Theses and Dissertations

December 2015

Porphyrin as a Spectroscopic Probe of Net Electric Fields in Heme Proteins

Hannah Elizabeth Wagie
University of Wisconsin-Milwaukee

Follow this and additional works at: https://dc.uwm.edu/etd
Part of the Biophysics Commons, and the Physical Chemistry Commons

PORPHYRIN AS A SPECTROSCOPIC PROBE OF

NET ELECTRIC FIELDS IN HEME PROTEINS

by

Hannah E. Wagie

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Chemistry

at

The University of Wisconsin–Milwaukee

December 2015

ABSTRACT

PORPHYRIN AS A SPECTROSCOPIC PROBE OF NET ELECTRIC FIELDS IN HEME PROTEINS

by

Hannah E. Wagie

The University of Wisconsin–Milwaukee, 2015
Under the Supervision of Professor Peter Geissinger

Heme proteins have diverse functions as well as varied structures but share the same organic, conjugated cofactor. Similarly varied approaches have been taken to deduce how heme can take on different roles based on its protein environment. A unique approach is to view the protein matrix as a constellation of point charges that generates a defined, reproducible, net internal electric field that has influence over the electronic properties of the heme cofactor. This work considers how porphyrins, the basic chromophore building block of heme, can be used as a native spectroscopic sensor of internal electric field at the active site of heme proteins.

First, a number of approaches to model the electrostatic nature of protein structure are described. One approach based on Coulomb's law is used to estimate the net electric field in myoglobin, easily placing the internal electric field on the order of MV/cm. A closer inspection of myoglobin structure reveals that slight changes in position or strategic mutations can cause appreciable change in the field magnitude and direction. Then, the idea of a porphyrin probe is further developed, followed by a theoretical and spectral characterization

of porphyrins substituted into heme proteins for use in emission spectroscopy as non-emissive heme must be replaced by other porphyrin analogs with higher quantum yield.

Once the porphyrin–protein system has been established as the guest–host system of interest, the hole-burning Stark spectroscopy method was used to quantitatively measure the magnitude and direction of the internal electric field vector generated by the protein. The collected Stark spectra had a more established classical analysis method for analysis, but a major aspect of this work is a quantum-mechanical analysis method that has been advanced for more practical and widespread usage. This novel quantum-mechanical approach to the method has promise for greater accuracy for internal electric field determination as well as the ability to resolve the field into spatial components in order to determine not just field magnitude but also direction. The results from the new analysis of experimental data for myoglobin of the in-plane components of the field places both at 1.7 MV/cm. Finally, two *ab initio* excited-state methods, CIS and TDDFT, were used to calculate electronic state energies and transition dipole moment values in support of this new quantum-mechanical analysis method. The two methods are described thoroughly with presentation of benefits and drawbacks to each method.

To Estelle Violet Wagie and her new brother, that the evidence of effort documented in this volume might help you to cultivate the courage to follow where you are led, even if it means you must blaze the trail.

TABLE OF CONTENTS

LIST OF FIGURES

b) Intensity of Soret peak over 5 hours reveals a number of steps with individual rates. The change in slope for the plot is indicated by arrows.

LIST OF TABLES

LIST OF ABBREVIATIONS

α..polarizability

β...hyperpolarizability

ε...molar absorptivity or energy

λ_{em}..emission wavelength

λ_{ex}...excitation wavelength

μ..[transition] dipole moment

A..absorbance

APD..avalanche photodiode

a.u. ....atomic units

CIS...configuration interaction–singles

cps...counts per second

DFT..density functional theory

DMSO...dimethyl sulfoxide

$\vec{E}$..electric field

f...local field factor

G09...*Gaussian09*

TDDFT…………………………………………………………….time-dependent density functional theory

UV–VIS………………………………………………………………………………………….ultraviolet–visible

V / cm…………………………………………………………………………………….volts per centimeter

ZnPPIX………………………………………………………………………..…………zinc protoporphyrin IX

ZnPPIX DME……………………………………………………… zinc protoporphyrin IX dimethyl ester

ZPL……………………………………………………………………………………….zero-phonon line

ACKNOWLEDGMENTS

Chapter 1

Introduction

1.1 Introduction to net electrostatics in proteins

While electrostatic interactions are part of most investigations of the determinant of the function of biological systems (e.g. proteins), the level of complexity of describing these interactions is more often than not limited to assigning individual amino acids the characteristics of polar or non-polar, charged or neutral. This both limits the volume in which the effect influence is noted within a protein structure as well as the mechanism by which electrostatic interactions are noted to be effective (attraction or repulsion of a ligand). On the other hand, many reports of protein research invoke a more global electrostatic environment provided by the sum of the entire protein environment to explain protein function. These net electrostatic environments are also often given a rather vague description of "positive" or "negative," because tools to quantify them are still in development.

This work contributes towards the goal of developing a method to measure net electric fields generated by heme proteins using hole-burning Stark spectroscopy, and from these come to a more detailed description of the electrostatic environment of active sites. Not only would this allow for a more detailed understanding of protein function in terms of contributions of the various amino acids and the importance of their location with respect to the active site, but allow for synthesis of artificial (i.e. non-biological) molecular environments that bestow a certain set of functional characteristics on a molecule of interest (e.g., synthesis of artificial blood substitutes).

The measurement of internal electric field at a microscopic level inside of a protein, however, requires a distinct approach from the routine measurement of such a field at a macroscopic level, e.g., across a semiconductor. The measurement of electric field around a macroscopic object can be performed by introducing another object as a measurement tool, e.g., a metal rod as a probe, in its vicinity. For microscopic objects, however, the introduction of a probe to measure electric field must be careful to not disturb the object itself, although for molecular probes the effect of the probe on the object to be measured may not be negligible. For *internal* electric field, it is implied that the region of interest is deep within some structure. This limits the tools to sense the field to native reporters (see examples in Chapter 3). Then, the necessarily remote sensing of the field leaves few choices for techniques with spectroscopy as a clear option to provide a window into the native probe's state in the presence of the electric field.

Examples of the areas in protein science that invoke net electrostatic properties and interactions as part of functional mechanisms are

- redox potential in metalloproteins (including heme proteins)[1]

- ligand stabilization and discrimination[2]

- electron transfer (especially in photosynthetic complexes) [3]

- general catalytic function[4]

- and protein folding[5] and dynamics[6]

Three studies that call on net electrostatics mechanistically in a particularly compelling manner are described here. They have specifically been chosen as non-heme proteins because heme protein examples will arise later in the manuscript.

