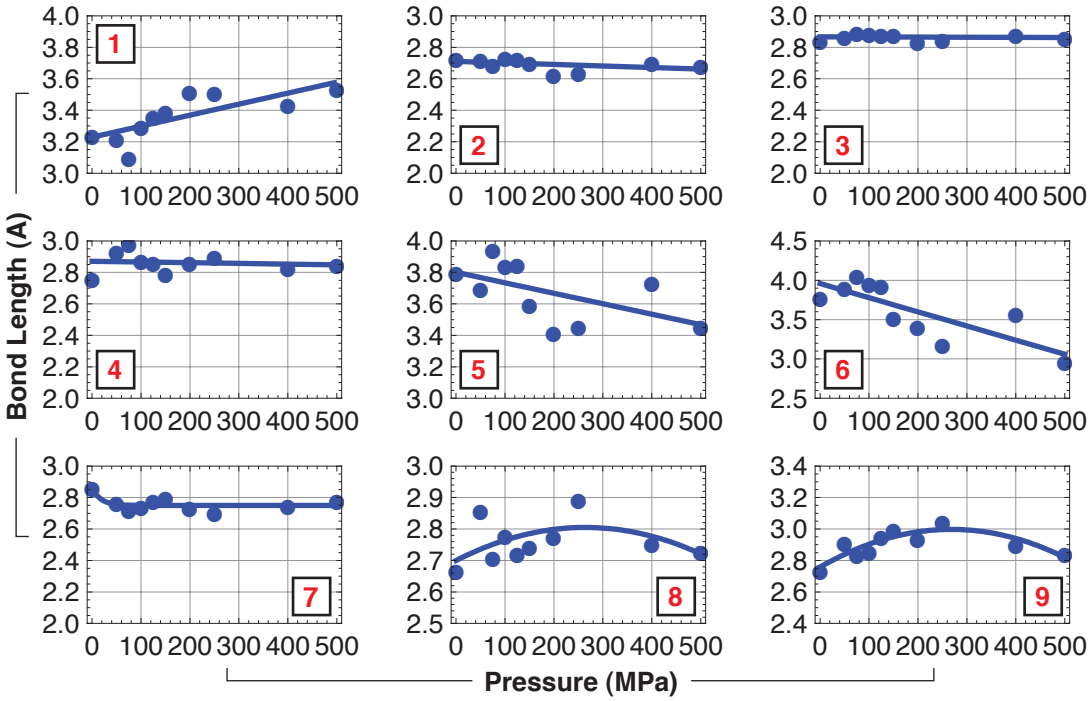
quantum yield of the molecule [196]. Niwa *et al.* [155] noted that certain analogs of the *Aequorea* fluorescent protein chromophore (isolated from the β-barrel) absorb light yet are non-fluorescent at room temperature. However, when these chromophore analogs and their solvent are frozen to cryogenic temperatures, they become fluorescent [155]. Computer simulations indicate that in a vacuum environment the chromophore is fluorescent, while in a liquid environment, the excited state of the molecule de-excites by quenching [156]. Mauring *et al*. [196] demonstrated that the application of high pressure to an *Aequorea* Blue Fluorescent Protein (BFP) increased the quantum yield of the molecule. The central residue of the BFP chromophore is a histidine, a shorter residue than the tyrosine normally found at the center of an *Aequorea* fluorescent protein chromophore. This shortened chromophore is unable to link to the chromophore cavity hydrogen bonding network. This loss of bonding interactions results in the destabilization of the BFP chromophore excited state. As a result, BFPs display lower quantum yields than other *Aequorea* fluorescent proteins [196]. Mauring *et al*. [196] attributed the pressure-induced increase in the quantum yield to the re-attachment of the BFP chromophore to the chromophore cavity hydrogen bonding network as distances from residue to residue in the cavity were reduced as the molecule compressed.

Figure 5.11: A. Hydrogen-bonding network in Citrine chromophore cavity. B. Variation of lengths in chromophore cavity with pressure.

A: Hydrogen Bonding Network in Chromophre Cavity



B: Variation in Bond Lengths in Hydrogen Bonding Network



While the behavior of Citrine and EYFP (Enhanced YFP) [53] under high pressure is the opposite of BFP: the fluorescence intensity reduces, the structural mechanism may be the same: the expansion or contraction of chromophore cavity hydrogen-bonding network. This speculation is supported by the existence of the REACh1, a YFP with reduced quantum yield developed by Ganesan *et al*. [197]. The H148V mutation in REACh1 removes a critical link in the hydrogen-bonding network attaching the phenolic oxygen of the main chromophore to the His148 δ_1-nitrogen and results in an 82% reduction in the fluorescence intensity with only a 15% drop in absorption [197].

In order to investigate the structural origin of the fluorescence intensity reduction of Citrine with increasing pressure (figure 5.10), the length of several of the hydrogen bonds in chromophore cavity were investigated as a function of pressure. A simplified schematic of the hydrogen-bonding network stabilizing the Citrine chromophore is shown in figure 5.11A. Plots of the averaged length of the hydrogen bonds anchoring the main chromophore are shown in figure 5.11B.

Most of the hydrogen bonds in the network show very small changes. The distance between the Leu68 amide nitrogen and the solvent molecule HOH-2 (bond 3 in figure 5.11) shows no variation with pressure. The chromophore oxygen-2 to Arg96 amide nitrogen distance (bond 2 in figure 5.11) varies by less than 0.1 Å over 500 MPa, as does the Tyr203 phenolic oxygen to HOH-2 distance (bond 4 in figure 5.11). The Ser205 γ-oxygen to the HOH-1 solvent molecule distance (bond 7 in figure 5.11) compresses by 0.1 Å over the pressure range 0.1 to 50 MPa, and then remains constant up to a pressure of 500 MPa. The bond between the chromophore phenolic oxygen and HOH-1 (bond 8 in figure 5.11) increases in length by ≈ 0.1 Å at a pressure ≈ 250 MPa, and then recompresses to its original length by 500 MPa. The bond between the

carbonyl oxygen of Asn146 and HOH-1 (bond 9 in figure 5.11) also expands by ≈ 0.25 Å at a pressure of 250 MPa, and the recompresses by 0.2 Å by 500 MPa.

Three bonds that do show large variations with pressure are those between the His148 δ_1-nitrogen and the chromophore phenolic oxygen (bond 1 in figure 5.11), the Glu222 ε_1-oxygen and the HOH-2 solvent molecule (bond 5 in figure 5.11), and chromophore nitrogen-2 and Glu222 ε_2-oxygen (bond 6 in figure 5.11). The bond 1 (figure 5.11) distance increases by 0.4 Å over 500 MPa, while bond 5 and bond 6 (figure 5.11) compress from 3.8 to 3.4 Å and 4.0 to 3.0 Å over 500 MPa, respectively. Ganesan *et al*. [197] demonstrated that elimination of bond 1 by mutation results in a YFP mutant with much lower quantum yield. We speculate that the pressure-induced increase in this bond length is responsible for the reduction in fluorescence intensity of Citrine shown in figure 5.10.

5.4 Discussion

The deformation of the Citrine scaffold can be thought of as the slight bending or buckling of the β-barrel scaffold. A cartoon representation of the deformation of the Citrine scaffold under pressure is shown in figure 5.12. This cartoon representation accentuates the relative movement of the two clusters composing the β-barrel and 3-10 helix. As the pressure applied to the molecule is increased, the main chromophore remains anchored inside cluster 2, attached to the central helix. However, the side of the β-barrel wall containing the perturbing tyrosine 203 ring and stabilizing His148 residue, moves away from cluster 2 (the side of the β-barrel wall containing Tyr203 bulges out), carrying these two important residues with it. The large expansion of the distance from the central 3-10 helix to the wall of β-barrel in the vicinity of Tyr203 can be seen in the highlighted region in the distance difference matrix in figure 5.3.

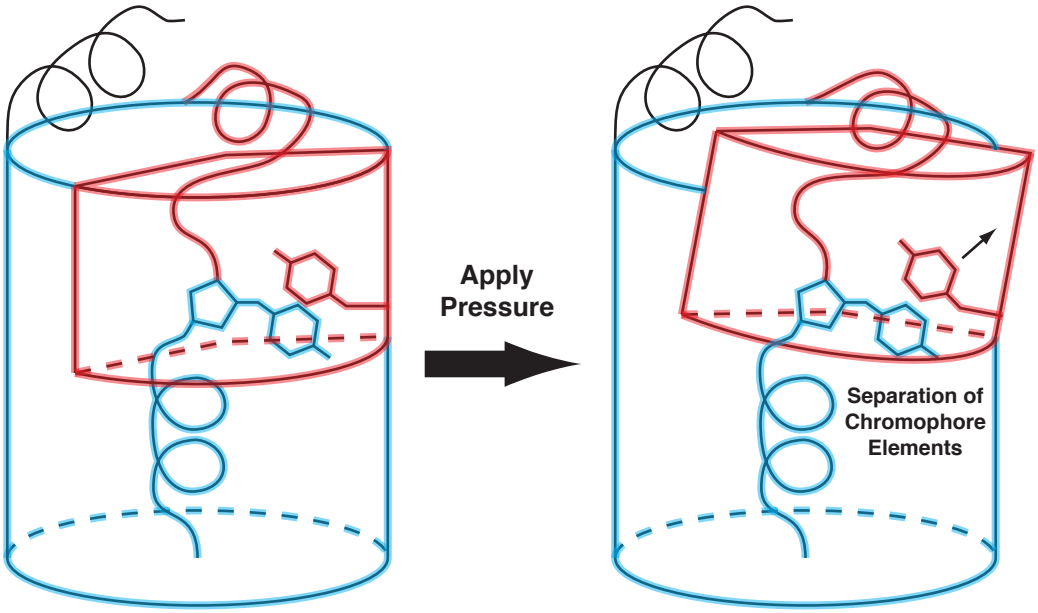


Figure 5.12: Cartoon representation of the bending of the Citrine scaffold.

It is interesting to note that the fluorescence peak intensity of all *Aequorea* fluorescent proteins studied under pressure [5, 53] increases with applied pressure, with the exception of the YFP types. Verkhusha *et al*. [53] measured the fluorescence peak and peak intensity of EYFP (Enhanced Yellow Fluorescent Protein), EGFP (Enhanced Green Fluorescent Protein), ECFP (Enhanced Cyan Fluorescent Protein) and DsRed (a red fluorescent protein that is structurally highly similar to GFP, yet is derived from a corallimorpharian from the genus *Discosoma* [198, 199]) under high hydrostatic pressure at room temperature. Verkhusha *et al*. [53] found that of the four fluorescent proteins studied, only the fluorescence peak of EYFP (a close relation of Citrine), shifted under the application of pressure. The fluorescence peak intensities of EGFP, ECFP and DsRed all increased with the application of pressure, while only EYFP showed a decrease in fluorescence intensity with increasing pressure [53]. At cryogenic temperatures, (see figure 5.10), we also observed a reduction in the fluorescence intensity of Citrine with increasing freezing pressure.

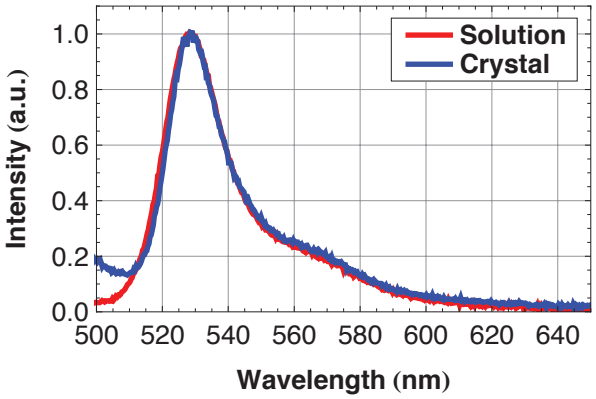


Figure 5.13: Comparison of the fluorescence spectra of a flash frozen Citrine solution and a very small, flash frozen Citrine crystal. The Citrine solution is composed of 1 mg/mL Citrine and 5% PEG 3350, 50 mM Na Acetate, 50 mM NH_4 Acetate, pH 5.0, the same buffer conditions as the crystal.

We speculate that the unique spectral response of the YFPs under pressure, at room temperature [53] and at cryogenic temperatures (figure 5.10), may be due to a YFP structure that predisposes the YFP-type molecules to the bending behavior observed here.

We do not believe that the crystalline environment significantly affects the fluorescence properties of the Citrine molecule. The fluorescence spectra of a very small, flash frozen Citrine crystal and a solution sample of Citrine are highly similar. These two spectra are plotted together in figure 5.13 for comparison. The similarity of these two spectra suggests that the crystalline environment does not significantly perturb the fluorescence properties of Citrine. As pressure-induced unfolding is not expected to occur in the solution state of citrine below at least 900 MPa [148, 193], we also believe that the pressure effects captured in the crystalline state represents those in the solution state within the explored pressure range. This is reasonable as the

deformation of citrine in the crystal state is not a response to a directionally applied external force. As opposed to typical small molecule crystals, the fluid water channels intrinsic to most protein crystals constitute approximately half of the crystal mass and transmit applied hydrostatic pressure to each protein. In other words, there is no pressure gradient across the volume of the crystal: identical, uniform hydrostatic pressure is effectively transmitted by the solvent channels to each molecule. A force originating at the surface of the crystal is not transmitted through covalent or metal bonds from molecule to molecule as happens in a small molecule crystal or a metal. Thus the protein deformation in response to pressure is not due to an externally applied force *per se*; rather it results from pressure-dependent interactions of the molecule. For example, it is known that the degree of ionization [200], hydration [162, 200], and hydrogen bonding [200, 201] of many amino acid residues are pressure dependent. It has also been shown that pressure changes the water occupancy of internal cavities [28, 29, 202, 203]. One expects that as these interactions change with pressure, the conformation of the protein will change in response. The fact that pressure is transmitted to the individual molecules in the crystal also argues that the resultant deformations are similar to those expected for the same molecule in solution, and differ only in so far as molecular contacts in the crystal change the surface exposure to water or constrain large deviations in protein structure.

All *Aequorea* fluorescent protein display a floppy β-strand composed of residues 143 to 148. These residues do not form β-sheet interactions with the rest of the β-barrel scaffold. At room temperature, these residues appear to bulge slightly more out into solution in the Citrine structure when compared with the corresponding residues in the EGFP structure. It may be that this “weakness” predisposes the Citrine β-barrel wall to further bulge under pressure, rather than compressing. Additionally, at room

temperature, the main chromophore of YFP protrudes slightly further, by ≈ 0.9 Å, into the chromophore cavity than the chromophore in the EGFP structure [52]. These two pre-existing deformations, possibly induced by the presence of the perturbing tyrosine 203 residue, may slightly weaken the YFP structure. These "weaknesses" may predispose the YFP scaffold to the bending behavior seen under high-pressure cryocooling conditions in this chapter. This bending behavior may be analogous to macroscopic structures failing at the weakest point.

Although the structure of Citrine and the response of its structure to high pressure are no doubt subtly different to that of wild-type Green Fluorescent Protein (wtGFP), the extension in path length from the phenolic oxygen to the ε-oxygen of Glu222 may explain (or at least provide a framework for explaining) the subtle shift in peak position of wtGFP observed under high pressure [204]. In the Citrine structures presented here, the distance from the phenolic oxygen of the main chromophore to the HOH-1 solvent molecule to the ε_2-oxygen of Glu222 increases by approximately 0.5 Å from ambient pressure to 500 MPa. The increase in this distance from the main chromophore through the HOH-1 solvent molecule to Glu222 should have no impact on the fluorescence properties of Citrine as excited state proton transfer from the main chromophore to Glu222 appears to play no role in the fluorescence mechanism of *Aequorea* fluorescent protein mutants containing de-protonated chromophores (phenolate anion chromophores; this class of molecules includes mutants containing the S65T or S65G mutation such as Citrine and EGFP [2]). However, excited state proton transfer does play a role in the fluorescence mechanism of wtGFP, and the extension of the proton transfer path may alter the fluorescence properties of wtGFP [205, 206].

An important observation from that was presented in chapter 4 [5] is that the fluorescence peak of Citrine asymptotes to ≈ 510 nm, the fluorescence peak of Enhanced Green Fluorescent Protein (EGFP) under ambient conditions, as the pressure applied to the molecule is raised to above 350 MPa. This shift to the green of the fluorescence peak is due to the removal of the perturbation of the tyrosine 203 ring is from the main chromophore [5]. The results presented here, on the effect of high pressure on the scaffold of the Citrine molecule indicate that the sliding motion of the perturbing tyrosine 203 phenol ring relative to the main chromophore is actuated by the separation of two clusters that compose the β-barrel wall and central 3-10 helix. This observation may provide a structural explanation for the observation of Blum *et al*. [7] that a single molecule of another YFP [1, 2] occasionally becomes dark, and then adopts a spectrum closely resembling the spectrum of bulk EGFP [2, 142, 159], before returning to a spectrum close to that of bulk YFP. It is known that protein molecules are fluctuating molecules that adopt a number of structural conformations [207]. We speculate that at room temperature, YFPs may occasionally adopt a configuration structurally highly similar to the bent state seen at high-pressure. The bent scaffold of this state may transiently stabilize the main chromophore-Tyr203 separation needed for the molecule to fluoresce in the green. This observation may have wider implications for understanding the protein structure-function relationship. Single molecule experiments indicate that single enzymes display a range of catalytic rates [54, 55, 57-60, 208, 209], and switch between them, in much the same way that a single YFP molecule displays a range of spectra and switches between them [7]. This result suggests that the high-pressure cryo-cooling technique may be useful for accessing the atomic structures of transiently populated enzyme states that are presently only observable in single molecule experiments. It is possible that the dark

state of YFP observed by Blum *et al.* [7] is one in which the stabilizing interaction between the main chromophore and His148 is completely lost.