The first report, "Electrostatic Steering of Substrate to Acetylcholinesterase: Analysis of Field Fluctuations," [7] is a molecular dynamics study in support of previous experimental studies. The well-known neurotransmitter acetylcholine is hydrolyzed by the enzyme acetylcholinesterase as a way to end the neurological signal. The enzyme acts at a site deep inside its structure and the ligand must traverse down a "gorge" gated by a "bottleneck" opening. However, this enzyme acts extremely swiftly at nearly diffusion rate. It has been shown computationally in agreement with experimental observations that electrostatic "steering" of acetylcholine increases the rate of reaction for this system by two orders of magnitude. This electric field is thought to originate from charge distributions around the opening to the long path to the active site. Mutant studies of individual amino acids in this area do not significantly affect the reaction rate, leading to the conclusion that it is not just specific points in the protein that cause this large-scale effect but "rather from the whole constellation of partial charges in the enzyme." In fact, this study goes beyond a static electric field to suggest that the fluctuations of the field also contributes to the effect of funneling acetylcholine down a particular path to the active site. The idea of a dynamic field arises naturally in this work when the statistical variations in myoglobin structure are more closely examined in Chapter 3.

Another enzyme whose catalytic rate is diffusion limited has a mechanism that relies not on the electrostatics of a single amino acid but of a network. "Faster Superoxide Dismutase (SOD) Mutants Designed by Enhancing Electrostatic Guidance"[8] does use site-specific mutants at four particular charged residue locations, however, the trend indicates that only by acting in concert are the kinetics enhanced. The human Cu,Zn–SOD enzyme examined in this study acts to disarm the harmful superoxide radical, O_2^-, by using it to produce dioxygen, O_2, and hydrogen peroxide,

H_2O_2. Like acetylcholinesterase, the superoxide substrate of SOD seems to be guided to the copper-ion active site electrostatically. Looking at the electrostatics of the wild-type enzyme in a more traditional way, the authors noted that the overall negative charge of the enzyme should prevent the negatively charged superoxide to reach the active site. The four-residue hydrogen-bonded network at the entrance to the opening to the active site include two negatively charged glutamates. Surprisingly, mutations of these residues to positively charged glutamine residues in fact *slows* the kinetic rate of enzyme. The conclusion was that although the individual charges on these residues is usually viewed as disadvantageous, the individual electrostatic characteristics were essential in forming a network that stabilized the configuration of the residues around the active-site entrance: "structural interactions of electrostatically important side chains orient the electrostatic fields and provide a basis for understanding the resultant rates of mutant enzymes." In this work, another discussion in Chapter 3 about electrostatic structure and Coulomb's law also emphasizes the importance of not just charge but also position of the charge.

Finally, a 2014 study, "Extreme Electric Fields Power Catalysis in the Active Site of Ketosteroid Isomerase"[9] has an experimental approach similar to the one used in this work in that it utilizes molecular probes as sensors of electric fields (see Chapter 4) with Stark spectroscopy (see Chapter 6). The enzyme of interest in this report, ketosteroid isomerase (KSI), is part of steroid biosynthesis and degradation. It acts by a proton transfer that results in a large increase in dipole moment along the carbonyl of a keto-intermediate. However, this high-energy formation step is performed with one of largest unimolecular rate constants known for an enzyme. One of the postulates for how the enzyme accomplishes this is that the protein itself generates and applies an electric field in a direction that mitigates such strong charge separation

and serves to stabilize the intermediate. Experimentally, the study measured the magnitude of the electric field at various points within the KSI structure with a carbonyl (CO) probe and compared the values with that at the active site. It was found the field strength at the active site was "extreme" compared to other sites in the enzyme, on the order of 1.5 MV/cm. Mutants to the active site decreased the field strength and it was found that the catalytic rate of the enzyme had a linear dependence on the magnitude of the electric field. The study also examined the source of such a large field. Two individual amino acids with –OH groups seemed to be candidates for large contributors to the field, but like the superoxide dismutase study, it was concluded that on their own, the two residues could not account for the magnitude of the measured fields. As concluded in the previous two studies, it is the overall field produced by the structure that seems to be the important agent for function.

1.2   Previous efforts in the measurement of internal electric fields in proteins

Given the importance of these electric fields for biological function, the development of methods to measure these fields were required. The challenge is to obtain information in a small region of space, i.e. the active site, within the much larger volume occupied by the entire biological systems under study. Thus, experimental approaches had to be found that selective probe the active site regions for these electric fields. Fortuitously, for heme cofactors, their optical absorption is in many cases quite distinct from that of the rest of the considered biosystem. By tuning a (laser) light source selective to the heme absorption wavelength, these can be selectively addressed with the molecular environment (i.e. the amino acids) registered as a perturbation to the heme electronic states.   Alternatively, characteristic vibrational

spectroscopic signature may also be useful, as shown by Boxer et al.[10] who considered the CO stretch frequency when CO was bound to heme.

The use of electronic absorption measurements to determine electronic fields originated in the Kohler group at the University of California, Riverside and continued at the University of Wisconsin–Milwaukee in the Geissinger and Woehl groups. The interest in measuring internal electric fields with high-resolution Stark spectroscopy was at first applied to polyenes in *n*-alkane matrices (e.g., octatetraene in hexane) at low temperature.[11] Using the spectral hole-burning technique to resolve the Stark shift in those solid matrices, the surprising results were that the non-polar bulk solvent produced a net electric field of considerable magnitude–on the order of MV/cm! Analysis was carried out at both molecular and atomic resolution to quantify the field along the length of linear polyene probe molecules. The extension of that work to biological systems turned naturally to heme proteins, where the porphyrin macrocycle is spectrally related to linear polyenes.[12] The protein matrix was also recognized as a complicated but conserved polymer that would likely produce a specific internal electric field vector that could be measured reproducibly and that may have significant functional relevance for the protein as indicated by the studies mentioned above. The initial proteins of interest were myoglobin and cytochrome *c* and successful internal electric field measurements were made in these systems.[13] In addition, a new method to analyze Stark spectra was pioneered taking a quantum-mechanical approach to the problem.[11a, 14] This advance showed promise to improve accuracy over the classical approach as well as providing a way to deduce directional information from the internal electric field vector as well as magnitude.

Another research group interested in measuring internal electric field in biological systems with spectroscopy is the Boxer group at Stanford University. Earlier work by this group was related to ours with a focus on large photosynthetic complexes with embedded chlorophyll molecules (and other porphyrin-related chromophores) using hole-burning Stark spectroscopy to shed light on the effect of internal electric fields on electron transfer.[1c, 3c-e, 15] The group has also employed other spectroscopic methods, namely electroabsorption[16] and the vibrational Stark spectroscopy,[17] for the study of both photosynthetic and other systems, e.g., green fluorescent protein,[18] myoglobin,[1c, 19] human aldose reductase.[20] Most recently, the group's focus has shifted to the calibration and implementation of diatomic ligand probes, carbonyl[21] and nitrile,[22] as described in the ketosteroid isomerase study above. A discussion of the use of different molecular probes and advantages of porphyrin over others is found in Chapter 4.

The efforts described in this volume sought to address some shortcomings that arose with the use of porphyrin probes to measure internal electric fields in proteins. Experimentally, a gain in spectral resolution was sought to improve accuracy of the measured field value and a new experimental set-up with a more sophisticated cryostat, laser system, optics, and detection was purchased (see Chapter 6). Other porphyrin probes, especially metalloporphyrins, were desired to test the theory of reproducibility of the internal electric field in previously tested proteins as well as gauge the effects of a reaction field (induced by a metal center) and a "self" field generated by charged propionate groups (see Chapters 4 and 5). Finally, the implementation of the promising quantum-mechanical Stark analysis previously had practical challenges that were overcome with additional insight into the problem (see Chapter 7) as well as better computational resources (see Chapter 8).

1.3   Manuscript organization

This manuscript has been organized topically to include both theory and experimental or computational work in each chapter with the idea that the reader has reference information nearby when looking at a particular section. The computational and experimental sections generally fall into separate chapters, with the exception being the materials and methods sections, Chapter 2, as simply a listing of instrumentation and computational tools that can be referenced as needed.

The experimental chapters are Chapters 4, 5, and 6. Chapter 4, "Porphyrin Probes," describes the idea of molecular probes (like that in the ketosteroid isomerase study) and specifically the use of porphyrin as a probe. It also presents procedures for substituting porphyrin for heme in heme proteins, as this work clearly distinguishes molecular probe from protein environment. The discussion of porphyrins continues in Chapter 5, "Spectral Characterization of Porphyrins in Heme Proteins," with the well-established theory of porphyrin absorption spectra (Gouterman's "four-orbital model"). Then the changes to absorption and emission spectra of porphyrins in different environments are detailed. The experimental method to measure net electric fields in heme proteins is laid out in Chapter 6, "Hole-Burning Stark Spectroscopy," with limited results.

The computational chapters, Chapters 3, 7, and 8, represent the bulk of this work's contribution to measuring electric fields in proteins. The problem of modeling net electrostatics in proteins is tackled in Chapter 3, "Calculation of Internal Electric Field in Proteins." This separation of the protein (environment) from the porphyrin (probe) is again representative of

the approach this work takes to measuring net electric fields. Theoretical descriptions of how to treat the protein as a field-generating matrix are given and quantitative results of calculations based on a Coulomb's law as well as a quantum-mechanical approach are presented. Then, after the experimental chapters, the manuscript returns to the analysis of Stark spectroscopy data in order to extract the net electric field value that is the goal of this work. Chapter 7, "Quantum-Mechanical Stark Analysis," describes a major advancement in a unique quantum-mechanical approach to the analysis of the measurement technique detailed in Chapter 6. This opens the door for the more widespread implementation of an analysis method that is potentially more accurate for internal electric field determination over the more typical classical analysis. Finally, Chapter 8, "*Ab initio* Calculations of Ground and Excited-State Properties of Porphyrins," contains a collection of computational work that includes essential excited-state input for the quantum-mechanical Stark analysis method given in Chapter 7. In addition, some calculations on the ground-state of porphyrins to begin to quantify the effect of the field on the electronic structure of the cofactor. The manuscript wraps up with a summary of conclusions and future directions for this work in Chapter 9.

1.4   References

1.   (a) Kassner, R. J., Effects of Nonpolar Environments on the Redox Potentials of Heme Complexes. *Proceedings of the National Academy of Sciences* **1972,** *69* (8), 2263-2267; (b) Kassner, R. J., Theoretical model for the effects of local nonpolar heme environments on the redox potentials in cytochromes. *Journal of the American Chemical Society* **1973,** *95* (8), 2674-2677; (c) Varadarajan, R.; Zewert, T.; Gray, H.; Boxer, S., Effects of buried ionizable amino acids on the reduction potential of recombinant myoglobin. *Science* **1989,** *243* (4887), 69-72; (d) Gunner, M. R.; Honig, B., Electrostatic control of midpoint potentials in the cytochrome subunit of the Rhodopseudomonas viridis reaction center. *Proceedings of the National Academy of Sciences* **1991,** *88* (20), 9151-9155; (e) Gunner, M. R.; Alexov, E.; Torres, E.; Lipovaca, S., The importance of the protein in controlling the electrochemistry of heme metalloproteins: methods of calculation and analysis. *Journal of Biological Inorganic Chemistry* **1997,** *2* (1), 126-134; (f) Mauk, A. G.; Moore, G. R., Control of metalloprotein redox potentials: what does site-directed mutagenesis of hemoproteins tell us? *JBIC Journal of Biological Inorganic Chemistry* **1997,** *2* (1), 119-125.

2.   (a) Kepp, K. P.; Dasmeh, P., Effect of Distal Interactions on O2 Binding to Heme. *The Journal of Physical Chemistry B* **2013,** *117* (14), 3755-3770; (b) Springer, B. A.; Sligar, S. G.; Olson, J. S.; Phillips, J., George N., Mechanisms of Ligand Recognition in Myoglobin. *Chemical Reviews* **1994,** *94*, 699-714; (c) Olson, J. S.; Phillips Jr, G. N., Myoglobin discriminates between O2, NO, and CO by electrostatic interactions with the bound ligand. *JBIC Journal of Biological Inorganic Chemistry* **1997,** *2* (4), 544-552; (d) Lee, L.-P.; Tidor, B., Optimization of binding electrostatics: Charge complementarity in the barnase-barstar protein complex. *Protein Science* **2001,** *10* (2), 362-377; (e) Ormos, P.; Szaraz, S.; Cupane, A.; Nienhaus, G. U., Structural Factors Controlling Ligand Binding to Myoglobin: A Kinetic Hole-Burning Study. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (12), 6762-6767; (f) Guven, G.; Atilgan, A. R.; Atilgan, C., Protonation States of Remote Residues Affect Binding-Release Dynamics of the Ligand but Not the Conformation of Apo Ferric Binding Protein. *The Journal of Physical Chemistry B* **2014**; (g) Gerencser, L. S.; Boros, B. t.; Derrien, V.; Hanson, Deborah K.; Wraight, Colin A.; Sebban, P.; Marati, P. t., Stigmatellin Probes the Electrostatic Potential in the QB Site of the Photosynthetic Reaction Center. *Biophysical Journal* **2015,** *108* (2), 379-394; (h) van der Kamp, M. W.; Mulholland, A. J., Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry* **2013**.