An interesting feature of the high-pressure state of Citrine is the penetration of a water molecule into a cavity close to the main chromophore. This water molecule joins another molecule already present in the cavity at ambient pressure. This additional water molecule has not been observed in any ambient pressure structure of Citrine, either at room temperature (PDB accession code 1HUY [1]) or at 100 K (PDB accession codes 3DPW and 3DQO (chapter 4) [5]) but consistently appears in structures at pressures of 50 MPa and above. We do not believe that the presence of this water molecule directly affects the spectral properties of Citrine, for instance by quenching the chromophore, as it appears at low pressure, before many of the spectral changes to the Citrine molecule have occurred. However, this additional water molecule may indirectly affect the spectrum of Citrine, by assisting in actuating the bending of the molecule. The presence of this water molecule may be a necessary, but insufficient, condition for initiation of the transition to the green shifted state of Citrine.

5.5 Conclusions

The analysis presented in this chapter demonstrates that the scaffold of the Citrine molecule deforms under high-pressure. Rather than isotropically compressing, the high-pressure deformation of Citrine is non-homogeneous. We speculate that the response to pressure of Citrine may differ from the structural response of non-tyrosine 203 bearing *Aequorea* fluorescent proteins due to the presence of a pre-existing weakness in the β-barrel scaffold of Citrine in the vicinity of introduced tyrosine 203 mutation.

The Citrine β-barrel is approximately divided into two atomic clusters that move and rotate relative to one another under pressurization. The two elements of the Citrine chromophore, the perturbing tyrosine 203 ring and the main chromophore, are each attached to different clusters. The relative motion and rotation of the two clusters composing the β-barrel and central 3-10 helix of the Citrine molecule causes a deformation in the Citrine scaffold that is reminiscent of buckling or bending, actuating the separation of these two elements of the Citrine chromophore, resulting in a fluorescence shift of the molecule. We speculate that this bending deformation only occurs for tyrosine 203 bearing mutants of the Green Fluorescent Protein. In addition to actuating the fluorescence shift of the molecule, the bending of the Citrine scaffold also perturbs the hydrogen-bonding network stabilizing the main chromophore. The most important consequence of this hydrogen-bonding network perturbation is an increase in the distance from the main chromophore phenol to the His148 side chain. We speculate that the increase in the His148 to main chromophore distance results in a destabilization of the excited state of the chromophore and the consequent dimming of Citrine seen under high-pressure cryo-cooling conditions [5] and under high hydrostatic pressure [53].

The bent or buckled state of the Citrine scaffold may be highly structurally similar to the transiently observed blue-shifted state of YFP at ambient pressure [7]. This suggests that high-pressure X-ray crystallography may offer the possibility of solving the structures of transiently occupied enzymatic states and understanding the structural basis of their differing catalytic rates.

CHAPTER 6

DISCUSSION, CONCLUSIONS AND FUTURE EXPERIMENTS

6.1 Introduction

The experimental results and analysis presented in this thesis comprise one of only a small number of experimental examples that demonstrate the detailed coupling of deformations of the scaffold of a protein molecule to the active site and then to the function of the molecule.

This chapter first details an attempt to model the fluorescence shift of Citrine under pressure. This model uses the high-pressure atomic structures of Citrine presented in chapters 4 and 5 in combination with the Extended Hückel model of the fluorescence spectrum of Citrine presented in chapter 3.

The chapter then turns to future experiments with other *Aequorea* fluorescent proteins, to test inferences that have been drawn from the analysis of Citrine data, with the aim of learning more about the mechanical stability of protein structure, details of the mechanism of protein allostery and possibly more about the detailed mechanism of catalysis.

The chapter then reports upon experiments with other protein systems that we have investigated in the course of this thesis and summarizes the results thus far. Finally, the chapter suggests possible protein engineering experiments that have been inspired by our experiences with Citrine.

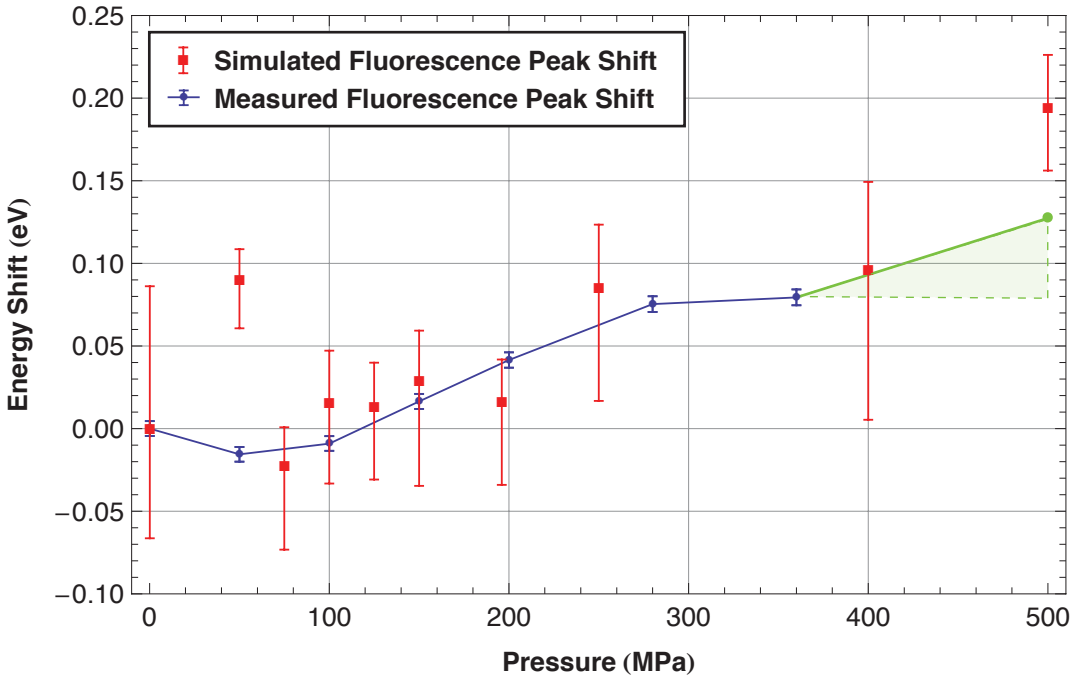


Figure 6.1: Calculated and observed fluorescence peak energy shift for Citrine under high-pressure cryocooling conditions. The final measured point, at 500 MPa, corresponds to the fluorescence peak energy shift of mEGP relative to Citrine at 0.1 MPa (527 versus 500 nm). The green shaded area should be considered as the location of the fluorescence energy shift of Citrine between 350 and 500 MPa. Calculation of the error bars on the calculated peak shifts is discussed in section 6.2.

6.2 Extended Hückel Calculation of Citrine's Fluorescence Peak Shift

6.2.1 Comparison of Fluorescence Band-gap Energy Shifts

In order to test the extended Hückel theory (EHT) model of Citrine's fluorescence peak shift that was discussed in chapter 3, and to provide a more substantial physical and chemical basis to this peak shift, we used the high-pressure structures of the Citrine as inputs to the YAEHMOP code, and calculated the corresponding fluorescence band-gap between the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) of the chromophore.

For the calculation presented here, we used the structures that were derived from refinement procedure 2, discussed in chapter 4, using the highest resolution diffraction data available in the refinement process.

The results of the extended Hückel computation are shown in figure 6.1. Figure 6.1 shows the measured and average calculated shift in energy of the fluorescence peak from its value at ambient pressure. The value the measured fluorescence band-gap shift at 500 MPa was taken as that of mEGFP relative to Citrine at 0.1 MPa (500 nm versus 527 nm) (figure 4.37). We consider this to be a reasonable estimate of the upper value fluorescence band-gap shift.

The calculated shifts shown in figure 6.1, $\Delta E_{\mathrm{c}}(P)$, from the ambient pressure HOMO-LUMO band-gap are compared with the measured energy shifts at each pressure,

$$\Delta E_{\mathrm{c}}(P) = E_{\mathrm{c}}(P) - E_{\mathrm{c}}(P_0). \tag{6.1}$$

For the measured fluorescence band-gap, the energy shift from ambient pressure, ΔE_{m}, was calculated from the fluorescence wavelength data shown in figure 4.37,

$$\Delta E_{\mathrm{m}}(P) = hc\left(\frac{1}{\lambda_{\mathrm{m}}(P)} - \frac{1}{\lambda_{\mathrm{m}}(P_0)}\right), \tag{6.2}$$

where P_0 is 0.1 MPa.

The measured and calculated fluorescence band-gap shifts, $\Delta E_{\mathrm{m}}(P)$ and $\Delta E_{\mathrm{c}}(P)$ respectively, are presented in terms of energy, rather than wavelength as the measured

and calculated baseline energies and wavelengths are different. We believe that it is unreasonable to expect this simple extended Hückel model of the Citrine fluorescence spectrum to correctly calculate the baseline HOMO-LUMO band-gap given the limited structure input into the model, and that much more sophisticated quantum chemical methodologies appear necessary to accurately estimate the absolute fluorescence peak of the *Aequorea* fluorescent proteins [210]. However, we do believe it reasonable to expect that this simple perturbation model could accurately calculate deviations from this baseline energy. However, if the results of this perturbation calculation were presented in term of wavelength, the calculated shift would appear inaccurate due to differences in baseline energies. To illustrate

$$\begin{aligned}
\Delta\lambda(P) &= \lambda(P) - \lambda(P_0) \\
\Delta\lambda(P) &= \frac{hc}{E_0 - \Delta E} - \frac{hc}{E_0} \\
\Delta\lambda(P) &= \frac{hc}{E_0^2}\left(\frac{E_0 - E_0 + \Delta E}{1 - \Delta E/E_0}\right) \cdot \\
\Delta\lambda(P) &= \frac{hc\Delta E}{E_0^2}\left(1 - \Delta E/E_0\right)^{-1}
\end{aligned} \tag{6.3}$$

Binomially expanding the inverse term,

$$\Delta\lambda(P) = \frac{hc\Delta E}{E_0^2}\left(1 + \Delta E/E_0 + \ldots\right). \tag{6.4}$$

As $\Delta E/E_0$ is on the order of 1/100th for wavelength shifts of several nanometers from baseline wavelengths of several hundred nanometers,

$$\Delta\lambda(P) \approx \frac{hc\Delta E}{E_0^2}. \tag{6.5}$$

Thus, as wavelength shifts are scaled by the inverse square of the baseline energy, comparison of two energy shifts from differing baseline energies can be misleading. The comparison of energy shifts, while less intuitive for an optical phenomenon such as a shift in fluorescence peak, is unambiguous.

Errors on the calculated fluorescence energy band-gap were estimated by a Monte Carlo procedure in which the main chromophore was randomly translated and rotated inside the error volume allowed by Cruickshank's positional uncertainty formula (equation 2.32). Translations and rotations were performed with the PyMOL molecular graphics and manipulation program (DeLano Scientific LLC, Palo Alto, CA, USA). For the simulation results presented in figure 6.1, 1000 random simulated perturbations were generated for each high-pressure structure. The HOMO-LUMO energy gap was calculated for each slightly perturbed structure with the extended Hückel theory program YAEHMOP. The highest and lowest energy band-gaps calculated in this procedure were recorded and taken as estimates upper and lower limits on the HOMO-LUMO band-gap for each high pressure Citrine structure, given the estimated positional uncertainty on the structure.

The calculated data points shown in figure 6.1 are the average of the calculated HOMO-LUMO band-gaps for the unperturbed structures at each pressure. The error bars shown in figure 6.1 correspond to the average highest and lowest values of the

energy band-gap calculated during this random translation and rotation procedure at each pressure. The baseline energy of the simulated band-gaps, was taken as the value of the energy band-gap of structure citrine0001_2 (PDB accession code 3DPW).

In general, the extended Hückel theory (EHT) procedure finds reasonable agreement between the measured and calculated shifts in the fluorescence band-gap. This result suggests that the mechanism of the fluorescence peak shift of Citrine under high-pressure cryocooling is the removal of the overlap of the orbitals of the tyrosine 203 phenol from the orbitals of the main chromophore. The result also supports the observed structural deformation of Citrine under high-pressure, suggesting that the deformation motion is real, rather than an artifact of refinement and data collection.

6.2.2 Limitations of Extended Hückel Theory Model

However, despite the apparent success of EHT in calculating the energy shift of the fluorescence peak of Citrine due to the observed structural perturbation under high pressure, it is appropriate to discuss the limitations of EHT and possible alternative mechanisms for the fluorescence shift.

Firstly, as we have noted before, the EHT model of the Citrine chromophore does not accurately calculate the Citrine fluorescence band-gap energy. At ambient pressure, the calculated fluorescence band-gap corresponds to a wavelength of ≈ 643 nm, while the observed fluorescence peak is 527 nm. As the errors on the calculated peak shift are comparable in magnitude to the peak shift (figure 6.1), it is difficult to confirm the accuracy of the calculation. Additionally, it is difficult to confirm if the model produces the correct results for the right reasons.

This discrepancy between the calculated and observed ambient pressure fluorescence band-gap may in part be due to the limited structure input into EHT model of Citrine's fluorescence spectrum. However, the Citrine scaffold is believed to not significantly tune the absorbance of the chromophore [156]. This suggests that the discrepancy in calculated and observed band-gaps is largely due to limitations intrinsic to EHT. Most importantly, EHT does not calculate the effects of electron-electron interaction. This limitation is particularly serious, as it does not permit calculation of changes in energy levels due to photo-excitation of the chromophore. This limitation also excludes the possibility of identifying small conformational changes to the chromophore upon photo-excitation.

More sophisticated methodologies have been applied to the fluorescence spectrum of the Green Fluorescent Protein chromophore. Toniolo *et al*. [156] used a combined semi-empirical quantum mechanical/molecular mechanical (QM/MM) simulation to explore the effects of solvation on the fluorescence lifetime of the chromophore. Toniolo *et al*. [156] found that solvation reduced the fluorescence lifetime of the chromophore by more than an order of magnitude, explaining the reduced quantum yield of GFP chromophore analogs in solution [155]. These simulations suggested that rotation of the chromophore phenol ring relative to the imidazolinone ring was responsible for non-radiative de-excitation of the chromophore.

Sinicropi *et al*. [210] used a combined QM/MM simulation to calculate the absorption and emission peaks of the wtGFP chromophore. Sinicropi *et al*. [210] used the CHARMM (Chemistry at HARvard Molecular Mechanics) molecular dynamics package to calculate molecular motion, and an *ab initio* CASSCF/CASPT2 (Complete Active Space Self-Consistent Field/Complete Active Space with Second-order

Perturbation Theory) molecular wavefunction calculation. These calculations [210] achieved notable accuracy in estimating the positions of the GFP chromophore absorption and emission peaks under a variety of chromophore cavity conditions.

The high-pressure structures of Citrine and associated spectral information should provide an important test of these more advanced quantum chemical models of fluorescence.

6.3 Future Experiments on Fluorescent Proteins

The experiments on the behavior of Citrine under high pressure discussed in this thesis suggest several immediate avenues for experimentation with other members of the *Aequorea* fluorescent protein family.

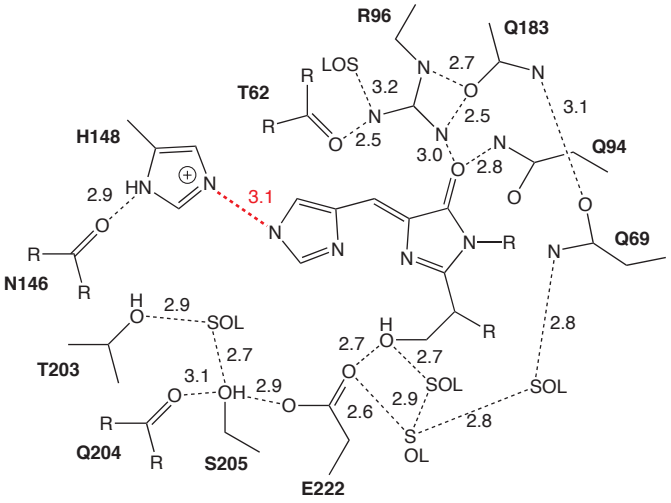
6.3.1 The Blue Fluorescent Protein

A first choice of an experimental system after Citrine might be the effects of pressure on the function and structure of the *Aequorea* Blue Fluorescent Protein (BFP). As discussed in chapter 5, Mauring *et al.* [196] demonstrated that the fluorescence intensity of the Blue Fluorescent Protein increases linearly with pressures up to 570 MPa. Of all of the *Aequorea* fluorescent proteins tested under pressure, only the Yellow Fluorescent Protein types display a decrease in fluorescence intensity with rising pressure at room temperature [53, 196].