3.   (a) Purchase, R.; Völker, S., Spectral Hole Burning: Examples from Photosynthesis. *Photosynth. Res.* **2009,** *101* (2), 245-266; (b) Jankowiak, R.; Reppert, M.; Zazubovich, V.; Pieper, J.; Reinot, T., Site Selective and Single Complex Laser-Based Spectroscopies: A Window on Excited State Electronic Structure, Excitation Energy Transfer, and Electron-Phonon Coupling of Selected Photosynthetic Complexes. *Chem. Rev. (Washington, DC, U. S.)* **2011,** *111* (8), 4546-4598; (c) Boxer, S. G.; Goldstein, R. A.; Lockhart, D. J.; Middendorf, T. R.; Takiff, L., Excited

States, Electron-Transfer Reactions, and Intermediates in Bacterial Photosynthetic Reaction Centers. *Journal of Physical Chemistry* **1989,** *93*, 8280-8294; (d) Boxer, S. G., Mechanisms of Long-Distance Electron Transfer in Proteins: Lessons from Photosynthetic Reaction Centers. *Annual Review of Biophysics and Biophysical Chemistry* **1990,** *19*, 267-299; (e) Lockhart, D. J.; Hammes, S.; Franzen, S.; Boxer, S. G., Electric Field Effects on Emission Line Shapes When Electron Transfer Competes wlth Emission: An Example from Photosynthetic Reaction Centers. *Journal of Physical Chemistry* **1991,** *95* (2217-2226); (f) Okamura, M. Y.; Feher, G., Proton Transfer in Reaction Centers from Photosynthetic Bacteria. *Annual Review of Biochemistry* **1992,** *61* (1), 861-896; (g) Rivera, E.; Montemayor, D.; Masia, M.; Coker, D., Influence of Site-Dependent Pigment-Protein Interactions on Excitation Energy Transfer in Photosynthetic Light Harvesting. *The Journal of Physical Chemistry B* **2013**; (h) Franken, E. M.; Neerken, S.; Louwe, R. J. W.; Amesz, J.; Aartsma, T. J., A Permanent Hole Burning Study of the FMO Antenna Complex of the Green Sulfur Bacterium *Prosthecochloris aestuarii*. *Biochemistry* **1998,** *37* (15), 5046-5051; (i) Kaila, V. R. I.; Send, R.; Sundholm, D., The Effect of Protein Environment on Photoexcitation Properties of Retinal. *J. Phys. Chem. B* **2012,** *116* (7), 2249-2258.

4.   (a) Herschlag, D.; Natarajan, A., Fundamental Challenges in Mechanistic Enzymology: Progress toward Understanding the Rate Enhancements of Enzymes. *Biochemistry* **2013**; (b) Xiang, Y.; Duan, L.; Zhang, J. Z. H., Protein's electronic polarization contributes significantly to its catalytic function. *The Journal of Chemical Physics* **2011,** *134* (20), -; (c) Liu, C. T.; Layfield, J. P.; Stewart, R. J.; French, J. B.; Hanoian, P.; Asbury, J. B.; Hammes-Schiffer, S.; Benkovic, S. J., Probing the Electrostatics of Active Site Microenvironments along the Catalytic Cycle for Escherichia coli Dihydrofolate Reductase. *Journal of the American Chemical Society* **2014**; (d) Dutta, B. J.; Bhattacharyya, P. K., Reactivity and Aromaticity of Nucleobases are Sensitive Towards External Electric Field. *The Journal of Physical Chemistry B* **2014**.

5.   Baumketner, A., Electric Field as a Disaggregating Agent for Amyloid Fibrils. *The Journal of Physical Chemistry B* **2014**.

6.   (a) Ji, C.; Mei, Y.; Zhang, J. Z. H., Developing Polarized Protein-Specific Charges for Protein Dynamics: MD Free Energy Calculation of pKa Shifts for Asp26/Asp20 in Thioredoxin. *Biophysical Journal* **2008,** *95* (3), 1080-1088; (b) Vivian, J. T.; Callis, P. R., Mechanisms of Tryptophan Fluorescence Shifts in Proteins. *Biophys. J.* **2001,** *80* (5), 2093-2109.

7.   Wlodek, S. T.; Shen, T.; McCammon, J. A., Electrostatic Steering of Substrate to Acetylcholinesterase: Analysis of Field Fluctuations. *Biopolymers* **2000,** *53* (3), 265-271.

8.   Getzoff, E. D.; Cabelli, D. E.; Fisher, C. L.; Parge, H. E.; Viezzoli, M. S.; Banci, L.; Hallewell, R. A., Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature* **1992,** *358* (6384), 347-351.

9.   Fried, S. D.; Bagchi, S.; Boxer, S. G., Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science* **2014,** *346* (6216), 1510-1514.

10.   Park, E. S.; Andrews, S. S.; Hu, R. B.; Boxer, S. G., Vibrational Stark Spectroscopy in Proteins: A Probe and Calibration for Electrostatic Fields. *The Journal of Physical Chemistry B* **1999,** *103* (45), 9813-9817.

11.   (a) Gradl, G.; Kohler, B. E.; Westerfield, C., Electric field splitting of the octatetraene 1 1Ag-->2 1Ag transition in n-hexane. *Journal of Chemical Physics* **1992,** *97* (9), 6064; (b) Kohler, B. E.; Woehl, J. C., Measuring internal electric fields with atomic resolution. *The Journal of Chemical Physics* **1995,** *102* (20), 7773-7781; (c) Kohler, B. E.; Woehl, J. C., Effects of internal electric fields and electrostatic potentials on optical spectra of linear polyenes. *Synthetic Metals* **1997,** *84* (1), 859-860; (d) Kohler, B. E.; Woehl, J. r. C., Effects of Electrostatic Fields and Potentials on the Electronic Energies of Conjugated Organic Molecules. *The Journal of Physical Chemistry A* **1999,** *103* (14), 2435-2445; (e) Geissinger, P., Quantitative Measurement of Internal Molecular Electric Fields in Solids. In *Anisotropic Organic Materials: Approaches to Polar Order*, Glaser, R., Kaszynski, Piotr, Ed. American Chemical Society: Washington, DC, 2002; (f) Woehl, J. C. Measuring Internal Electrostatic Fields and Potentials at Molecular and Atomic Resolution using Hole-Burning Spectroscopy. University of California, Riverside, Riverside, CA, 1996.

12.   Gouterman, M., Study of the Effects of Substitution on the Absorption Spectra of Porphin. *The Journal of Chemical Physics* **1959,** *30* (5), 1139-1161.

13.   (a) Geissinger, P., Kohler, Bryan E., Woehl, Jorg C., Electric Field and Structure in the Myoglobin Heme Pocket. *J. Phys. Chem.* **1995,** *99*, 16527-16529; (b) Geissinger, P.; Kohler, B.; Woehl, J. C., Measuring internal electric fields in proteins using hole-burning spectroscopy. *Proceedings of the International Conference on Lasers '96* **1997**, 319-332; (c) Geissinger, P.; Kohler, B. E.; Woehl, J. C., Experimental determination of internal electric fields in ordered systems: Myoglobin and cytochrome C. *Synthetic Metals* **1997,** *84* (1), 937-938; (d) Geissinger, P.; Woehl, J. C.; Kohler, B., Internal electric fields and inhomogeneous broadening: protoporphyrin-IX in myoglobin. *Molecular Crystals Liquid Crystals* **1996,** *283*, 69-74; (e) Geissinger, P.; Woehl, J. C.; Kohler, B., Transition frequency and internal electric field for protoporphyrin-IX in myoglobin. *Molecular Crystals Liquid Crystals* **1996,** *283*, 249-254.

14.   (a) Kohler, B.; Woehl, J. C., Classical and quantum mechanical models for Stark experiments. *Molecular Crystals Liquid Crystals* **1996,** *291*, 119-134; (b) Geissinger, P., Woehl, Jorg C., Prince, Barry J., A quantum-mechanical model for the determination of internal electric fields at protein active sites from the Stark effect on persistent spectral holes. *Journal of Luminescence* **2004,** *107*, 220-229; (c) Geissinger, P.; DuPrey, M. R.; Schwabacher, I. J.; Woehl, J. C.; Prince, B. J., Evaluation and Properties of a Model for the Determination of Internal Electric Fields in Proteins from the Stark Effect of Spectral Holes. *Optics & Spectroscopy* **2005,** *98* (5), 669-674.

15.   (a) Boxer, S. G.; Lockhart, D. J.; Middendorf, T. R., Photochemical Hole-Burning in Photosynthetic Reaction Centers. *Chemical Physics Letters* **1986,** *123* (6), 476-482; (b) Hammes, S. L.; Mazzola, L.; Boxer, S. G.; Gaul, D. F.; Schenck, C. C., Stark spectroscopy of the Rhodobacter

sphaeroides reaction center heterodimer mutant. *Proceedings of the National Academy of Sciences* **1990,** *87* (15), 5682-5686; (c) Steffen, M. A.; Lao, K.; Boxer, S. G., Dielectric Asymmetry in the Photosynthetic Reaction Center. *Science* **1994,** *264* (5160), 810-816; (d) Moore, L. J.; Zhou, H.; Boxer, S. G., Excited-State Electronic Asymmetry of the Special Pair in Photosynthetic Reaction Center Mutants: Absorption and Stark Spectroscopy. *Biochemistry* **1999,** *38* (37), 11949-11960; (e) Treynor, T. P.; Andrews, S. S.; Boxer, S. G., Intervalence Band Stark Effect of the Special Pair Radical Cation in Bacterial Photosynthetic Reaction Centers. *Journal of Physical Chemistry B* **2003,** *107*, 11230-11239; (f) Chuang, J. I.; Boxer, S. G.; Holten, D.; Kirmaier, C., High Yield of M-Side Electron Transfer in Mutants of Rhodobacter capsulatus Reaction Centers Lacking the L-Side Bacteriopheophytin. *Biochemistry* **2006,** *45* (12), 3845-3851; (g) Kanchanawong, P.; Dahlbom, M. G.; Treynor, T. P.; Reimers, J. R.; Hush, N. S.; Boxer, S. G., Charge Delocalization in the Special-Pair Radical Cation of Mutant Reaction Centers of Rhodobacter sphaeroides from Stark Spectra and Nonadiabatic Spectral Simulations. *The Journal of Physical Chemistry B* **2006,** *110* (37), 18688-18702; (h) Carter, B.; Boxer, S. G.; Holten, D.; Kirmaier, C., Trapping the PB Initial Intermediate State of Charge Separation in Photosynthetic Reaction Centers from Rhodobacter capulatus. *Biochemistry* **2009,** *48*, 2571-2573.

16.   Bublitz, G. U., Boxer, Steven G., Stark Spectroscopy: Applications in Chemistry, Biology, and Materials Science. *Annu. Rev. Phys. Chem.* **1997,** *48* (42), 213-240.

17.   Boxer, S. G., Stark Realities. *Journal of Physical Chemistry* **2009,** *113* (2972-2983).

18.   Oltrogge, L. M.; Boxer, S. G., Short Hydrogen Bonds and Proton Delocalization in Green Fluorescent Protein (GFP). *ACS Central Science* **2015,** *1* (3), 148-156.

19.   (a) Decatur, S. M.; Belcher, K. L.; Rickert, P. K.; Franzen, S.; Boxer, S. G., Hydrogen Bonding Modulates Binding of Exogenous Ligands in a Myoglobin Proximal Cavity Mutants. *Biochemistry* **1999,** *38* (34), 11086-11092; (b) Pond, A. E.; Roach, M. P.; Sono, M.; Rux, A. H.; Franzen, S.; Hu, R.; Thomas, M. R.; Wilks, A.; Dou, Y.; Ikeda-Saito, M.; Ortiz de Montellano, P. R.; Woodruff, W. H.; Boxer, S. G.; Dawson, J. H., Assignment of the Heme Axial Ligand(s) for the Ferric Myoglobin (H93G) and Heme Oxygenase (H25A) Cavity Mutants as Oxygen Donors Using Magnetic Circular Dichroism. *Biochemistry* **1999,** *38* (23), 7601-7608.

20.   Xu, L.; Cohen, A. E.; Boxer, S. G., Electrostatic Fields near the Active Site of Human Aldose Reductase: 2. New Inhibitors and Complications Caused by Hydrogen Bonds. *Biochemistry* **2011,** *50* (39), 8311-8322.

21.   Fried, S. D.; Bagchi, S.; Boxer, S. G., Measuring Electrostatic Fields in Both Hydrogen-Bonding and Non-Hydrogen-Bonding Environments Using Carbonyl Vibrational Probes. *Journal of the American Chemical Society* **2013,** *135* (30), 11181-11192.

22.   (a) Andrews, S. S.; Boxer, S. G., Vibrational Stark Effects of Nitriles I. Methods and Experimental Results. *The Journal of Physical Chemistry A* **2000,** *104* (51), 11853-11863; (b)

Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G., Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment with Theory. *Science* **2006,** *313* (5784), 200-204.

Chapter 2

Materials & Methods

2.1   Spectroscopic tools

Spectral characterization is referenced throughout this text. The commercial instruments used to do so are described here in general terms. Details may be given in context of a particular experiment. Both absorption and emission spectroscopy were employed.

2.1.1   Absorption spectroscopy

Ultraviolet-visible (UV-VIS) absorption spectra were obtained with a Lambda 650 dual-beam spectrophotometer (Perkin-Elmer, Waltham, MA). Typically, a spectrum was obtained in scanning mode in 1.0 nm increments at a 0.2 s integration time. The visible region was excited by a halogen lamp while the UV region was excited by a deuterium lamp. Samples were contained in a 1-cm cuvette. For aqueous samples, disposable methacrylate cuvettes (Sigma-Aldrich, Milwaukee, WI) were regularly employed. For organic solvents, quartz cuvettes (Starna Cells, Atascadero, CA) were used. For spectra that reached <350 nm, a Spectrosil cuvette (Aldrich Chemical, Milwaukee, WI) was required, e.g., monitoring tyrosine absorbance at 280 nm for protein samples. A blank, consisting solely of the analyte's solvent, was placed in the second beam path to correct for any solvent absorption.