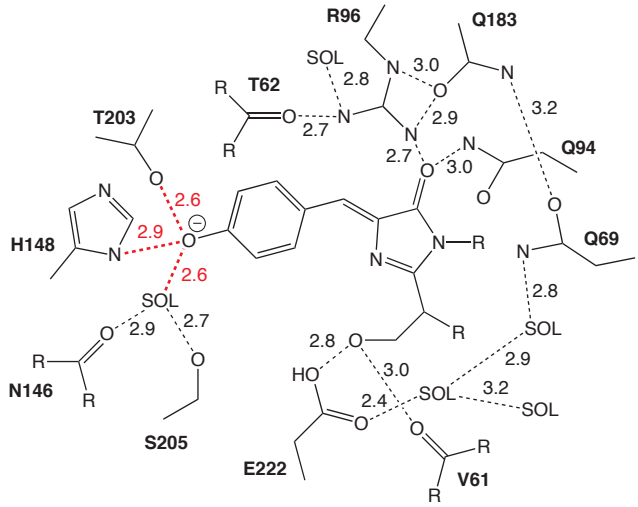
Mauring *et al*. [196] speculated that the increase in fluorescence intensity of BFP under pressure was due to a decrease in the distance between His148, proximate to the chromophore on the β-barrel wall, and the BFP chromophore, increasing the stabilization of the excited state of the chromophore. This increased excited state

Figure 6.2: A: Hydrogen bonding network in Blue Fluorescent Protein chromophore cavity. Adapted from Wachter *et al*. [52]. B: Hydrogen bonding network in Enhanced Green Fluorescent Protein chromophore cavity. Adapted from adapted from Brejc *et al.* [159]. C: Hydrogen bonding network in wild-type Green Fluorescent Protein chromophore cavity. Adapted from Ormö *et al*. [160]. Arrows indicate the direction of proton transfer following photon absorption by wtGFP. Bonds stabilizing the chromophore are shown as thick, red lines.

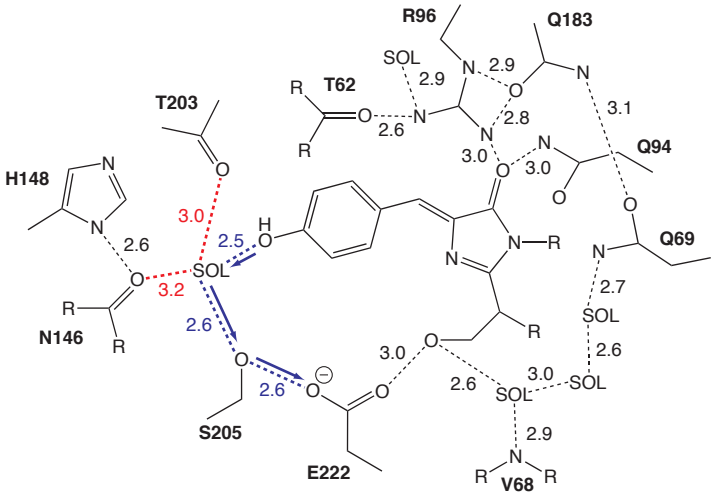
A: Blue Fluorescent Protein Hydrogen Bonding Network



B: Enhanced Green Fluorescent Protein Hydrogen Bonding Network



C: Wild-Type Green Fluorescent Protein Hydrogen Bonding Network



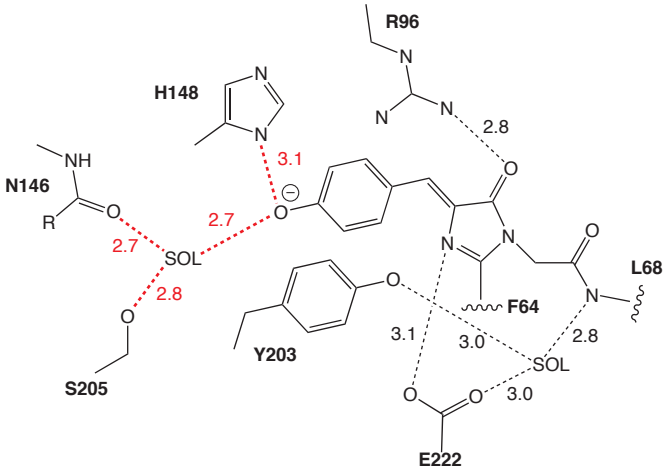


Figure 6.3: Room temperature and pressure Citrine chromophore cavity hydrogen bonding network. Network adapted from Wachter *et al.* [8]. Room temperature, room pressure hydrogen bonding distances from Citrine structure by Griesbeck *et al.* [7] (PDB accession code 1HUY). Bond lengths are measured in angstroms. Hydrogen bonds that stabilize the main chromophore are shown as thick, red lines.

stabilization reduces the probability of non-radiative decay and increases the quantum yield of the BFP chromophore. This speculation implies that the side of the β-barrel wall containing His148 moves closer to the chromophore and the central 3-10 helix rather than being pushed away from it, as is the case with Citrine, shown in figures 5.9 and 5.12.

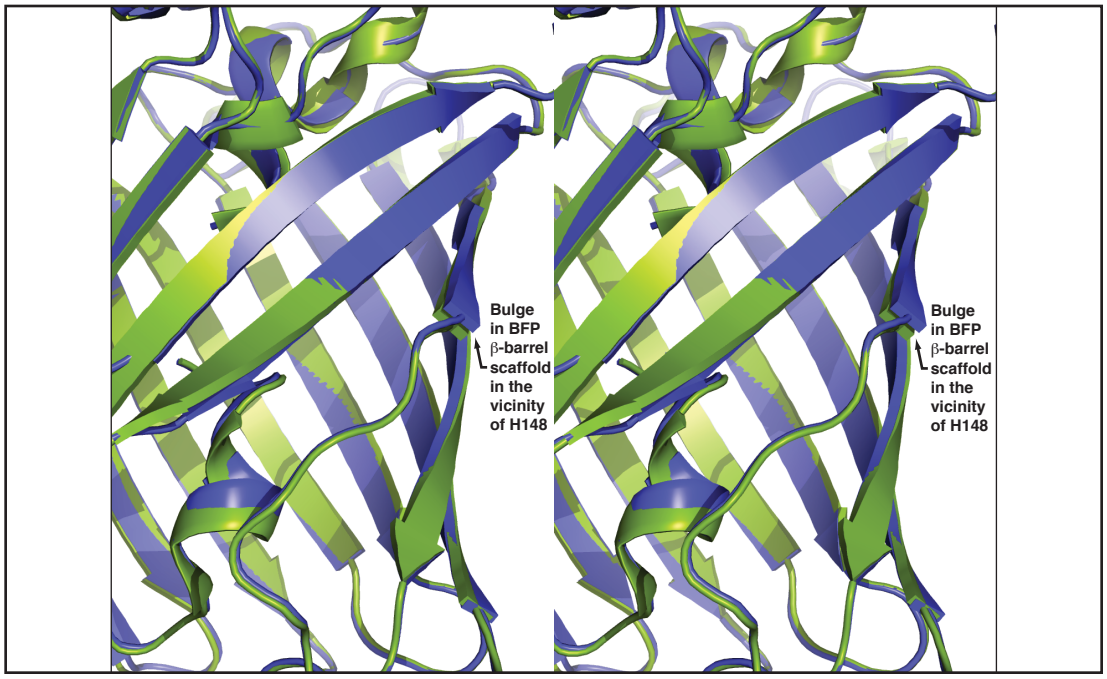


Figure 6.4: Ribbon diagram showing a bulge of β-barrel scaffold of BFP in the vicinity of His148 (PDB accession code 1BFP). Room temperature atomic coordinates by Wachter *et al*. [52]. For comparison, the scaffold of Citrine at room temperature by Griesbeck *et al*. [1] is shown alongside. The BFP model is shown in blue, and the Citrine structure is shown in yellow-green.

A diagram of the BFP chromophore cavity hydrogen bonding network, adapted from Wachter *et al*. [211] is shown in figure 6.2A. For comparison, the chromophore cavity hydrogen bonding networks of Enhanced Green Fluorescent Protein (EGFP); adapted from Brejc *et al.* [159] and of wild-type GFP (wtGFP); adapted from Ormö *et al.* [160], are shown in figures 6.2B and C respectively. The hydrogen bonding network present in the chromophore cavity of Citrine is re-presented for comparison in figure 6.3. In the case of EGFP, wtGFP and Citrine the hydrogen bonding network makes multiple stabilizing interactions with the main chromophore. However, in the case of BFP, the hydrogen bonding network makes only a single, weakly stabilizing

interaction, originating from His148, with the chromophore. In the case of Citrine, the links in this hydrogen bonding network, especially the bond between the main chromophore and His148, expand with pressure, reducing the stabilization of the main chromophore, and causing a reduction in the fluorescence intensity of Citrine, both at room temperature and under high-pressure cryocooling.

However, all other *Aequorea* fluorescent proteins studied under pressure show an increase in fluorescence intensity with increasing pressure [53, 196]. These observed increases in fluorescence intensity suggest that the links in these hydrogen bonding networks compress, increasing the stabilization of the main chromophore. This implies that the response of the β-barrel scaffold of the Yellow Fluorescent Protein mutants differs from that of the other types of *Aequorea* fluorescent protein mutants. This suggestion is supported by slight differences in the β-barrel scaffold of the YFP mutants and other types of *Aequorea* fluorescent protein.

Firstly, comparison of the room temperature β-barrel scaffolds of BFP [211] and of Citrine [1] indicates that the section of β-barrel scaffold containing His148 slightly protrudes into solution in the BFP structure. This salient bulge of the BFP β-barrel scaffold is shown in figure 6.4. It is reasonable to expect that this bulge is compressible. Compression of this bulge may be the only structural deformation required in order to actuate the compression stabilizing hydrogen bond between His148 and the BFP chromophore proposed by Mauring *et al*. [196].

However, this bulge around His148 is not present in EGFP, and so does not explain the observation that the fluorescence intensity of EGFP increases under pressure, while that of Citrine falls. It is possible that the bending deformation motion under pressure

seen in the high-pressure atomic structures of Citrine does not occur in EGFP, nor in other *Aequorea* fluorescent protein mutants.

All *Aequorea* fluorescent proteins display a floppy β-strand composed of residues 143 to 148. These residues do not form β-sheet interactions with the rest of the β-barrel scaffold. At room temperature, these residues appear to bulge slightly more into solution in the Citrine structure when compared with the corresponding residues in the EGFP structure. Additionally, at room temperature, the main chromophore of YFP protrudes slightly further, by ≈ 0.9 Å, into the chromophore cavity than the chromophore in the EGFP structure [52]. These two pre-existing deformations, possibly induced by the presence of the perturbing tyrosine 203 residue, may slightly weaken the YFP structure. These two "weaknesses" may predispose the YFP scaffold to the bending behavior seen under high-pressure cryocooling conditions that was discussed in chapter 5. This bending behavior may be analogous to macroscopic structures failing at the weakest point.

The possibly differing responses of different *Aequorea* fluorescent protein mutants suggest several experiments that may shed light on the influence of small mutations on the flexibility of the protein scaffold.

Firstly, it would be interesting to measure the fluorescence intensity of BFP under high-pressure cryocooling conditions. Any changes in the fluorescence intensity of BFP could be further investigated by solving the structure of BFP under high-pressure cryocooling. Changes in the fluorescence intensity of BFP could be correlated with changes in the hydrogen bonding network present in the BFP chromophore cavity. Finally, analysis of the deformation of the BFP scaffold may reveal the structural

mechanism for deformations in the BFP chromophore cavity hydrogen bonding network.

Additionally, it would be interesting to solve the structure of EGFP under high-pressure cryocooling conditions. These structures could be analyzed to understand changes in the chromophore cavity hydrogen bonding network, and perhaps most interestingly, changes in the overall scaffold of the molecule. While it is interesting to speculate about the nature of the high-pressure scaffold deformation of other members of the *Aequorea* fluorescent protein family, the true behavior is almost certainly extremely rich.

6.3.2 Wild-type Green Fluorescent Protein

Although the structure of Citrine and the response of its structure to high pressure are no doubt subtly different to that of wild-type Green Fluorescent Protein (wtGFP), changes in the hydrogen bonding network of Citrine, discussed in chapter 5, may provide a framework for explaining the subtle shift in the position of the fluorescence peak of wtGFP observed under high pressure observed by Oger *et al.* [204]. A diagram of the wtGFP chromophore cavity hydrogen bonding network is shown in figure 6.2C.

Wild-type Green Fluorescent Protein (wtGFP) is unusual amongst the *Aequorea* Fluorescent Proteins, in that excited state proton transfer (ESPT) plays an important role in its fluorescence mechanism. Under neutral solution conditions, the wtGFP chromophore can be found in either a neutral form, with a protonated chromophore, or an anionic chromophore, with a de-protonated chromophore. These two forms of the chromophore correspond to two bands present in the absorption spectrum of wtGFP: 395-397 nm for the protonated form, and 470-475 nm for the de-protonated form [2].

Following absorption of light by the neutral chromophore, the proton is transferred from the phenol of the chromophore to a solvent molecule close to the chromophore, to the γ-oxygen of Ser205, and then to the ε-oxygen of Glu222. The mutation of the chromophore residue Ser65 in wtGFP to Thr65 in EGFP lowers the pK_a of the chromophore, favoring the anionic form of the chromophore and disrupts wtGFP hydrogen bonding network, changing it to the network shown in figure 6.2B. The mutation S65T also shifts the fluorescence peak of the molecule from 504 nm to 511 nm.

Oger *et al*. [204] noted that when expressed in *E. coli*, under high pressures up 600 MPa, the fluorescence peak of wild-type GFP shifts by ≈ 0.5 nm per 100 MPa of applied pressure. Leiderman *et al*. [205] studied the effects of pressure up to 1100 MPa on the fluorescence peak and proton transfer rates in wtGFP. Leiderman *et al*. [205] concluded that the rate of proton transfer from the chromophore phenol to the solvent molecule in the chromophore cavity (figure 6.2C) was insensitive to pressure. However, Leiderman *et al*. [205] noted that recombination of the proton and excited state of the chromophore was increased by almost two orders of magnitude by the application of pressures up to 1100 MPa.

It would be interesting to solve the structure of wtGFP under high pressure cryocooling conditions, and correlate changes in the chromophore cavity hydrogen bonding network with changes in the fluorescence peak of the molecule, and with the proton recombination rate modification reported by Leiderman *et al*. [205]. The red-shifting of the fluorescence peak of wtGFP under pressure, suggests that the hydrogen bonding network in the chromophore cavity of wtGFP may be becoming more like that of EGFP.

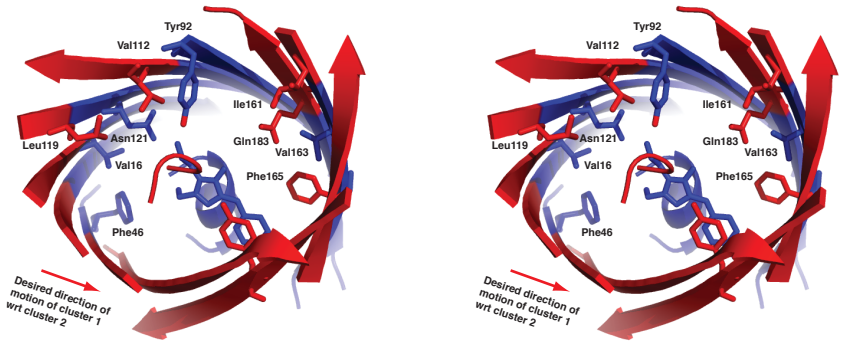


Figure 6.5: Speculated mutation sites to induce bending of Citrine scaffold under ambient conditions. We speculate that it may be possible to induce the bending of the Citrine scaffold seen in figure 5.9, and drawn as a cartoon in figure 5.12, by introducing small to large mutations on the left side of the molecule, and large to small mutations on the right side.

6.3.3 Modifying the Structure of Citrine

In chapter 5, it was noted that the pressure-induced bent state of the Citrine scaffold, shown as a cartoon in figure 5.12, may be structurally similar to the transiently observed, blue shifted state of the YFP (Yellow Fluorescent Protein) molecule seen in single molecule experiments at room temperature and pressure by Blum *et al.* [7].

We speculate that the structural information available on Citrine at high pressure may inspire structural modifications to Citrine that may permit the molecule to remain in the blue-shifted transient state for longer, or even permanently stabilize it, at room temperature and pressure. A cartoon representation of the bending deformation of Citrine under high-pressure was shown in figure 5.12. The bending of the scaffold is caused by relative motion and rotation of the two clusters of atoms that compose the Citrine scaffold that was shown in figure 5.9. It may be possible to induce this bending

under ambient conditions by repacking the Citrine scaffold by mutating or adding residues along the interface of these two clusters. We note that the identification of possible mutation sites in a cluster interface is not possible with conventional structural methods.

Figure 6.5 shows several possible sites where we speculate that mutations may be introduced into the Citrine scaffold. To replicate the bending seen at high pressure, it is required that cluster 1 move down and to the right in figure 6.5. We speculate that this may be achieved by simultaneously mutating residues on the left hand interface between clusters 1 and 2 to higher molecular weight residues, and mutating those on the right hand interface to lower molecular weight residues, while keeping other properties of the residue as similar as possible. On the left hand side, possible mutations may be Phe46 to Tyr or Trp, Val16 to Leu, Ile or Met, Asn121 to Gln, Val112 to Leu, Ile or Met, Tyr92 to Trp. On the right hand side, these mutations may be Ile1161 to Val, Ala or Gly, Gln183 to Asp, Cys, Thr or Ser, Val163 to Ala or Gly, while leaving Phe165 unchanged as it is the smalles aromatic amino acid residue. The success of this strategy would depend upon the perturbations of these mutations being small enough to not affect the folding of the β-barrel scaffold. If structural perturbations due to pressure of the character seen in Citrine are observed in catalytic proteins, it may be possible to use an approach similar to this to suggest mutations to alter the rate of catalysis.