Absorption spectra was also occasionally collected with an Agilent 8453 photodiode array (Agilent, Santa Clara, CA). This instrument does not scan but acquires a broadband spectrum of the entire UV-VIS region in approximately 100 ms. It had the advantage of avoiding a lamp change

in the region <380 nm. In addition, the quick acquisition time also lent this instrument to kinetics experiments. Samples were contained in cuvettes as above, however, in single-beam fashion with no simultaneous blank.

2.1.2 Emission spectroscopy

Fluorescence spectra were obtained with a Fluorolog-3 fluorometer (Jobin–Yvon [Horiba], Edison, NJ). The instrument was used in two modes: 1) emission scan, where a single excitation wavelength was employed while fluorescence detection was scanned over a specified wavelength range, and 2) fluorescence excitation scan, where a single emission wavelength was detected while the source excitation wavelength was scanned. Parameters like integration time and slit width were specified per experiment. For room temperature acquisition, a 1-cm cuvette with four transparent walls as described above was employed. For low temperature (77 K) acquisition, the instrument-specific liquid nitrogen dewar assembly (Horiba, Edison, NJ) was inserted into the sample compartment where the sample was immersed in liquid nitrogen. Spectra was collected at a 90° angle for liquid samples. For solid samples, either frozen or in solid matrix, spectra was occasionally collected in "front-face reflection" (22° angle from incident beam) mode if scattering losses were high.

2.2 Porphyrin sample preparation tools

2.2.1 Ultrapure water

All porphyrin and substituted protein preparations used distilled water purified with a Barnstead Nanopure ultrapure water system (Thermo Scientific, Marietta, OH). The system produced water with a resistivity of 18.2 MΩ, indicating low ion content, and relatively free of

organic contaminants, RNase, DNase, and bacteria. This was accomplished with a UV lamp, ultrafilter, and final 0.2 micron filter.

2.2.2   Reagents

All reagents and solvents without origin specification in this text were obtained from either Sigma–Aldrich (Milwaukee, WI), Fisher Scientific (Pittsburgh, PA), or VWR International (Radnor, PA). Reagents were employed without further purification. Glycerol for spectroscopy samples was purchased at ultrapure, spectrophotometric grade from VWR (Alfa Aesar label).

2.2.3   pH

Porphyrin and substituted proteins were created in solutions with carefully controlled pH. A VWR Symphony benchtop pH meter (VWR International, Radnor, PA) was employed for pH measurement and adjustment. The meter was able to measure accurately to 0.002 pH unit. The measurements were typically calibrated with three standard buffers (VWR): pH 4.00, pH 7.00, pH 10.00. VWR Symphony probes used were either a glass, refillable, Ag/AgCl referenced with a separate epoxy automatic temperature correction (ATC) probe, or a combination gel, non-refillable, Ag/AgCl pH probe with integrated ATC capability.

2.2.4   Phosphate buffer

Protein solutions were typically buffered with phosphate buffer. Using either sodium or potassium salts, the HPO_4^{2-} ("dibasic phosphate") / $H_2PO_4^-$ ("monobasic phosphate") system provided a pH range appropriate for working with myoglobin and hemoglobin. Stock solutions of 1 M dibasic and 1 M monobasic phosphate were prepared. Then, the appropriate proportions were mixed to achieve the desired pH. Most experiments were performed at pH 7.3, which

17

required a mixture of 77% dibasic and 23% monobasic solution. Then, the resulting 1 M buffer was diluted to 0.1 or 0.05 M for use with protein solutions. Finally, the pH was checked with a pH meter and adjusted to 0.1 pH unit accuracy.

2.2.5   Centrifugation

Centrifugation was employed in preparation of substituted proteins for purification purposes as well as to recover precipitated protein. A Sorvall RC6 Plus floor-model centrifuge with Fiberlite composite rotor (Thermo Scientific, Marietta, OH) was used for samples >1 mL. Typically, an SH-3000 swinging bucket rotor was used at a maximum of 4400 rpm. The internal cooler was set to 4 °C for protein preparation. For samples <1 mL, an IEC Micromax microcentrifuge (Thermo Scientific, Marietta, OH) was used, which was kept to a temperature of approximately 4 °C or less in a walk-in cooler.

2.2.6   High-performance liquid chromatography

For some purification steps in substituted protein preparation, an AKTA high-performance liquid chromatography system (HPLC) (GE Healthcare Life Sciences, Pittsburgh, PA) was employed. Pre-packed ion-exchange and size-exclusion columns were implemented in the system. Details of column type and operating conditions are discussed in context of particular experiments. All samples and buffers were filtered with a 0.45 micron or smaller pore size before loading onto the HPLC.

2.2.7   Long-term protein storage

Once prepared, porphyrin-substituted proteins were stored in buffered solution in polymer microcentrifuge at low temperature in a CryoPro liquid-nitrogen storage dewar (VWR

International, Radnor, PA). Individual samples were placed in labeled aluminum ladles, which were submerged in cryogen.

2.3   Computational tools

2.3.1   *Gaussian09* & *GaussView 5*

For all *ab initio*, density functional theory, and semi-empirical calculations of molecular and excited-state energies and properties, *Gaussian09*, Rev. A.02 (Gaussian Inc., Wallingford, CT)[1] was used. For visualizing results and manipulating structures, *GaussView 5* was used. The Linux version of *Gaussian09* (*G09*) was installed and accessed on two high-performance computing clusters: 1) "Avi" (University of Wisconsin–Milwaukee research) and 2) "Cleve" (Department of Chemistry & Biochemistry at UWM). In addition, the programs were also installed and accessed on a desktop computer in the Department of Chemistry & Biochemistry.

Details of practical usage of Gaussian09 can be found in Appendix A. Here, the basics of crafting an input file, are presented. Theoretical background of these methods are discussed in Chapter 8.

2.3.2   *Stark06* program

A program that calculated internal electric field values from hole-burning Stark spectra, *Stark06*, was developed "in-house" by Dr. Barry J. Prince in 2003. The quantum-mechanical Stark analysis method described in Chapter 7 was implemented in this version of the program. Details of the program can be found in Appendix B, including examples of input files.

2.4   References

1.   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.02*, Gaussian, Inc.: Wallingford, CT, 2009.

Chapter 3

Calculation of Internal Electric Fields in Proteins

3.1 Introduction

As we have illustrated the role of net electrostatics in protein function in Chapter 1, we now turn to ways to predict the net electric field value as a way to validate the experimental values for heme proteins produced by the Stark spectroscopy method. Various methods have been established to model and quantify electrostatic properties in proteins.[1] The methods can be divided into macroscopic and microscopic approaches. As computational resources become increasingly powerful and available, the methods have followed suit, becoming more and more detailed.

Widely used approaches[4] to quantify electrostatics of macromolecules include, on the macroscopic end, the Tanford–Kirkwood treatment (introduced in 1957),[5] which views a protein as a substance with a single dielectric constant containing charged groups embedded on the surface of the spherical macromolecule. On the microscopic end, the Protein-Dipole–Langevin-Dipole (PDLD) approach[6] treats both protein and solvent as a collection of dipoles. In the middle of the spectrum are a variety of so-called Poisson–Boltzmann equation (PBE) solvers, such as the currently free APBS[7] and DelPhi[8] programs. These PBE solvers improve upon a simple macroscopic models, e.g., taking the irregular shape of the protein into account rather than approximating it as spherical; including dipoles at specific locations in the protein interior. Overall, they use some combination of treating the solvent as a dielectric as well as the protein

with assigned dielectric constant(s). Boundary conditions are also usually implemented in PBE solvers with the goal of realistically describing the charge distribution on the surface an in active sites of the protein.

There are many, many types of electrostatic interactions in microscopic descriptions[9] arising from the best description of a particular region of a protein, e.g., monopole, dipole, multipole, etc., with each interaction bearing its own mathematical description. In addition, the non-covalent interactions sought with by these models can also be organized into short- (Born–Mayer [$1/r^{12}$]; Van der Waals [$1/r^6$]; hydrogen bonding), medium- (Coulombic [$1/r$], with its derivative being electric field), and long-range (solvent effects)[10]. It is important to define the aim of the works amongst the variety of electrostatic pursuits. The experimental method described in Chapter 6 specifically measures the $1/r^2$ electric field at the heme active site in heme proteins, which is generally in the interior of the protein. In addition, it is the *net* field at heme over its surface in the heme pocket, as opposed to any electrostatic mapping of the protein matrix itself. Thus, it is truly the medium-range electrostatic "experience" of the heme provided by the protein host that is described by our measurement.

The following describes first, three approaches to modeling the net electric field in heme proteins that were explored to validate experimental results. Then, unique aspects of protein structure in general and then specifically of myoglobin, the model heme protein used in this work, are discussed. Finally, results for calculations are presented for internal electric field in myoglobin.

3.2 Modeling net electrostatic fields in proteins

The experimental portion of this work views the net electrostatic environment inside of a heme protein as an electric field vector with both magnitude and direction.[11] This is a result of the individual electrostatic structure unique to a particular protein that was discussed in the Introduction (Chapter 1). In attempting to model this electrostatic environment, we take on two other viewpoints as well: a point-charge model, where each individual atom has a specific partial atomic charge and position with respect to the heme active site; and a dielectric constant model, where the entire structure sums as a bulk material to influence the heme active site like a solvent. Each of these three electrostatic models has a unique perspective as discussed in the following.

3.2.1   Point charge model

Coulomb's law forms the basis of the point charge model:

$$\phi = \frac{q}{4\pi\varepsilon_0 r} \tag{3.1}$$

where ϕ is the electrostatic potential, q is the magnitude of the point charge, ε_0 is the dielectric permittivity in free space, and r is the distance between the point charge and the point where potential is measured. The first derivative of this quantity is then the electric field, which, as a vector, takes on the form:

$$\vec{\varepsilon} = \sum \frac{q}{4\pi\varepsilon_0 r^2} \cdot \frac{\hat{r}}{r} \tag{3.2}$$

where $\vec{r}$ indicates a vector. The expression sums over all point charges in the vicinity of the point in space where the field is measured. See illustration in Fig. 3.1.

For proteins, the point charges are individual atoms in the protein matrix. In this model, assignment of magnitude of the point charge, q, is dependent on the partial atomic charge method used, which will be necessarily arbitrary (see Section 3.5.2 below). Partial atomic charge distribution takes into account that neutral atoms and charged ions in a molecule, especially a macromolecule with salt bridges, hydrogen bonds, etc., will not carry charges of exactly zero or exactly ±1, ±2, etc.

The position of these point charges, r, however, is often better defined based on structures from x-ray crystallography or nuclear magnetic resonance (NMR) methods. The position is taken as that of center of the nucleus of each atom. In heme proteins, the distance that r refers to is with reference to the iron center in the heme active site, where electric field will be calculated in this work. Yet, protein structure, too, has some variability that will also be discussed in further detail in Section 3.5.1 below.

This point-charge approach has precedence in the development of a number of force fields implemented in molecular mechanics algorithms. When applied to proteins, the exercise results in the tabulation of bond lengths, bond angles, and partial atom charge values for individual amino acids.[12]
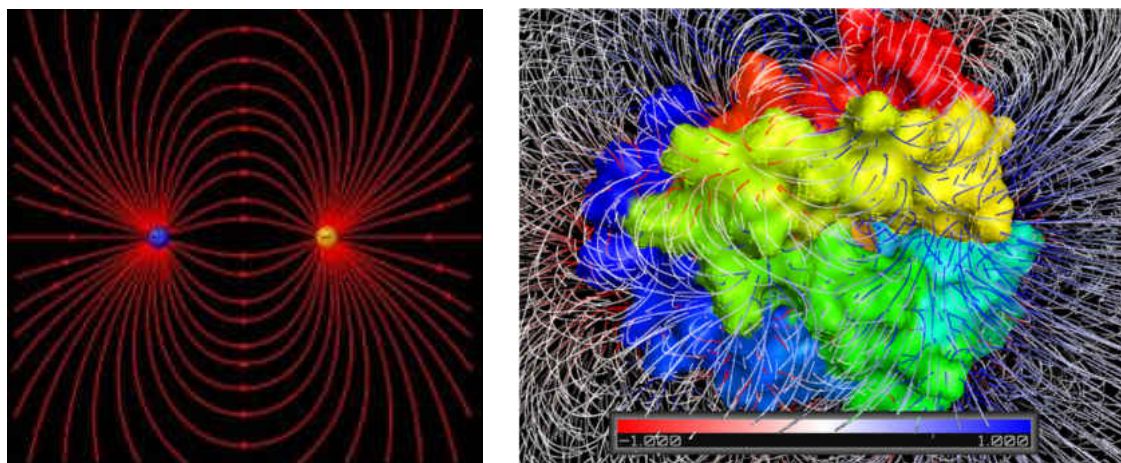
Figure 3.1. a) Illustration of one positive and one negative point charges with electric field lines formed between them. b) Surface model of myoglobin with electric field lines emanating from the surface as visualized with PyMOL.[2] Field lines were calculated with the APBD plugin for PyMOL.