6.4 Morphinone Reductase

6.4.1 Introduction

We sought to apply the techniques and lessons learned with Citrine to other proteins, specifically those that perform catalysis. Following the experiments with Citrine

detailed in chapters 2 to 5, we sought a second protein system that also displayed pressure sensitivity. For this second system, we desired a protein whose behavior, unlike Citrine, was well modeled under high pressure, but for which experimental confirmation of the proposed structural deformations under high pressure was lacking. Additionally, we sought to extend the high-pressure cryo-cooling analysis to proteins that catalyzed chemical reactions.

A protein molecule that met these requirements was Morphinone Reductase. The effects of pressure on the NADH oxidation rate of Morphinone Reductase, and a structural mechanism for this rate constant enhancement were recently reported by Hay *et al.* [18] and were summarized in chapter 1. We attempted to solve the structure of Morphinone Reductase, complexed with the non-oxidizable substrate analog $NADH_4$, under high-pressure cryocooling conditions to confirm the model by Hay *et al*. [18].

6.4.2 Crystallizing Morphinone Reductase

We received purified, frozen wild-type Morphinone Reductase and several vials of frozen, lyophilized $NADH_4$ from Christopher Pudney, a member of Professor Nigel Scrutton's laboratory (University of Manchester, United Kingdom). To ensure that the package passed through U.S. customs with minimum delay, a letter on headed paper, clearly explaining the contents of the package, and attesting to its non-hazardous nature was clearly affixed to the exterior of the shipping box.

We attempted to crystallize this Morphinone Reductase following the crystallization conditions reported by Pudney *et al*. [38]: 50% saturated ammonium sulfate, 100 mM HEPES, pH 7.0 using sitting drop vapor diffusion. Pudney *et al*. [38] reported that

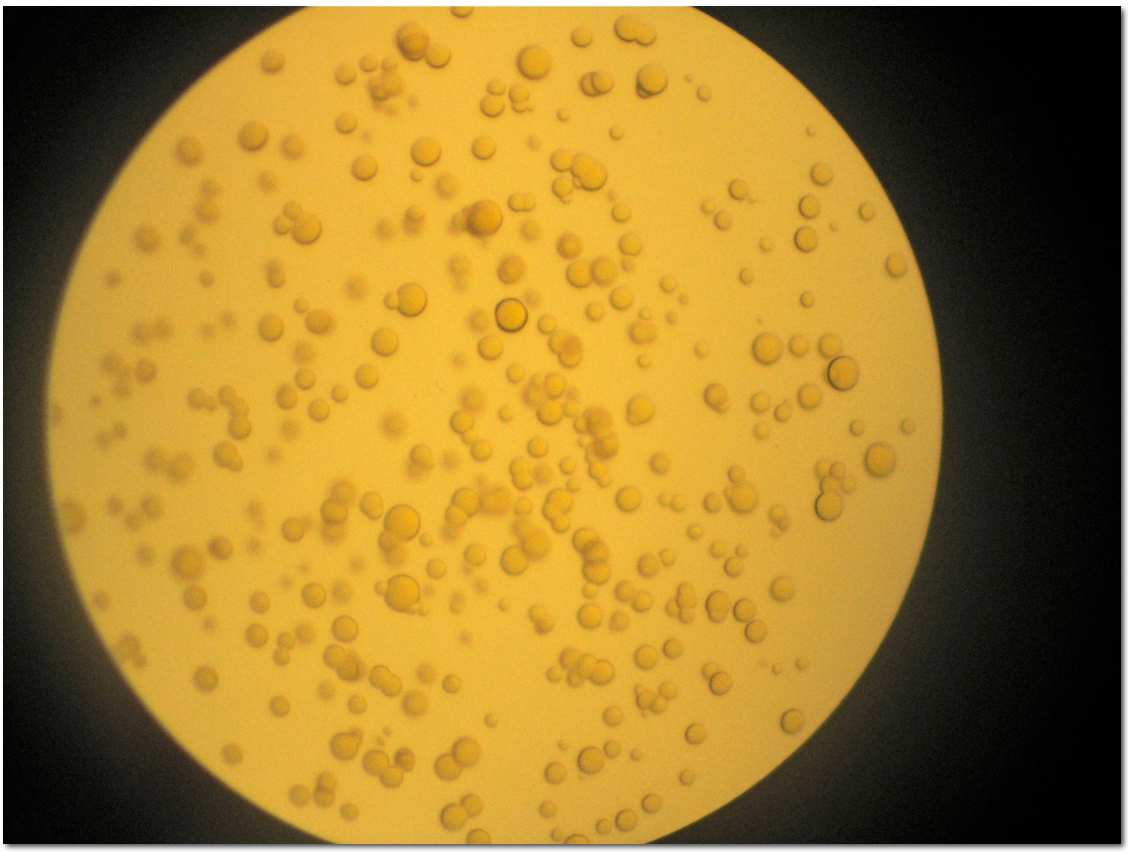


Figure 6.6: Spherulite Morphinone Reductase crystals grown in 45% saturated ammonium sulfate, 100 mM HEPES, pH 7.0.

crystals formed by this process diffracted to a maximum resolution of approximately 1.3 Å (PDB accession code 2R14).

To make the ammonium sulfate solution, a saturated ammonium sulfate solution was made by adding between 155 and 180 grams of ammonium sulfate powder (Catalog number 09978-1KG, Fluka BioUltra Ammonium Sulfate, Sigma-Aldrich, St. Louis, MO, USA) to 250 mL of deionized water. The mixture was boiled with a Bunsen flame, left to cool, and stored in a sterile container. This solution was used as a concentrate to make crystallization solutions. For example, for a 50% saturated

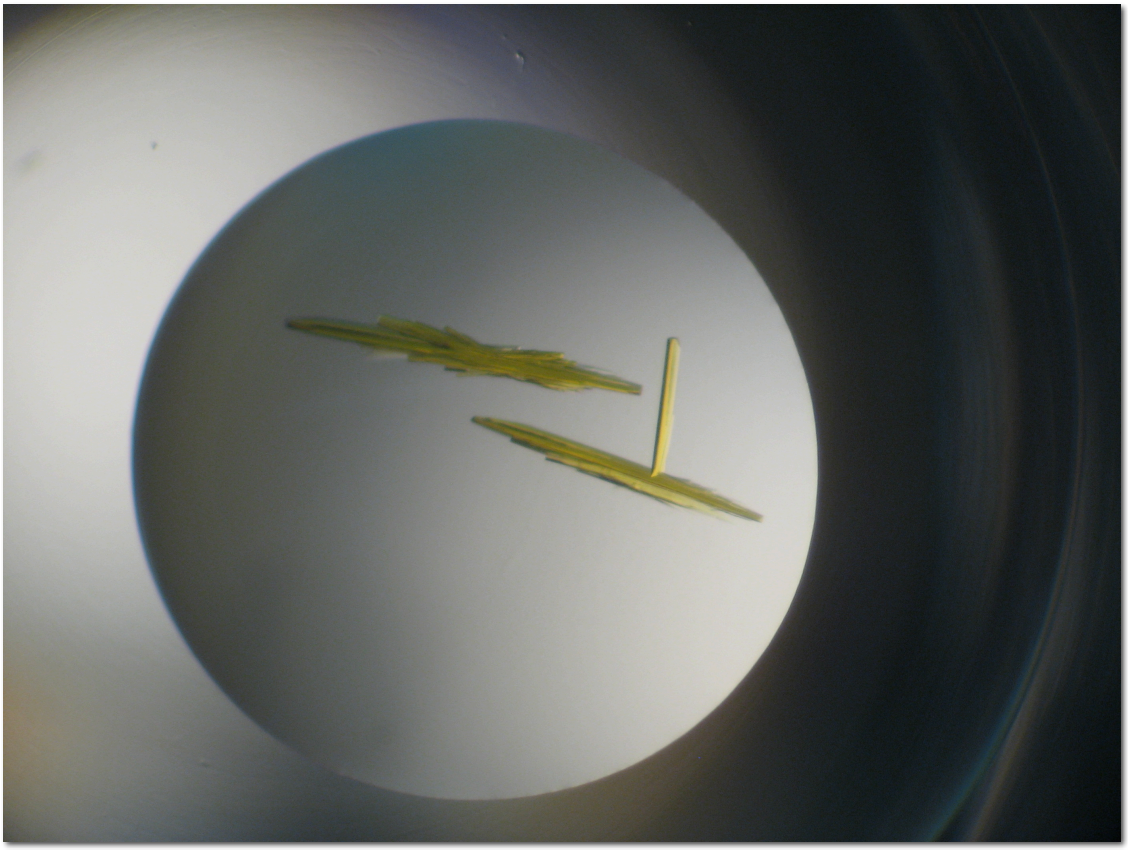


Figure 6.7: Brick-like crystals of Morphinone Reductase grown in 30% PEG 400, 100 mM HEPES, 200 mM $MgCl_2$, pH 7.0.

ammonium sulfate solution, this concentrate composed half of the volume of the final solution. Crystallization solutions were filtered with a 0.22 µm cellulose acetate filter.

Unfortunately, screening around the conditions reported by Pudney *et al*. [38] did not yield diffraction quality protein crystals. Crystallization experiments consistently yielded spherulites, thought to be droplets formed by fine needle-like crystals that radiate from a single point. Simultaneous crystallization attempts using another aliquot of the same purification batch of Morphinone Reductase by Dr David Leys at the University of Manchester found similar results. A photograph of these spherulites is shown in figure 6.6.

We then attempted to crystallize Morphinone Reductase following the conditions reported by Barna *et al*. [97]: 15-25% v/v PEG 550; 100 mM HEPES; 100 mM NaCl; 1 mM dithiothreitol; pH 6.5-7.6, and Moody *et al*. [96]: 28% v/v PEG 400; 100 mM HEPES; 200 mM $MgCl_2$; pH 7.5 and 30% v/v PEG 400 3; 100 mM HEPES; 200 mM $MgCl_2$; pH 7.5.

As recommended by Moody *et al*. [96], the Morphinone Reductase solution was dialyzed into deionized water and concentrated to 10 mg/mL. We found that crystals of adequate appearance grew at 30-35% PEG 400, 100 mM HEPES, 200 mM $MgCl_2$, pH 7.0-7.5. Moody *et al*. [96] reported diffraction from crystals grown under these conditions that extended only to 2.45 Å. Barna *et al*. [97] reported that crystals grown in PEG 550 diffracted to a maximum resolution of only 2.14 Å. A photograph of crystals grown in PEG 400 is shown in figure 6.7.

As discussed in chapter 1, Hay *et al*. [18] anticipated a pressure-induced reduction in the $NADH_4$ to FMN distance in Morphinone Reductase of ≈ 0.7 Å by 200 MPa. The Morphinone Reductase dataset reported by Barna *et al*. [97] contained 23,548 unique reflections with a completeness of 85.1%. The model of Morphinone Reductase by Barna *et al*. [97] contains 2917 atoms, and has an *R*-factor of 20.5%. Cruickshank's coordinate precision estimation formula (equation 2.30) indicates a coordinate uncertainty of

$$\begin{aligned}\sigma\left(x,B_{\text{avg}}\right) &= 1.0\left(N_i/(n-p)\right)^{1/2} C^{-1/3} R d_{\min} \\ \sigma\left(x,B_{\text{avg}}\right) &= 1.0\left(2917/\left(23548-2917\times 4\right)\right)^{1/2} \times 0.851^{-1/3} \times 0.205 \times 2.14\,\text{Å} \\ \sigma\left(x,B_{\text{avg}}\right) &= 0.229\,\text{Å}\end{aligned} \quad . \quad (6.6)$$

This suggests that the uncertainty on the largely unrestrained distance between the $NADH_4$ and FMN will be

$$\begin{aligned} \Sigma(\Delta x) &= \sqrt{3}\left(2 \times 0.229^2\right)^{1/2} \text{Å} \\ \Sigma(\Delta x) &\approx 0.6 \text{Å} \end{aligned}, \tag{6.7}$$

indicating that it may be difficult to definitively identify the deformation motion speculated to exist by Hay *et al*. [18] using diffraction data that extends to only ≈ 2.14 Å.

For crystals grown in ammonium sulfate, the estimated standard coordinate uncertainty is only 0.05 Å, giving an uncertainty on an unrestrained distance of only 0.12 Å. For this reason, we pursued attempts to crystallize Morphinone Reductase under the conditions reported by Pudney *et al*. [38].

We used micro-seeding from the Morphinone Reductase crystals grown in PEG 400 in an attempt to grow crystals under the ammonium sulfate conditions reported by Pudney *et al*. [38]. These attempts were unsuccessful.

As we were unable to grow apparently diffraction quality crystals of Morphinone Reductase in ammonium sulfate, we took crystals grown in PEG 400 and then soaked them in the ammonium sulfate well solution (45% saturated ammonium sulfate, 100 mM HEPES, pH 7.0). We, and David Leys working independently, found this soak was necessary in order for the $NADH_4$ substrate analog to bind to the Morphinone Reductase. Binding can be confirmed by visual inspection; a purified Morphinone Reductase solution is slightly yellow-green in color. However, when bound to $NADH_4$, the protein solution turns to a tobacco-brown color.

Before freezing under high pressure or at room pressure, the crystals were soaked in a mixture of well solution (45% saturated ammonium sulfate, 100 mM HEPES, pH 7.0) containing 20% v/v glycerol and approximately 10 mM to 100 mM $NADH_4$. To make the ≈ 100 mM $NADH_4$ solution, we took vials containing ≈ 10 mg of $NADH_4$ (molecular weight of 667.4) and dissolved the contents in 150 μL of 45% saturated ammonium sulfate well solution. This solution was used to soak Morphinone Reductase crystals for several minutes prior to high-pressure cryocooling.

The high-pressure cryocooled crystals of Morphinone Reductase were taken to CHESS (Cornell High Energy Synchrotron Source) station F2, and X-ray diffraction data was collected from them. Unfortunately, the maximum resolution of these X-ray diffraction datasets was much lower than that used to derive the original Morphinone Reductase-$NADH_4$ structure by Pudney *et al*. [38]. However, these datasets were refined by a refinement procedure similar to refinement procedure 2 used to refine the high-pressure Citrine atomic models (chapter 4), and several atomic models were derived. Quality indicators, and the FMN N5 to $NADH_4$ NC4 atom distance, are shown for each structure in table 6.1.

Table 6.1: Quality indicators for high-pressure Morphinone Reductase structures. *n* is the number of reflections used to derive each dataset, *C* is the overall completeness of the dataset. ESU B is the estimated standard uncertainty on the B-factor of the structure. ESU R is the estimated standard coordinate uncertainty based upon the *R*-factor of the structure (equation 2.30). ESU R_{Free} is the ESU based upon R_{Free} (equation 2.31). *d* is the FMN N5 to $NADH_4$ C4 atom distance. *d* ESU is the estimated standard uncertainty on *d*.

Structure	*P* (MPa)	*R*	R_{Free}	*n*	Hi Res (Å)	Low Res (Å)	*C*	Mean B (Å^2)	Bond Lengths RMSD (Å)	Angles RMSD (°)	ESU B (Å^2)	Atoms	ESU R_{Free} (Å)	ESU *R* (Å)	r (Å)	r ESU (Å)
2R14	0.1	0.157	0.1806	94812	1.4	12	100	16.72	0.017	1.837	2.996	3569	0.05	0.053	3.452	0.122
MR_NADH4_AS_0500_1	50	0.211	0.2653	17199	2.45	41.27	100	31.76	0.021	1.982	9.73	2958	0.274	0.386	3.402	0.671
MR_NADH4_AS_1000_1	100	0.183	0.224	18792	2.48	48.17	100	30.3	0.023	2.133	7.146	3062	0.228	0.311	3.405	0.558
MR_NADH4_AS_1000_3	100	0.194	0.2629	16598	2.07	38.35	100	24.67	0.019	1.827	5.408	3027	0.237	0.329	3.206	0.581
MR_NADH4_AS_2100_3	210	0.155	0.198	27610	2.18	39.53	100	28.66	0.029	2.254	3.7	3290	0.151	0.162	3.333	0.37

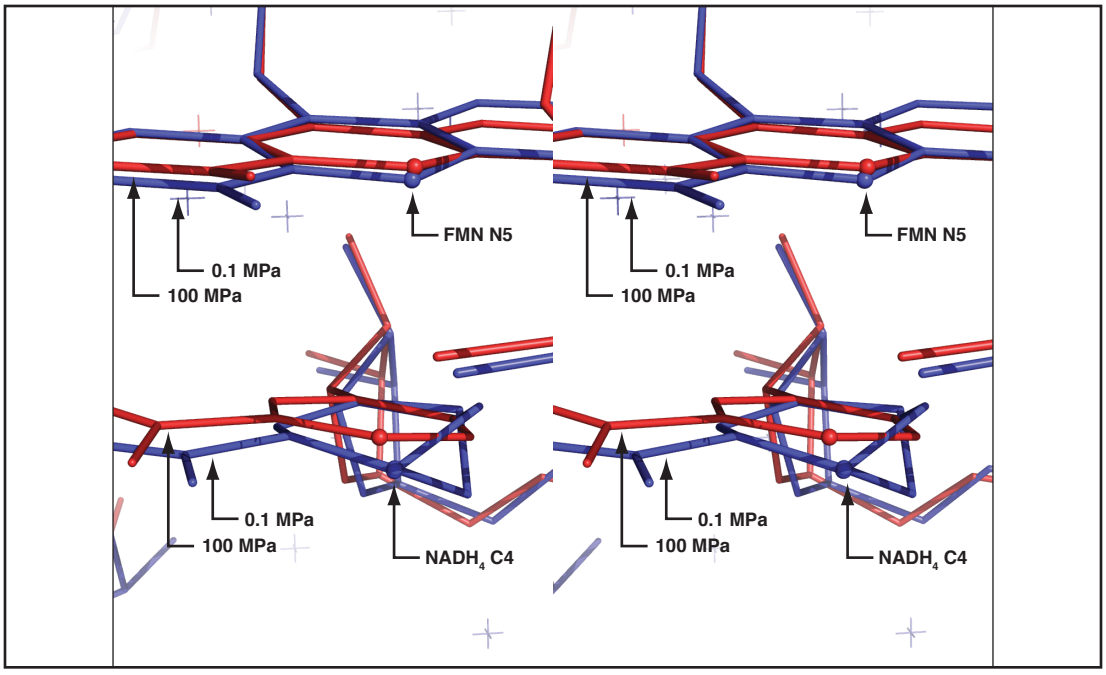


Figure 6.8: Stereo view of possible deformation motion observed at active site of Morphinone Reductase. The flash-frozen ambient pressure structure of Morphinone Reductase by Pudney *et al.* is shown in blue (PDB accession code 2R14). A high-pressure cryocooled structure of Morphinone Reductase at 100 MPa is shown in red (MR_NADH4_-AS_1000_3). Note that the distance between the FMN N5 and $NADH_4$ NC4 atom shrinks slightly with pressure. Note that in the ambient pressure structure of Morphinone Reductase, two conformations of the $NADH_4$ NC5 atom can be resolved, while only one conformation, can be resolved in the high-pressure structure.

The results presented in table 6.1 show a small reduction in the distance *d*, between atoms with increasing pressure. However, this reduction in distance does not appear to be simply correlated with pressure. Additionally, we estimate the errors on this distance to be large in comparison to the apparent reduction in distance. To clarify, the distances shown in table 6.1, that between the FMN N5 to $NADH_4$ NC4 atoms, is not

the tunneling distance, *r*, discussed by Hay *et al.* [18] and in chapter 1. The tunneling distance, *r*, is the distance between proton sites attached to the FMN N5 to $NADH_4$ NC4 atoms, rather than the distance, *d*, between the FMN N5 to $NADH_4$ NC4 atom centers. However, it seems reasonable to assume that the proton-nitrogen and proton-carbon bond lengths should not be affected by pressure; thus, changes in *d*, should be very closely equal to changes in *r*.

It is interesting to note that for one of these structures at 100 MPa, there is a very notable reduction in the distance between the FMN N5 to $NADH_4$ C4 atoms, reminiscent of the reduction in this distance proposed by Hay *et al*. [18]. This deformation motion is shown in figure 6.8. However, this deformation motion is not as clearly present in the other high-pressure atomic models of Morphinone Reductase. It is also notable that the reduction in the atom center distance, *d*, is far smaller than that proposed by Hay *et al*. [18], of approximately 0.7 Å over 200 MPa.

However, these small contractions are consistent with recent molecular dynamics (MD) simulation results by Hay *et al.* [212] that suggest that the reduction in the FMN N5 to $NADH_4$ C4 distance compresses by only ≈ 0.2 Å as the pressure applied to the MR-NADH4 complex is increased from 0.1 to 200 MPa. Additionally, the errors on the FMN N5 to $NADH_4$ C4 distance are comparable to the spread in this distance implied by MD simulations [212].

These results, while intriguing, should be taken with some skepticism, due to the limited precision available in the high-pressure structure of Morphinone Reductase. Firstly, it may be that the compressibility of the Morphinone Reductase active site is lower under high-pressure cryocooling conditions than at room temperature, the

condition for which the model by Hay *et al*. [18] is proposed. Secondly, the structure of the Morphinone Reductase-$NADH_4$ complex may respond differently than the Morphinone-Reductase-NADH complex under high pressure.

To resolve these issues, and shed more light on the behavior of Morphinone Reductase several issues should be addressed. Firstly, the error on the FMN N5 to $NADH_4$ C4 atom distance will need to be reduced considerably. This will require the production of higher quality crystals of Morphinone Reductase. Unfortunately, at the present time, the crystals used to derive the high resolution Morphinone Reductase model appear to be irreproducible, by myself, and by the originator of these crystals; Dr David Leys (University of Manchester, United Kingdom). We further purified, by gel filtration chromatography, and then re-concentrated our aliquot of Morphinone Reductase. This suggests that an unknown contaminant was present in the purification batch of Morphinone Reductase used to produce the original batch of crystals that diffracted to 1.3 Å. To promote the growth of the diffraction quality crystals from future purification batches of Morphinone Reductase, I would recommend the use of an additive screen (Hampton Research of Aliso Viejo, CA, USA sells an additive screen under catalog number HR2-428), to mimic the effects of the possible unknown contaminant.

The second issue that should be addressed is the equivalence of the activity of Morphinone Reductase under high-pressure cryocooling conditions and under room temperature, high-pressure conditions. It is very important to remember that the behavior of Citrine under high-pressure cryocooling conditions and under room temperature, high pressure conditions is very different: at room temperature, the fluorescence peak of Citrine shifts to the red with increasing pressure, while under

high-pressure cryocooling the fluorescence peak shifts to the blue. It is certainly likely that the structural response of Morphinone Reductase under high-pressure cryocooling is different from its response under high-pressure at room temperature. Unfortunately, it is presently not possible to repeat the high-pressure stopped flow measurements that Hay *et al*. [18] performed under high-pressure cryocooling conditions. However, Professor Nigel Scrutton (University of Manchester, United Kingdom) proposed that we measure the absorbance of the π-π charge-transfer complex formed between $NADH_4$ and Morphinone Reductase [37, 212]. The absorbance of this complex is sensitive to the FMN N5 to $NADH_4$ C4 distance, and could be used to validate the results of high-pressure crystallography on Morphinone Reductase and compare them with room temperature results. The micro-spectrophotometer that we constructed for use with Citrine, and that discussed in chapter 4, is capable of performing these absorbance measurements on high-pressure cryocooled Morphinone Reductase samples.

Additionally, should larger crystals of Morphinone Reductase become available, it may be possible to use the high-pressure beryllium X-ray cell to determine the structure of MR under high pressure at room temperature. This will allow closer correlations with the work by Hay *et al*. [18, 212].

6.5 Cellulase Activity Under High Pressure

6.5.1 Introduction

Cellulases are a class of enzymes that degrade the biological carbohydrate polymer cellulose. Cellulose is a highly abundant, natural source of glucose [213]. Cellulose may be hydrolyzed to glucose, and this glucose may be fermented to ethanol for use as a transportation fuel [214, 215]. It is thought that the derivation of glucose from cellulosic material is environmentally sustainable, as the crops necessary to provide this material may be grown on low quality land with minimal artificial fertilizer input [214, 216]. However, present techniques to degrade cellulosic material are energy intensive, posing a barrier to the widespread production of cellulosic ethanol [216]. For this reason, attempts to understand the mechanism of cellulases, and suggest means to enhance their activity have received considerable attention [214, 216].

We sought to investigate if the activity of cellulases could be modified by the application of pressure. The combination of information on a pressure-modified rate constant and high-pressure structural information might shed further light on the mechanism of action of cellulases. Additionally, high-pressure structures of cellulase may provide suggestions for mutations to induce acceleration of the activity of these enzymes at room temperature and pressure.

6.5.2 Assay for Cellulase Activity Under High Pressure

We sought a convenient high-pressure assay for cellulase. We tested two cellulase substrate-analogs: Resorufin-β-D-Cellobioside [217] (Catalog number M1245, Marker Gene Technologies Inc., Eugene, OR, USA) and EnzChek (Catalog number EE33953, Molecular Probes, Eugene, OR, USA). Both of these substrate analogs release a fluorescent dye into solution upon digestion by cellulase. The rate of increase of this

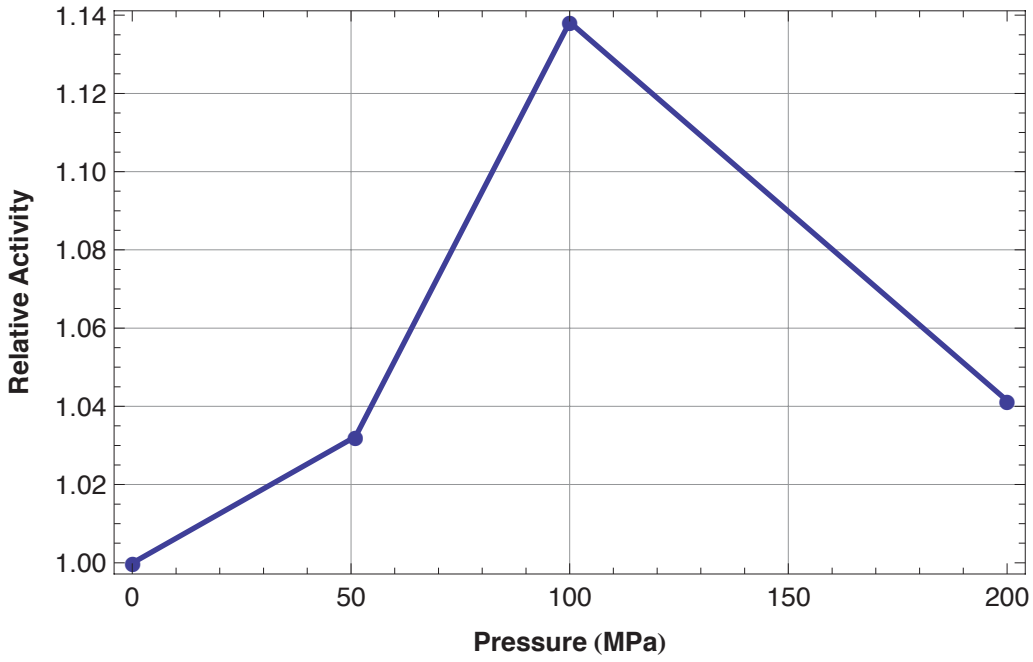


Figure 6.9: Apparent rate constant variation of *Thermobifida fusca* Cel9A cellulase under hydrostatic pressure at room temperature. Error bars on this plot are difficult to quantify at the time of writing. The plot is shown to catalyze future work on cellulase.

fluorescence signal can be taken as a proxy for the catalytic rate of the cellulase. Both of these substrate-analogs degrade even in the absence of cellulase. We found that EnzChek showed an unacceptably high degradation rate in the absence of cellulase, and was also prohibitively expensive. Thus, we decided to use Resorufin-β-D-Cellobioside in a high-pressure assay for cellulase activity.

The Resorufin-β-D-Cellobioside substrate-analog contains cellobioside, a dimer of two D-glucose molecules, complexed with the fluorescent dye resorufin. When resorufin is complexed with cellobioside, it is quenched, and largely non-fluorescent. When cellulase hydrolyzes the Resorufin-β-D-Cellobioside, the resorufin is released into solution and is no longer quenched. Thus, the fluorescence signal from the

solution increases. The rate of increase of the fluorescence signal can be interpreted as the rate of action of the cellulase enzyme.

We attempted to measure the rate constant of two cellulases under high pressure: tomato Endo-β-1,4-glucanase, SlCel9C1 cellulase [213] and the highly active *Thermobifida fusca* Cel9A cellulase [215]. We found that the rate of degradation of Resorufin-β-D-Cellobioside by the tomato cellulase was indistinguishable from the background degradation rate. Thus, we focused our efforts on the *T. fusca* Cel9A cellulase [215].

6.5.3 Results

We measured the fluorescence signal from a solution of 0.01 μM *T. fusca* Cel9A cellulase and 0.025 mM Resorufin-β-D-Cellobioside (Molecular weight 537.47) in 100 mM sodium acetate, pH 5.0, under high pressure in a commercial high-pressure fluorescence cell (ISS, Urbana-Champaign, IL, USA) using a Chronos spectrophotometer (ISS). Fluorescence excitation was at 571 nm, and emission was monitored at 585 nm. The temperature of the high-pressure cell was maintained at 25 ˚C with a thermostatic re-circulating water bath (Neslab, Thermo Scientific, Waltham, MA, USA).

We found a small rate constant increase of cellulase, of ≈ 14% under a pressure of 100 MPa of the *Thermobifida fusca* Cel9A cellulase [215]. A plot of the apparent rate constant of the *T. fusca* cellulase versus pressure at room temperature is shown in figure 6.9. This plot shows an apparent increase in the activity of cellulase up to a pressure of 100 MPa, and then a decrease as the pressure rises to 200 MPa.

6.5.4 Discussion

It should be emphasized that results on cellulase presented here should be treated with skepticism until many more careful control experiments have been performed. Firstly, the apparent rollover in the activity of the *T. fusca* cellulase at pressures exceeding 100 MPa may be the result of an increasing of the optical density of the reaction buffer due to the release of resorufin. What appears to be a rollover in catalytic activity, may in fact be a large increase in activity. When the reaction vessel was placed in the high-pressure cell, the reaction buffer had a light, translucent orange color, due to the presence of free resorufin. However, after the experiment was completed, and the reaction vessel was removed from the high-pressure cell, the appearance of the reaction buffer was more like that of red wine.

Secondly, at room temperature, resorufin cellobioside has a very low, but still detectable breakdown rate. At room temperature and pressure, the fluorescence signal of the r Resorufin-β-D-Cellobioside will show a slight increase even in the absence of any carbohydrate active enzymes. This puts a limit on the detectability of very low reaction rates by this method. The breakdown rate of Resorufin-β-D-Cellobioside at elevated pressures has not been quantified, and should be prior to drawing any definitive conclusions about the behavior of cellulase under high pressure.

Most importantly, Resorufin-β-D-Cellobioside is a highly non-realistic test substrate for the cellulase. Cellobioside contains only two glucose molecules. Breakdown of this molecule by cellulase, is an unrealistic analog of the real-world substrate that cellulase will be required to digest. In general, the activity of cellulases against soluble substrates is not a good proxy for the activity against more realistic, insoluble

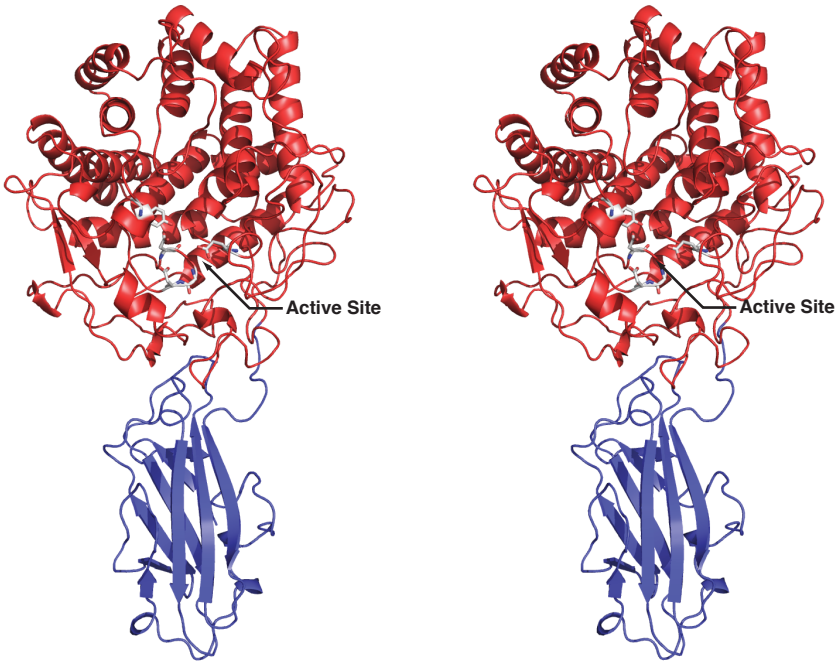


Figure 6.10: Ribbon diagram of the of *T. fusca* Cel9A cellulase [215]. Atomic model by Sakon *et al*. [218] (PDB accession code 1TF4). The catalytic domain, residues 1-454, is colored in red and the binding domain, residues 455-605, is colored in blue. The active site residues are shown as sticks. A close up of the active site can be seen in figure 6.11.

substrates [214]. For this reason, the measurement of the activity of cellulase under high pressure against more realistic substrates should be measured.

However, if these more realistic assays do show an increase, even a modest one, in the activity of cellulase under pressure, it is worth considering a high-pressure crystallographic experiment on the *T. fusca* cellulase. The structure of the *T. fusca* cellulase is known to 1.9 Å resolution (PDB accession code 1TF4) [218]. This model was derived from an X-ray diffraction dataset of 120,169 unique reflections with a of 94%. The model contains 11,077 non-hydrogen atoms and has a crystallographic *R*-factor of 20%. This suggests a coordinate uncertainty, from Cruickshank's formula

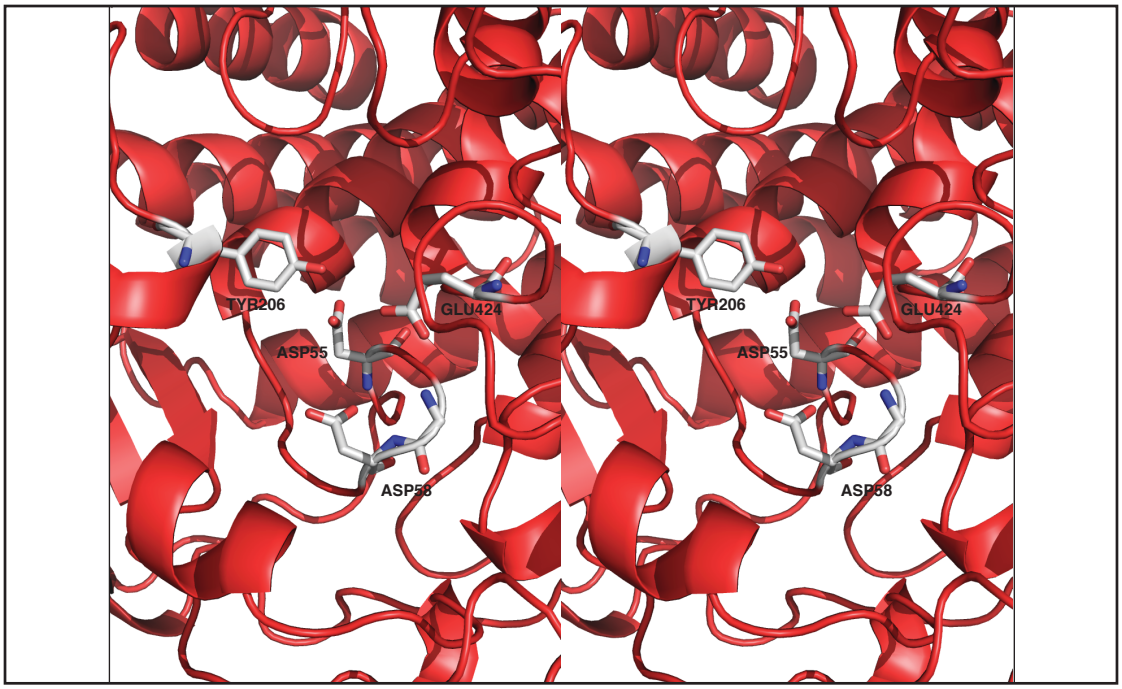


Figure 6.11: Stereo view of the proposed catalytic site of *T. fusca* Cel9A cellulase [218]. Atomic model by Sakon *et al.* [218] (PDB accession code 1TF4).

(equation 2.30) of approximately 0.15 Å. A stereo diagram of the *T. fusca* Cel9A cellulase is shown in figure 6.10. A close up of the proposed active site of the Cel9A cellulase is shown in figure 6.11.

Preliminary results of single molecule experiments on cellulases suggest that, like several other enzymes studied by single molecule techniques [7, 54, 55, 57, 58], individual cellulase molecules display several different rate constants [216]. If the deformation motion of cellulase under pressure is reminiscent of that of Citrine: deformation into state resembling a transiently occupied state at room temperature and pressure; and this state is highly active; it may be possible to insert mutations into the cellulase scaffold that induce the protein to adopt a more highly active state at room temperature and pressure.

6.6 Speculations on Hydrogenases

The ability to catalyze reactions at room temperature and pressure is an extremely attractive feature of protein molecules. Many of the industrial catalysts in use today require high temperatures and high pressures, requiring high energy inputs.

Molecular hydrogen is a highly attractive transportation fuel that can be oxidized in a fuel cell to form only water as an exhaust. Hydrogen oxidation produces no carbon dioxide or additional exhaust emissions such as the ozone precursors acetaldehyde and formaldehyde [219], nitrogen oxide, unburned hydrocarbons and nitrogen dioxide [220] that are produced by ethanol and ethanol-gasoline burning engines. However, present methods of producing H_2 are less attractive: the light hydrocarbon steam reforming reaction and coal or heavy hydrocarbon gasification require high temperatures, usually well in excess of 1000 ˚C to proceed [221]. The electrolysis of water is in principle more desirable. However, electrolysis relies upon expensive electrodes to act as catalysts, limiting its wide application especially in developing nations [222]. Mineable resources of hydrogen are scarce [223].

The need for onsite, on demand generation is particularly great for H_2, as climate modeling indicates that leakage may contribute to climatic disruption [224].

Photosynthetic production of hydrogen using engineered microorganisms is thought to be a desirable novel hydrogen generation technology [225]. A key challenge in achieving this goal is the design and engineering of a hydrogenase enzyme that catalyzes the reduction of protons to H_2 using electrons supplied by a photosystem that can operate in an aerobic industrial environment. Photosynthetic H_2 production was demonstrated by Ihara *et al.* [226] using a cyanobacterial photosystem linked to a

hydrogenase with a nickel-iron (NiFe) active site [226]. For industrial synthesis of H_2, hydrogenases with an iron-iron (FeFe) active site are much more appropriate: they are almost 1000 times more active [227] and require much less complex assembly than NiFe-hydrogenases [92]. However, the greatest obstacle to the use of FeFe-hydrogenases is their intolerance to aerobic environments.

It is thought that O_2 migrates through gas channels in the hydrogenase structure that also transport H_2 [228-231] and irreversibly oxidizes the active site [232]. FeFe-hydrogenases have a hydrogen generation rate of approximately 10^4 H_2 molecules per second [233] and are inactivated by a single O_2 molecule. This implies that in order to have a half-life of 24 hours, the hydrogenase must permit only one O_2 molecule to enter its active site for approximately every 10^9 H_2 molecules generated. For comparison, a K^+ ion channel membrane protein permits the passage of 1 Na^+ ion for every 10^4 K^+ ions that cross the channel [234]. The difference in average radius between O_2 and H_2 is only 0.3 Å, implying that the dimensions of the gas channels of a hydrogenase may need to be modified on the sub-angstrom scale in order to achieve a high level of discrimination between H_2 and O_2 while retaining the high catalytic rate of the enzyme. This subtle modulation of the gas channel dimensions may be possible by repacking the core of the hydrogenase.

If high-pressure structures of hydrogenases were to reveal a reduction in the gas volume, the structures could be analyzed in a fashion similar to that employed to understand the deformation of Citrine under pressure: a clustering analysis. If such a clustering analysis is successful (it is important to remember that this clustering analysis may not be appropriate for all proteins under pressure: it does not appear to work for T4 lysozyme) it may reveal points in the structure amenable to the

introduction of “wedge” residues that may subtly repack the core of the protein, and subtly reduce the volume of the gas channels at room temperature and pressure.

6.7 Conclusions

The experiments presented in this thesis provide an explicit demonstration of the continuous linkage of the scaffold of a protein molecule, its active site and the activity; in this case fluorescence, of the molecule.

These experiments link a slight bending, of approximately 2˚ over 500 MPa of applied pressure, of the β-barrel scaffold of the yellow fluorescent protein molecule Citrine to a deformation of approximately 0.8 Å in its active site. This bending motion under pressure is due to the relative motion of two clusters of atoms of differing compressibility that compose the scaffold of the Citrine molecule. This deformation of the β-barrel scaffold is communicated to the aromatic rings directly involved in fluorescence.

The bending motion of the β-barrel scaffold actuates the separation of the two elements of the Citrine chromophore: the main chromophore and the perturbing tyrosine 203 ring. The main chromophore and tyrosine 203 phenol slide apart by ≈ 0.8 Å under a pressure of 500 MPa. This sub-angstrom separation of the main chromophore and the tyrosine 203 ring removes the perturbing influence of the tyrosine 203 phenol on the main chromophore, and allows the fluorescence spectrum of Citrine to return to the green under high pressure. Extended Hückel modeling of the shift in energy of the fluorescence peak of Citrine indicates that the mechanism of perturbation of the main chromophore by the tyrosine 203 phenol is orbital overlap leading to destabilization of the highest molecular orbital (HOMO) of the main

chromophore, and narrowing in the energy of the HOMO-LUMO transition from which fluorescence photons are emitted.

Dietz *et al*. [151-153] have performed single molecule unfolding experiments on Green Fluorescent Protein molecules using optical tweezers. We speculate that a similar experiment could be performed where mechanical forces, such as those produced by the tip of an atomic force microscope (AFM) or optical tweezers, could be used distort the protein to test if it results in a similar spectral shift to that observed under high pressure. While this would be a very difficult experiment to perform, it is in principle feasible.

High-pressure studies of protein structure and function offer the possibility of probing the energy landscapes of protein active sites and offering new insights into the structural basis of enzymatic catalysis. Experiments of this kind may provide further insight into models of protein function, the coupling between protein active sites and the protein scaffold, giving additional insights into the structural basis of allosteric behavior in protein molecules. For example, the pressure dependent behavior of Citrine provides a test of models of protein fluorescence [156, 210, 235].

This type of simultaneous study of the activity and structure of protein molecules under progressive deformation will provide challenging tests for, and may allow the improvement of, many models of protein function. This chapter has suggested several prospects for future investigation: other members of *Aequorea* fluorescent protein family, Morphinone Reductase, cellulases and hydrogenases. An improvement in protein function prediction algorithms may have important benefits for rational protein design.

It is likely that small, less than 1 Å, structural deformations in the active sites of many proteins, may have measurable effects on their functions. It is also possible that seemingly small refinement errors, on the order 0.1 Å, in atomic structures of proteins derived from low resolution data or by homology modeling of known structures, could lead to substantial errors on the predicted catalytic activities of these proteins [236]. These results also suggest that to achieve maximum efficiency in their design goal, designed macromolecules may need to be engineered with sub-angstrom structural accuracy.

These experiments may also provide structural insights into the results of single molecule experiments, by allowing the trapping of normally transiently occupied protein states, as we believe was shown with Citrine. Finally, high pressure perturbation may provide a means to explore catalytic rate enhancement, and a way to achieve it, by suggesting sites at which strain-inducing mutations [237, 238] may be introduced into the protein scaffold, to mimic the effects of high pressure under ambient conditions, that would not be highlighted by single protein structures, random mutagenesis or directed evolution methods.

REFERENCES

1. Griesbeck, O.*, et al.*, *Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications.* J Biol Chem, 2001. **276**(31): p. 29188-94.

2. Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44.

3. Kim, C.U., R. Kapfer, and S.M. Gruner, *High-pressure cooling of protein crystals without cryoprotectants.* Acta Crystallogr D Biol Crystallogr, 2005. **61**(Pt 7): p. 881-90.

4. Urayama, P.K., G.N. Phillips, and S.M. Gruner, *Probing substates in sperm whale myoglobin using high-pressure crystallography.* Structure, 2002. **10**(1): p. 51-60.

5. Barstow, B.*, et al.*, *Alteration of citrine structure by hydrostatic pressure explains the accompanying spectral shift.* Proc Natl Acad Sci U S A, 2008. **105**(36): p. 13362-6.

6. Barstow, B.*, et al.*, *Coupling of Pressure-Induced Structural Shifts to Spectral Changes in the Yellow Fluorescent Protein.* In Submission, 2009.

7. Blum, C., A.J. Meixner, and V. Subramaniam, *Room Temperature Spectrally Resolved Single-Molecule Spectroscopy Reveals New Spectral Forms and Photophysical Versatility of Aequorea Green Fluorescent Protein Variants.* Biophysical Journal, 2004. **87**(6): p. 4172-4179.

8. Perutz, M.F.*, et al.*, *Stereochemistry of cooperative mechanisms in hemoglobin.* Accounts of Chemical Research, 1987. **52**: p. 555-65.

9. Daopin, S.*, et al.*, *Comparison of two crystal structures of TGF-beta2: the accuracy of refined protein structures.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 1): p. 85-92.

10. Suydam, I.T.*, et al.*, *Electric fields at the active site of an enzyme: direct comparison of experiment with theory.* Science, 2006. **313**(5784): p. 200-4.

11. Fersht, A. and G. Winter, *Protein engineering.* Trends Biochem Sci, 1992. **17**(8): p. 292-5.

12. Frauenfelder, H.*, et al.*, *Proteins and pressure.* J. Phys. Chem., 1990. **94**: p. 1024-1037.

13. Ueda, I.*, et al.*, *Does pressure antagonize anesthesia? High-pressure stopped-flow study of firefly luciferase and anatomy of initial flash.* Biophys J, 1999. **76**(1 Pt 1): p. 478-82.

14. Bruner and Hall, *Pressure effects on alamethicin conductance in bilayer membranes.* Biophys J, 1983. **44**(1): p. 39-47.

15. Unno, M.*, et al.*, *High-pressure flash photolysis study of hemoprotein: effects of substrate analogues on the recombination of carbon monoxide to cytochrome P450CAM.* Biochem., 1994. **33**(32): p. 9762-8.

16. Adachi, S. and I. Morishima, *The Effects Of Pressure On Oxygen And Carbon-Monoxide Binding-Kinetics For Myoglobin - A High-Pressure Laser Flash-*

Photolysis Study. Journal of Biological Chemistry, 1989. **264**(32): p. 18896-18901.

17. Miller, J.F.*, et al.*, *High-pressure-temperature bioreactor for studying pressure-temperature relationships in bacterial growth and productivity.* Biotechnol Bioeng, 1988. **31**(5): p. 407-13.

18. Hay, S., M.J. Sutcliffe, and N.S. Scrutton, *Promoting motions in enzyme catalysis probed by pressure studies of kinetic isotope effects.* Proc Natl Acad Sci USA, 2007. **104**(2): p. 507-12.

19. Wishart, J.F.*, et al.*, *High-Pressure Pulse-Radiolysis Study Of Intramolecular And Intermolecular Reduction Of Cytochrome-C By Ruthenium(II) Ammine Complexes.* Inorganic Chemistry, 1992. **31**(19): p. 3986-3989.

20. Meier, M.*, et al.*, *Pressure Effects On The Rates Of Intramolecular Electron-Transfer In Ruthenium-Modified Cytochrome-C - Role Of The Intervening Medium In Tuning Distant Fe2+Ru3+ Electronic Couplings.* Journal of the American Chemical Society, 1994. **116**(4): p. 1577-1578.

21. Bansch, B.*, et al.*, *Mechanistic Information From The First Volume Profile Analysis For A Reversible Intermolecular Electron-Transfer Reaction Involving Pentaammine(Isonicotinamide)Ruthenium And Cytochrome-C.* Inorganic Chemistry, 1994. **33**(21): p. 4744-4749.

22. Scott, J.R.*, et al.*, *Intramolecular electron-transfer reactions of cytochrome b(5) covalently bonded to ruthenium(II) polypyridine complexes: Reorganizational energy and pressure effects.* Inorganica Chimica Acta, 1996. **243**(1-2): p. 193-200.

23. Kundrot, C. and F. Richards, *Collection and Processing of X--Ray Diffraction Data from Protein Crystals at High Pressure.* Journal of Applied Crystallography, 1986. **19**: p. 208-213.

24. Kundrot, C. and F. Richards, *Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres.* J Mol Biol, 1987. **193**(1): p. 157-70.

25. Fourme, R.*, et al.*, *High-pressure protein crystallography (HPPX): instrumentation, methodology and results on lysozyme crystals.* Journal of synchrotron radiation, 2001. **8**(Pt 5): p. 1149-56.

26. Refaee, M.*, et al.*, *Pressure-dependent changes in the solution structure of hen egg-white lysozyme.* J Mol Biol, 2003. **327**(4): p. 857-65.

27. Williamson, M.P., K. Akasaka, and M. Refaee, *The solution structure of bovine pancreatic trypsin inhibitor at high pressure.* Protein Sci, 2003. **12**(9): p. 1971-9.

28. Collins, M.D.*, et al.*, *Cooperative water filling of a nonpolar protein cavity observed by high-pressure crystallography and simulation.* Proc Natl Acad Sci USA, 2005. **102**(46): p. 16668-71.

29. Collins, M.D.*, et al.*, *Structural rigidity of a large cavity-containing protein revealed by high-pressure crystallography.* J Mol Biol, 2007. **367**(3): p. 752-63.

30. Wilton, D.J.*, et al.*, *Pressure-induced changes in the solution structure of the GB1 domain of protein G.* Proteins, 2008. **71**(3): p. 1432-40.

31. Miyashita, O. and N. Go, *Pressure dependence of protein electron transfer reactions: Theory and simulation.* Journal of Physical Chemistry B, 1999. **103**(3): p. 562-571.

32. Prabhakaran, M.*, et al.*, *Cytochrome c: the effect of temperature and pressure from molecular dynamics simulations.* Journal of Physics and Chemistry of Solids, 2004. **65**(8-9): p. 1615-1622.

33. Galkin, O.*, et al.*, *Pressure effects on the proximal heme pocket in myoglobin probed by Raman and near-infrared absorption spectroscopy.* Biophys J, 1997. **73**(5): p. 2752-63.

34. French, C.E. and N.C. Bruce, *Bacterial morphinone reductase is related to Old Yellow Enzyme.* Biochem J, 1995. **312 (Pt 3)**: p. 671-8.

35. Bruce, N.C.*, et al.*, *Microbial degradation of the morphine alkaloids: identification of morphine as an intermediate in the metabolism of morphine by Pseudomonas putida M10.* Arch Microbiol, 1990. **154**(5): p. 465-70.

36. Bruce, N.C.*, et al.*, *Microbial degradation of the morphine alkaloids. Purification and characterization of morphine dehydrogenase from Pseudomonas putida M10.* Biochem J, 1991. **274 (Pt 3)**: p. 875-80.

37. Craig, D.H.*, et al.*, *Reductive and oxidative half-reactions of morphinone reductase from Pseudomonas putida M10: a kinetic and thermodynamic analysis.* Biochemistry, 1998. **37**(20): p. 7598-607.

38. Pudney, C.R.*, et al.*, *Mutagenesis of morphinone reductase induces multiple reactive configurations and identifies potential ambiguity in kinetic analysis of enzyme tunneling mechanisms.* J Am Chem Soc, 2007. **129**(45): p. 13949-56.

39. Isaacs, N.S., K. Javaid, and E. Rannala, *Reactions At High-Pressure .5. Effect Of Pressure On Some Primary Kinetic Isotope Effects.* Journal of the Chemical Society-Perkin Transactions 2, 1978(8): p. 709-711.

40. Isaacs, N.S., *Isotopic effects : recent developments in theory and experiment*, in *Isotopes in organic chemistry v. 6*, E. Buncel and C.C. Lee, Editors. 1984, Elsevier: Amsterdam ; New York. p. 67-105.

41. Schlichting, I.*, et al.*, *Crystal structure of photolysed carbonmonoxy-myoglobin.* Nature, 1994. **371**(6500): p. 808-12.

42. Kuriyan, J.*, et al.*, *X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 A resolution.* J Mol Biol, 1986. **192**(1): p. 133-54.

43. Phillips, S.E., *Structure and refinement of oxymyoglobin at 1.6 A resolution.* J Mol Biol, 1980. **142**(4): p. 531-54.

44. Kitagawa, T., *Biological applications of Raman spectroscopy*, T.G. Spiro, Editor. 1987, Wiley: New York. p. 97-132.

45. Rousseau, D.L. and J.M. Friedman, *Biological applications of Raman spectroscopy*, T.G. Spiro, Editor. 1987, Wiley: New York. p. 97-132.

46. Lim, M., T.A. Jackson, and P.A. Anfinrud, *Nonexponential protein relaxation: dynamics of conformational change in myoglobin.* Proc Natl Acad Sci U S A, 1993. **90**(12): p. 5801-4.

47. Kuczera, K.*, et al.*, *Nonexponential relaxation after ligand dissociation from myoglobin: a molecular dynamics simulation.* Proc Natl Acad Sci U S A, 1993. **90**(12): p. 5805-7.

48. Sage, J.T. and P.M. Champion, *Small substrate recognition in heme proteins*, in *Supramolecular Chemistry.*, K.S. Suslick, Editor. 1997, Elsevier: Amsterdam.

49. Stavrov, S.S., *The effect of iron displacement out of the porphyrin plane on the resonance Raman spectra of heme proteins and iron porphyrins.* Biophys J, 1993. **65**(5): p. 1942-50.

50. Ansari, A.*, et al.*, *Rebinding and relaxation in the myoglobin pocket.* Biophys Chem, 1987. **26**(2-3): p. 337-55.

51. Frauenfelder, H.*, et al.*, *The role of structure, energy landscape, dynamics, and allostery in the enzymatic function of myoglobin.* Proc Natl Acad Sci U S A, 2001. **98**(5): p. 2370-4.

52. Wachter, R.M.*, et al.*, *Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein.* Structure, 1998. **6**(10): p. 1267-77.

53. Verkhusha, V.V.*, et al.*, *Effect of high pressure and reversed micelles on the fluorescent proteins.* Biochim Biophys Acta, 2003. **1622**(3): p. 192-5.

54. Lu, H.P., L.Y. Xun, and X.S. Xie, *Single-molecule enzymatic dynamics.* Science, 1998. **282**(5395): p. 1877-1882.

55. Zhuang, X.W.*, et al.*, *Correlating structural dynamics and function in single ribozyme molecules.* Science, 2002. **296**(5572): p. 1473-1476.

56. van Oijen, A.M.*, et al.*, *Single-molecule kinetics of lambda exonuclease reveal base dependence and dynamic disorder.* Science, 2003. **301**(5637): p. 1235-1238.

57. Velonia, K.*, et al.*, *Single-enzyme kinetics of CALB-catalyzed hydrolysis.* Angewandte Chemie-International Edition, 2005. **44**(4): p. 560-564.

58. Flomenbom, O.*, et al.*, *Stretched exponential decay and correlations in the catalytic activity of fluctuating single lipase molecules.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(7): p. 2368-2372.

59. English, B.*, et al.*, *Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited.* Nat Chem Biol, 2006. **2**(2): p. 87-94.

60. Yang, H.*, et al.*, *Protein conformational dynamics probed by single-molecule electron transfer.* Science, 2003. **302**(5643): p. 262-266.

61. Coureux, P.D.*, et al.*, *Picometer-scale conformational heterogeneity separates functional from nonfunctional states of a photoreceptor protein.* Structure, 2008. **16**(6): p. 863-72.

62. Vanbeeumen, J.J.*, et al.*, *Primary Structure Of A Photoactive Yellow Protein From The Phototrophic Bacterium Ectothiorhodospira-Halophila, With Evidence For The Mass And The Binding-Site Of The Chromophore.* Protein Science, 1993. **2**(7): p. 1114-1125.

63. Baca, M.*, et al.*, *Complete Chemical-Structure Of Photoactive Yellow Protein - Novel Thioester-Linked 4-Hydroxycinnamyl Chromophore And Photocycle Chemist.* Biochemistry, 1994. **33**(48): p. 14369-14377.

64. Getzoff, E.D., K.N. Gutwin, and U.K. Genick, *Anticipatory active-site motions and chromophore distortion prime photoreceptor PYP for light activation.* Nature Structural Biology, 2003. **10**(8): p. 663-668.

65. Kiefer, F., *et al.*, *The SWISS-MODEL Repository and associated resources.* Nucleic Acids Res, 2009. **37**: p. D387-D392.

66. Arnold, K., *et al.*, *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.

67. Tramontano, A. and V. Morea, *Assessment of homology-based predictions in CASP5.* Proteins, 2003. **53 Suppl 6**: p. 352-68.

68. Komoto, J., *et al.*, *Effects of site-directed mutagenesis on structure and function of recombinant rat liver S-adenosylhomocysteine hydrolase. Crystal structure of D244E mutant enzyme.* J Biol Chem, 2000. **275**(41): p. 32147-56.

69. Kundrot, C. and F. Richards, *Effect of hydrostatic pressure on the solvent in crystals of hen egg-white lysozyme.* J Mol Biol, 1988. **200**(2): p. 401-10.

70. Urayama, P.K., *Techniques for High Pressure Macromolecular Crystallography and the Effects of Pressure on the Structure of Sperm Whale Myoglobin.* 2001.

71. Fourme, R., *et al.*, *High-pressure macromolecular crystallography (HPMX): status and prospects.* Biochim Biophys Acta, 2006. **1764**(3): p. 384-90.

72. Fourme, R., *et al.*, *Opening the high-pressure domain beyond 2 kbar to protein and virus crystallography--technical advance.* Structure, 2002. **10**(10): p. 1409-14.

73. Girard, E., *et al.*, *The first crystal structure of a macromolecular assembly under high pressure: CpMV at 330 MPa.* Biophys J, 2005. **88**(5): p. 3562-71.

74. Fourme, R.*, et al.*, *Using a quasi-parallel X-ray beam of ultrashort wavelength for high-pressure virus crystallography: implications for standard macromolecular crystallography.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 10): p. 1767-72.

75. Girard, E.*, et al.*, *Adaptation of the base-paired double-helix molecular architecture to extreme pressure.* Nucleic Acids Res, 2007. **35**(14): p. 4800-8.

76. Perutz, M.F., *Science is not a quiet life : unravelling the atomic mechanism of haemoglobin*. 1997, London: Imperial College Press. xxi, 636 p.

77. Perutz, M.F., *I wish I'd made you angry earlier : essays on science, scientists, and humanity*. Expanded ed. 2003, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. xv, 486 p.

78. Kendrew, J.C.*, et al.*, *A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.* Nature, 1958. **181**(4610): p. 662-6.

79. Perutz, M.F.*, et al.*, *Structure of Haemoglobin: A Three-Dimensional Fourier Synethesis At 5.5 Å Resolution Obtained by X-ray Analysis.* Nature, 1960. **185**: p. 416-422.

80. Kusse, B. and E. Westwig, *Mathematical physics : applied mathematics for scientists and engineers*. 1998, New York: Wiley. xi, 668 p.

81. Perutz, M.F., *I Wish I'd Made You Angry Earlier*. 2003.

82. Als-Nielsen, J. and D. McMorrow, *Elements of modern X-ray physics*. 2001, New York: Wiley. xi, 318 p.

83. Uson, I. and G.M. Sheldrick, *Advances in direct methods for protein crystallography.* Curr Opin Struct Biol, 1999. **9**(5): p. 643-8.

84. Hauptman, H., *Phasing methods for protein crystallography.* Curr Opin Struct Biol, 1997. **7**(5): p. 672-80.

85. Vagin, A.A. and A. Teplyakov, *MOLREP: an automated program for molecular replacement.* J. Appl. Crystallography, 1997. **30**: p. 1022-1025.

86. Jelsch, C.*, et al.*, *Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin.* Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3171-6.

87. Connon, S.A. and S.J. Giovannoni, *High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates.* Appl Environ Microbiol, 2002. **68**(8): p. 3878-85.

88. Bartlett, D.H., *Pressure effects on in vivo microbial processes.* Biochim Biophys Acta, 2002. **1595**(1-2): p. 367-81.

89. Rothschild, L.J. and R.L. Mancinelli, *Life in extreme environments.* Nature, 2001. **409**(6823): p. 1092-101.

90. Krimsky, S., *Genetic alchemy : the social history of the recombinant DNA controversy*. 1982, Cambridge, Mass.: MIT Press. xiii, 445 p.

91. King, P.W.*, et al.*, *Functional studies of [FeFe] hydrogenase maturation in an Escherichia coli biosynthetic system.* J Bacteriol, 2006. **188**(6): p. 2163-72.

92. Lenz, O.*, et al.*, *Requirements for heterologous production of a complex metalloenzyme: the membrane-bound [NiFe] hydrogenase.* J Bacteriol, 2005. **187**(18): p. 6590-5.

93. Tian, J.*, et al.*, *Accurate multiplex gene synthesis from programmable DNA microchips.* Nature, 2004. **432**(7020): p. 1050-1054.

94. Kong, D.S.*, et al.*, *Parallel gene synthesis in a microfluidic device.* Nucleic Acids Res, 2007.

95. Bergfors, T.M., *Protein crystallization : techniques, strategies, and tips : a laboratory manual*. IUL biotechnology series 1. 1999, La Jolla, Calif.: International University Line. xix, 306 p.

96. Moody, P.C.*, et al.*, *Crystallization and preliminary diffraction studies of morphinone reductase, a flavoprotein involved in the degradation of morphine alkaloids.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 5): p. 619-21.

97. Barna, T.*, et al.*, *Crystal structure of bacterial morphinone reductase and properties of the C191A mutant enzyme.* J Biol Chem, 2002. **277**(34): p. 30976-83.

98. Steller, I., R. Bolotovsky, and M.G. Rossmann, *An algorithm for automatic indexing of oscillation images using Fourier analysis.* J. Appl. Crystallography, 1997. **30**: p. 1036-1040.

99. Rossmann, M.G. and C.G.v.B. van Beek, *Data processing.* Acta Crystallogr D Biol Crystallogr, 1999. **55 (Pt 10)**: p. 1631-40.

100. Leslie, A., *Recent Changes to the MOSFLM Package for Processing Film and Image Plate Data.* 1992.

101. Collaborative Computational Project, N., *The CCP4 suite: programs for protein crystallography.* Acta Crystallogr D Biol Crystallogr, 1994. **50**(Pt 5): p. 760-3.

102. Kabsch, W., *Evaluation of Single-Crystal X-ray Diffraction Data from a Position-Sensitive Detector* Journal of Applied Crystallography, 1988. **21**: p. 916-924.

103. French, S. and K.S. Wilson, *On the Treatment of Negative Intensity Observations.* Acta Cryst., 1978. **A34**: p. 517-525.

104. Main, P. and M.G. Rossmann, *Relationships Among Structure Factors Due To Identical Molecules In Different Crystallographic Environments.* Acta Crystallographica, 1966. **21**: p. 67-&.

105. Navaza, J., *Amore - an automated package for molecular replacement.* Acta Crystallographica Section A, 1994. **50**: p. 157-163.

106. Navaza, J., *Implementation of molecular replacement in AMoRe.* Acta Crystallogr D Biol Crystallogr, 2001. **57**: p. 1367-1372.

107. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 3): p. 240-55.

108. Cohen, S.N.*, et al.*, *Towards complete validated models in the next generation of ARP/wARP.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2222-9.

109. Evrard, G.X.*, et al.*, *Assessment of automatic ligand building in ARP/wARP.* Acta Crystallogr D Biol Crystallogr, 2006. **63**(Pt 1): p. 108-17.

110. Morris, R.J., A. Perrakis, and V.S. Lamzin, *ARP/wARP's model-building algorithms. I. The main chain.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 6 Pt 2): p. 968-75.

111. Morris, R.J., A. Perrakis, and V.S. Lamzin, *ARP/wARP and automatic interpretation of protein electron density maps.* Meth. Enzymol., 2003. **374**: p. 229-44.

112. Perrakis, A.*, et al.*, *wARP: improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models.* Acta Crystallogr D Biol Crystallogr, 1997. **53**(Pt 4): p. 448-55.

113. Perrakis, A.*, et al.*, *ARP/wARP and molecular replacement.* Acta Crystallogr D Biol Crystallogr, 2001. **57**(Pt 10): p. 1445-50.

114. Laskowski, R.*, et al.*, *PROCHECK: a program to check the stereochemical quality of protein structures.* Journal of Applied Crystallography, 1993. **26**: p. 283-291.

115. Kim, C.U., Q. Hao, and S.M. Gruner, *Solution of protein crystallographic structures by high-pressure cryocooling and noble-gas phasing.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 7): p. 687-94.

116. Kim, C.U.*, et al.*, *Pressure Induced High-Density Amorphous Ice in Protein Crystals.* J. Appl. Crystallography, 2007. **41**: p. 1-7.

117. Kim, C.U., Q. Hao, and S.M. Gruner, *High-pressure cryocooling for capillary sample cryoprotection and diffraction phasing at long wavelengths.* Acta Crystallogr D Biol Crystallogr, 2007. **63**(Pt 5): p. 653-9.

118. Feynman, R.P., R.B. Leighton, and M.L. Sands, *The Feynman lectures on physics*. 1989, Redwood City, Calif.: Addison-Wesley.

119. Huang, B.*, et al.*, *Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy.* Science, 2008. **319**(5864): p. 810-3.

120. Hell, S.W., *Toward fluorescence nanoscopy.* Nat Biotechnol, 2003. **21**(11): p. 1347-55.

121. Hell, S.W., *Far-field optical nanoscopy.* Science, 2007. **316**(5828): p. 1153-8.

122. Bock, H.*, et al.*, *Two-color far-field fluorescence nanoscopy based on photoswitchable emitters.* Applied Physics B-Lasers and Optics, 2007. **88**(2): p. 161-165.

123. Gustafsson, M.G., *Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution.* Proc Natl Acad Sci U S A, 2005. **102**(37): p. 13081-6.

124. Rust, M.J., M. Bates, and X. Zhuang, *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)*. Nat Methods, 2006. **3**(10): p. 793-5.

125. Moerner, W.E., *Single-molecule mountains yield nanoscale cell images.* Nat Methods, 2006. **3**(10): p. 781-2.

126. Bates, M.*, et al.*, *Multicolor super-resolution imaging with photo-switchable fluorescent probes.* Science, 2007. **317**(5845): p. 1749-53.

127. Betzig, E.*, et al.*, *Imaging intracellular fluorescent proteins at nanometer resolution.* Science, 2006. **313**(5793): p. 1642-5.

128. Hess, S.T., T.P. Girirajan, and M.D. Mason, *Ultra-high resolution imaging by fluorescence photoactivation localization microscopy.* Biophys J, 2006. **91**(11): p. 4258-72.

129. Egner, A.*, et al.*, *Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters.* Biophys J, 2007. **93**(9): p. 3285-90.

130. Sharonov, A. and R.M. Hochstrasser, *Wide-field subdiffraction imaging by accumulated binding of diffusing probes.* Proc Natl Acad Sci U S A, 2006. **103**(50): p. 18911-6.

131. Prince, E. and P.T. Boggs, *Least Squares*, in *International Tables for Crystallography Volume C: Mathematical, Physical and Chemical Tables*, E. Prince, Editor. 2004, Kluwer Academic: Dodrecht/Boston/London. p. 678-688.

132. Messerschmidt, A., *X-ray crystallography of biomacromolecules : a practical guide*. 2007, Weinheim: Wiley-VCH. xiii, 304 p.

133. Cruickshank, D., *Remarks about protein structure precision.* Acta Crystallogr D Biol Crystallogr, 1999. **55 (Pt 3)**: p. 583-601.

134. Tickle, I.J., R.A. Laskowski, and D.S. Moss, *Error estimates of protein structure coordinates and deviations from standard geometry by full-matrix refinement of gammaB- and betaB2-crystallin.* Acta Crystallogr D Biol Crystallogr, 1998. **54**(Pt 2): p. 243-52.

135. Allen, F.H.*, et al.*, *Typical interatomic distances: organic compounds*, in *International Tables for Crystallography Volume C: Mathematical, Physical and Chemical Tables*, E. Prince, Editor. 2004, Kluwer Academic: Dodrecht/ Boston/London. p. 790-794.

136. Laskowski, R.A., D.S. Moss, and J.M. Thornton, *Main-chain bond lengths and bond angles in protein structures.* J Mol Biol, 1993. **231**(4): p. 1049-67.

137. Cruickshank, D.W.J., *The Accuracy Of Electron-Density Maps In X-Ray Analysis With Special Reference To Dibenzyl.* Acta Crystallographica, 1949. **2**(2): p. 65-82.

138. Brünger, A.T., *Free R value: a novel statistical quantity for assessing the accuracy of crystal structures.* Nature, 1992. **355**: p. 472-475.

139. Sparks, C.*, et al.*, *Submersible observations on the daytime vertical distribution of Aequorea ?forskalea off the west coast of southern Africa.* Journal of the Marine Biological Association of the United Kingdom, 2005. **85**: p. 519-522.

140. *Nature's Magic*, BBC Radio 4, 2004, 15 minutes, http://www.bbc.co.uk/radio4/ science/naturesmagic.shtml.

141. Heim, R., D.C. Prasher, and R.Y. Tsien, *Wavelength mutations and posttranslational autoxidation of green fluorescent protein.* Proc Natl Acad Sci USA, 1994. **91**(26): p. 12501-4.

142. Cubitt, A.B.*, et al.*, *Understanding, improving and using green fluorescent proteins.* Trends Biochem Sci, 1995. **20**(11): p. 448-55.

143. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein.* Biochemistry, 1997. **36**(22): p. 6786-91.

144. Prasher, D.C.*, et al.*, *Primary structure of the Aequorea victoria green-fluorescent protein.* Gene, 1992. **111**(2): p. 229-33.

145. Chalfie, M.*, et al.*, *Green fluorescent protein as a marker for gene expression.* Science, 1994. **263**(5148): p. 802-5.

146. Inouye, S. and F.I. Tsuji, *Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein.* FEBS Lett, 1994. **341**(2-3): p. 277-80.

147. Zolotukhin, S.*, et al.*, *A "humanized" green fluorescent protein cDNA adapted for high-level expression in mammalian cells.* J Virol, 1996. **70**(7): p. 4646-54.

148. Scheyhing, C.H.*, et al.*, *Temperature-pressure stability of green fluorescent protein: a Fourier transform infrared spectroscopy study.* Biopolymers, 2002. **65**(4): p. 244-53.

149. Ward, W.W.*, et al.*, *Spectral perturbations of the aequorea green-fluorescent protein.* Photochemistry Photobiology, 1982. **35**.

150. Ward, W.W. and S.H. Bokman, *Reversible denaturation of Aequorea green-fluorescent protein: physical separation and characterization of the renatured protein.* Biochem., 1982. **21**(19): p. 4535-40.

151. Dietz, H. and M. Rief, *Exploring the energy landscape of GFP by single-molecule mechanical experiments.* Proc Natl Acad Sci USA, 2004. **101**(46): p. 16192-7.

152. Dietz, H.*, et al.*, *Anisotropic deformation response of single protein molecules.* Proc Natl Acad Sci USA, 2006. **103**(34): p. 12724-8.

153. Dietz, H. and M. Rief, *Protein structure by mechanical triangulation.* Proc Natl Acad Sci USA, 2006. **103**(5): p. 1244-7.

154. Mickler, M.*, et al.*, *Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations.* Proc Natl Acad Sci USA, 2007. **104**(51): p. 20268-73.

155. Niwa, H.*, et al.*, *Chemical nature of the light emitter of the Aequorea green fluorescent protein.* Proc Natl Acad Sci U S A, 1996. **93**(24): p. 13617-22.

156. Toniolo, A.*, et al.*, *Conical intersection dynamics in solution: The chromophore of Green Fluorescent Protein.* Faraday Discuss, 2004. **127**: p. 149.

157. Chattoraj, M.*, et al.*, *Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer.* Proc Natl Acad Sci USA, 1996. **93**(16): p. 8362-7.

158. Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved green fluorescence.* Nature, 1995. **373**(6516): p. 663-4.

159. Brejc, K.*, et al.*, *Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein.* Proc Natl Acad Sci USA, 1997. **94**(6): p. 2306-11.

160. Ormö, M.*, et al.*, *Crystal structure of the Aequorea victoria green fluorescent protein.* Science, 1996. **273**(5280): p. 1392-5.

161. Yang, F., L. Moss, and G.N. Phillips, *The molecular structure of green fluorescent protein.* Nat Biotechnol, 1996. **14**(10): p. 1246-51.

162. Gross, M. and R. Jaenicke, *Proteins under pressure. The influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes.* Eur J Biochem, 1994. **221**(2): p. 617-30.

163. Neuman, R., W. Kauzmann, and A. Zipp, *Pressure Dependence of Weak Acid Ionization in Aqueous Buffers.* Journal of Physical Chemistry, 1973. **77**(22): p. 2687-2691.

164. Zipp, A. and W. Kauzmann, *Pressure Denaturation of Metmyoglobin.* Biochem., 1973. **12**(21): p. 4217-4228.

165. Lakowicz, J.R., *Principles of fluorescence spectroscopy*. 3rd ed. 2006, New York: Springer. xxvi, 954 p.

166. Mealli, C. and D.M. Proserpio, *MO Theory Made Visible.* Journal of Chemical Education, 1990. **67**: p. 399-402.

167. Cramer, C.J., *Essentials of computational chemistry : theories and models*. 2nd ed. 2004, Chichester, West Sussex, England ; Hoboken, NJ: Wiley. xx, 596 p.

168. Hoffmann, R., *An Extended Hückel Theory .I. Hydrocarbons*, in *J Chem Phys*. 1963. p. 1397-&.

169. Slater, J.C., *Atomic Shielding Constants.* Physical Review, 1930. **36**(1): p. 57.

170. Heikal, A.A.*, et al.*, *Molecular spectroscopy and dynamics of intrinsically fluorescent proteins: coral red (dsRed) and yellow (Citrine).* Proc. Natl. Acad. Sci. U.S.A., 2000. **97**(22): p. 11996-2001.

171. Landrum, E., Senior Thesis, *A Study Of The Effects Of Pressure On Protein Crystals*, 2008, Physics Department, Cornell University.

172. Davies, J. and D.I. Smith, *Plasmid-determined resistance to antimicrobial agents.* Annu Rev Microbiol, 1978. **32**: p. 469-518.

173. Shaw, K.J.*, et al.*, *Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes.* Microbiol Rev, 1993. **57**(1): p. 138-63.

174. Petry, *Design of Protective Structures*, 1950, Navy Docks P-51, Bureau of Yards and Docks, Dept. of Navy, Washington D.C.,

175. Timoshenko, S., *Strength of Materials*. Vol. 2. 1955.

176. Timoshenko, S., *Strength of Materials*. Vol. 1. 1955.

177. Frauenfelder, H.*, et al.*, *Thermal expansion of a protein.* Biochem., 1987. **26**(1): p. 254-61.

178. Hadfield, H. and H. J, *A Fast and Portable Microspectrophotometer for Protein Crystallography.* Journal of Applied Crystallography, 1993. **26**: p. 839-842.

179. Klink, B.U., R.S. Goody, and A.J. Scheidig, *A newly designed microspectrofluorometer for kinetic studies on protein crystals in combination with x-ray diffraction.* Biophys J, 2006. **91**(3): p. 981-92.

180. Zacharias, D.A*., et al.*, *Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells.* Science, 2002. **296**(5569): p. 913-6.

181. Collins, M.D., PhD Thesis, *High-Pressure X-ray Crystallography and Core Hydrophobicity of T4 Lysozyme*, 2006, Physics Department, Cornell University.

182. Kleywegt, G.J. and T.A. Jones, *Braille for Pugilists.* 1995.

183. Perrakis, A., R.J. Morris, and V.S. Lamzin, *Automated protein model building combined with iterative structure refinement.* Nat Struct Biol, 1999. **6**(5): p. 458-63.

184. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

185. Kabsch, W., *Solution For Best Rotation To Relate 2 Sets Of Vectors*, in *Acta Crystallogr A*. 1976. p. 922-923.

186. Schirmer, T. and P.R. Evans, *Structural basis of the allosteric behaviour of phosphofructokinase.* Nature, 1990. **343**(6254): p. 140-5.

187. Long, S.B.*, et al.*, *Atomic structure of a voltage-dependent K+ channel in a lipid membrane-like environment.* Nature, 2007. **450**(7168): p. 376-82.

188. Carey, F., F. Knowles, and Q. Gibson, *Effect of hydrostatic pressure on ligand binding to hemoglobin.* J Biol Chem, 1977. **252**(12): p. 4102-7.

189. Oliphant, T.E., *Guide to NumPy*, 2006, http://numpy.scipy.org/numpybook.pdf

190. Sanner, M.F., A.J. Olson, and J.C. Spehner, *Reduced surface: an efficient way to compute molecular surfaces.* Biopolymers, 1996. **38**(3): p. 305-20.

191. Kabsch, W. and C. Sander, *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features.* Biopolymers, 1983. **22**(12): p. 2577-637.

192. Li, X.*, et al.*, *Deletions of the Aequorea victoria green fluorescent protein define the minimal domain required for fluorescence.* J Biol Chem, 1997. **272**(45): p. 28545-9.

193. Herberhold, H.*, et al.*, *Characterization of the pressure-induced intermediate and unfolded state of red-shifted green fluorescent protein--a static and kinetic FTIR, UV/VIS and fluorescence spectroscopy study.* J Mol Biol, 2003. **330**(5): p. 1153-64.

194. Heremans, K. and L. Smeller, *Protein structure and dynamics at high pressure.* Biochim Biophys Acta, 1998. **1386**(2): p. 353-70.

195. Foote, J. and A. Raman, *A relation between the principal axes of inertia and ligand binding.* Proc Natl Acad Sci USA, 2000. **97**(3): p. 978-83.

196. Mauring, K.*, et al.*, *Enhancement of the fluorescence of the blue fluorescent proteins by high pressure or low temperature.* The journal of physical chemistry B, 2005. **109**(26): p. 12976-81.

197. Ganesan, S.*, et al.*, *A dark yellow fluorescent protein (YFP)-based Resonance Energy-Accepting Chromoprotein (REACh) for Forster resonance energy transfer with GFP.* Proc Natl Acad Sci U S A, 2006. **103**(11): p. 4089-94.

198. Yarbrough, D.*, et al.*, *Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0-\Å resolution.* Proc Natl Acad Sci USA, 2001. **98**(2): p. 462-7.

199. Gross, L.A.*, et al.*, *The structure of the chromophore within DsRed, a red fluorescent protein from coral.* Proc Natl Acad Sci U S A, 2000. **97**(22): p. 11990-5.

200. Mozhaev, V.*, et al.*, *High pressure effects on protein structure and function.* Proteins, 1996. **24**(1): p. 81-91.

201. Le Tilly, V.*, et al.*, *An infrared study of 2H-bond variation in myoglobin revealed by high pressure.* Eur J Biochem, 1992. **205**(3): p. 1061-5.

202. Ando, N.*, et al.*, *Structural and thermodynamic characterization of T4 lysozyme mutants and the contribution of internal cavities to pressure denaturation.* Biochemistry, 2008. **47**(42): p. 11097-109.

203. Hummer, G.*, et al.*, *The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins.* Proc Natl Acad Sci USA, 1998. **95**(4): p. 1552-5.

204. Oger, P.M., I. Daniel, and A. Picard, *Development of a low-pressure diamond anvil cell and analytical tools to monitor microbial activities in situ under controlled P and T.* Biochim Biophys Acta, 2006. **1764**(3): p. 434-42.

205. Leiderman, P.*, et al.*, *Effect of temperature on excited-state proton tunneling in wt-green fluorescent protein.* J Phys Chem B, 2008. **112**(4): p. 1232-9.

206. Lill, M.A. and V. Helms, *Proton shuttle in green fluorescent protein studied by dynamic simulations.* Proc Natl Acad Sci U S A, 2002. **99**(5): p. 2778-81.

207. Tobi, D. and I. Bahar, *Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 18908-13.

208. Min, W.*, et al.*, *Fluctuating enzymes: lessons from single-molecule studies.* Acc Chem Res, 2005. **38**(12): p. 923-31.

209. Kou, S.*, et al.*, *Single-molecule Michaelis-Menten equations.* The journal of physical chemistry B, 2005. **109**(41): p. 19068-81.

210. Sinicropi, A.*, et al.*, *Properties of the emitting state of the green fluorescent protein resolved at the CASPT2//CASSCF/CHARMM level.* J Am Chem Soc, 2005. **127**(33): p. 11534-5.

211. Wachter, R.M.*, et al.*, *Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein.* Biochem., 1997. **36**(32): p. 9759-65.

212. Hay, S.*, et al.*, *Barrier compression enhances an enzymatic hydrogen-transfer reaction.* Angew Chem Int Ed Engl, 2009. **48**(8): p. 1452-4.

213. Urbanowicz, B.*, et al.*, *A Tomato Endo-beta-1,4-glucanase, SlCel9C1, Represents a Distinct Subclass with a New Family of Carbohydrate Binding Modules (CBM49)*. Journal of Biological Chemistry, 2006. **282**(16): p. 12066-12074.

214. Demain, A.L., M. Newcomb, and J.H.D. Wu, *Cellulase, Clostridia, and Ethanol.* Microbiology and Molecular Biology Reviews, 2005.

215. Li, Y., D.C. Irwin, and D.B. Wilson, *Processivity, substrate binding, and mechanism of cellulose hydrolysis by Thermobifida fusca Cel9A*. Appl Environ Microbiol, 2007. **73**(10): p. 3165-72.

216. Personal communication, Wilson, D.B. to B. Barstow, 2008

217. Coleman, D.J., M.J. Studler, and J.J. Naleway, *A long-wavelength fluorescent substrate for continuous fluorometric determination of cellulase activity: resorufin-beta-D-cellobioside.* Anal Biochem, 2007. **371**(2): p. 146-53.

218. Sakon, J.*, et al.*, *Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca.* Nat Struct Biol, 1997. **4**(10): p. 810-8.

219. Jacobson, M.Z., *Effects of ethanol (E85) versus gasoline vehicles on cancer and mortality in the United States.* Environ Sci Technol, 2007. **41**(11): p. 4150-7.

220. Magnusson, R., C. Nilsson, and B. Andersson, *Emissions of aldehydes and ketones from a two-stroke engine using ethanol and ethanol-blended gasoline as fuel.* Environ Sci Technol, 2002. **36**(8): p. 1656-64.

221. Navarro, R.M., M.A. Pena, and J.L. Fierro, *Hydrogen production reactions from carbon feedstocks: fossil fuels and biomass.* Chem Rev, 2007. **107**(10): p. 3952-91.

222. Lewis, N.S., *Toward cost-effective solar energy use.* Science, 2007. **315**(5813): p. 798-801.

223. Vincent, K.A., A. Parkin, and F.A. Armstrong, *Investigating and exploiting the electrocatalytic properties of hydrogenases.* Chem Rev, 2007. **107**(10): p. 4366-413.

224. Tromp, T.K.*, et al.*, *Potential environmental impact of a hydrogen economy on the stratosphere.* Science, 2003. **300**(5626): p. 1740-2.

225. JASON, *Engineering Microorganisms for Energy Production*, 2006, The MITRE Corporation, McLean, Virginia 22102

226. Ihara, M.*, et al.*, *Light-driven Hydrogen Production by a Hybrid Complex of a [NiFe]-Hydrogenase and the Cyanobacterial Photosystem I.* Photochem Photobiol, 2006. **82**(3): p. 676.

227. Ghirardi, M.L.*, et al.*, *Hydrogenases and hydrogen photoproduction in oxygenic photosynthetic organisms.* Annual review of plant biology, 2007. **58**: p. 71-91.

228. Montet, Y.*, et al.*, *Gas access to the active site of Ni-Fe hydrogenases probed by X-ray crystallography and molecular dynamics.* Nat Struct Biol, 1997. **4**(7): p. 523-6.

229. Buhrke, T., *Oxygen Tolerance of the H2-sensing [NiFe] Hydrogenase from Ralstonia eutropha H16 Is Based on Limited Access of Oxygen to the Active Site.* Journal of Biological Chemistry, 2005. **280**(25): p. 23791-23796.

230. Cohen, J.*, et al.*, *Finding gas diffusion pathways in proteins: application to O2 and H2 transport in CpI [FeFe]-hydrogenase and the role of packing defects.* Structure, 2005. **13**(9): p. 1321-9.

231. Cohen, J.*, et al.*, *Molecular dynamics and experimental investigation of H(2) and O(2) diffusion in [Fe]-hydrogenase.* Biochem Soc Trans, 2005. **33**(Pt 1): p. 80-2.

232. De Lacey, A.L.*, et al.*, *Activation and inactivation of hydrogenase function and the catalytic cycle: spectroelectrochemical studies.* Chem Rev, 2007. **107**(10): p. 4304-30.

233. Frey, M., *Hydrogenases: hydrogen-activating enzymes.* Chembiochem, 2002. **3**(2-3): p. 153-60.

234. Doyle, D.A.*, et al.*, *The structure of the potassium channel: molecular basis of K+ conduction and selectivity.* Science, 1998. **280**(5360): p. 69-77.

235. Toniolo, A., Granucci, and T.J. Martinez, *Conical Intersections in Solution: A QM/MM Study Using Floating øccupation Semiempirical Configuration Interaction Wave Functions.* Journal of Physical Chemistry A, 2003. **107**(19): p. 3822-3830.

236. Depristo, M., P. Debakker, and T. Blundell, *Heterogeneity and Inaccuracy in Protein Structures Solved by X-Ray Crystallography.* Structure, 2004. **12**(5): p. 831-838.

237. Liu, R., W.A. Baase, and B.W. Matthews, *The introduction of strain and its effects on the structure and stability of T4 lysozyme.* J. Mol. Biol., 2000. **295**(1): p. 127-45.

238. Sagermann, M., L. Gay, and B.W. Matthews, *Long-distance conformational changes in a protein engineered by modulated sequence duplication.* Proc. Natl. Acad. Sci. U.S.A., 2003. **100**(16): p. 9191-5.