3.2.2 Electric field vector

Modeling the electrostatics of a protein matrix as an electric field vector also has its origin in Coulomb's law and also arises from the sum of contributions from a distribution of charges surrounding the point of measurement. The calculation of a molecular electric field vector, however, does not necessarily rely on assigning point charges but might use a wavefunction for the protein matrix instead in an *ab initio* calculation. From a quantum-mechanical perspective, electrostatic potential is given as:[13]

$$V(r) = \sum_A \frac{Z_A}{|r-R_A|} - \sum_{\mu\nu} P_{\mu\nu} \int \frac{\phi_\mu(r_1)\phi_\nu(r_1)}{|r-r_1|} \qquad (3.3)$$

where *V(r)* is the electrostatic potential at some point *r* (the heme iron in this work). The first part of the expression is the nuclear portion, which is a point charge: *Z* is the nuclear charge of atom *A* and R_A is the position of the nucleus of atom *A*. The second part of the expression is the

25

electronic portion, which is treated as a distribution: $P_{\mu v}$ is the density matrix element of atom A; ϕ_μ and ϕ_v are orbital basis functions at position r_1.

In experiment, however, a static structure is not likely and the resulting vector is an average quantity that takes into account the variability in a dynamic protein.[14] The vector itself might also be allowed dynamic nature, especially if the protein motion is periodic, e.g., protein "breathing."[15]

This viewpoint of the electrostatic structure also necessitates the specific term "internal electric field" vector ($\vec{\mathcal{E}}_{int}$) to differentiate it from a field that might exist outside the protein matrix itself ($\vec{\mathcal{E}}_{ext}$).[11] This external field might be applied as a perturbing field during an experiment (e.g., Stark spectroscopy) or may arise from naturally occurring environmental conditions (e.g., a solvent containing highly charged particles or an independent, charged structure in close proximity). The internal and external electric field vectors sum to yield the total electric field experienced at the heme protein's active site:

$$\vec{\mathcal{E}}_{int} + \vec{\mathcal{E}}_{ext} = \vec{\mathcal{E}}_{tot} \qquad (3.4)$$

The external field portion may be further modulated by the protein matrix itself, which will be further discussed in describing Stark spectroscopy in Chapter 6.

3.2.3 Dielectric constant

An entirely different approach to the electrostatic environment inside of heme proteins is to view the protein matrix with a singular, bulk parameter, the dielectric constant. It is very generally defined as the tendency of a material to undergo electric polarization from charge

displacement when an electric field is applied. From a microscopic point of view, high dielectric constant is correlated with polar materials and low dielectric constant with non-polar materials.

For our purposes, we can define the quantity as a ratio of the permittivity of the material, ε (in farads per meter, F/m), to the absolute electric permittivity of free space, ε_0, which is the ultimate insulator and theoretical minimum for a permittivity measurement. Dielectric constant is synonymous with relative permittivity, ε_r:

$$\varepsilon_r = \frac{\varepsilon}{\varepsilon_0} \tag{3.5}$$

This macroscopic quantity is related to the microscopic idea of polarizability, α, the tendency of a molecule's electron distribution to be polarized in the presence of an electric field, often generated by a dipole in the vicinity of the molecule. The relationship between macroscopic and microscopic for a liquid or solid material is given by the Clausius–Mosotti equation:

$$\frac{\varepsilon-1}{\varepsilon+2}\frac{M}{\rho} = \alpha\frac{N_A}{3\varepsilon_0} \tag{3.6}$$

Here, *M* is the molecular weight of a molecule of the material of interest; ρ is the material's density; N_A is Avogadro's number.

As mentioned earlier, the assignment of dielectric constant to proteins is integral to some methods for quantifying electrostatics in proteins and the determination of the value has become a pursuit in itself.[16] Typically, the value assigned is very low, ε=2.5–4, but for some purpose can be as high as ε=40.[4] However, it is strange to attempt to apply a continuum value to a discrete macromolecule because the dependence of the value on interacting dipoles assumes homogeneity. In addition, the assignment of a single value to any one protein is also

questionable, as the value likely changes from region to region (i.e., dipole to dipole) as well as over time in such a dynamic structure. Though the assignment of a macroscopic quantity to a microscopic structure is inherently flawed,[17] the usefulness of treating the protein matrix as a "solvent" will become clearer in our discussion of porphyrin spectra in Chapter 5 as well as with application to calculations of porphyrin electronic structure below. As such, we will use the variable ε_p with the understanding that the protein dielectric constant is an effective value.

3.3 Protein as a "glassy solvent"

The viewpoint of protein matrix as a solvent for heme is reinforced by experiments which reveal glass-like properties of the material. The experiments in this work are done at low temperature (liquid-nitrogen and liquid-helium temperatures) dissolved in solvents (e.g., glycerol–water mixture, DMSO, etc.) that freeze to form amorphous solids. These conditions impart the unexpected glass-like properties. From x-ray crystallography, the tendency of proteins to crystallize is well documented. The orderly structures the technique reveals reflect the specific tertiary structure that has been so important in elucidating function. However, there are also components of randomness in protein structure that is comparable to glassy (i.e., amorphous) materials, which are characterized by structural relaxation that occurs even at temperatures in the millikelvin realm. These structural relaxations can be seen spectroscopically in time-resolved experiments.[18][19]Changes in spectra can reveal the kinetics of small changes in the position of solvent atoms, side-groups, or molecules with respect to the chromophore. However, experiments that showed similarity of proteins to glasses also show that the extent of the randomness is not as great.[18c] We are left with a picture of overall organization of tertiary structure in proteins that nonetheless retains entropy in time and space.[20]

It is helpful to think of the miniscule changes in a glassy material in terms of an energy landscape that contains numerous local minima, most of them quite shallow.[21]   The same illustration has been applied to protein folding, with a deep global minimum representing the native folded structure (see Fig. 3.1). However, even inside the correct tertiary arrangement, the position of individual amino acids can have more than one option. Experiments at extremely low temperature show sampling of different conformations on an observable timescale for proteins as has been observed for glasses. This is represented by quantum-mechanical tunneling through the small barrier that separates one conformation from another. Experiments that raise the temperature of the system slightly and then re-cool the system also shows spectral changes that reveal conformational changes in both glasses and proteins.[18b] This is represented by the system obtaining enough energy as the temperature rises to overcome a larger energy barrier to then become trapped in the second minimum as the system is cooled again. The availability of a transition between one conformational state and another has been termed a "two-level system" (TLS).[22] For a macromolecule, there are an enormous number of two-level systems for its native structure. One major difference between proteins and glasses is the extent of coupling between spatial regions in the solid system. Proteins are necessarily coupled between regions of the macromolecule while individual regions of glasses are not. This can be illustrated by areas of smoothness on a protein potential energy surface while a glassy solid's is saturated with potential

wells.[23] With respect to internal electric fields in proteins, it is clear that this description of proteins with glassy characteristics implies that the electrostatics are also dynamic.
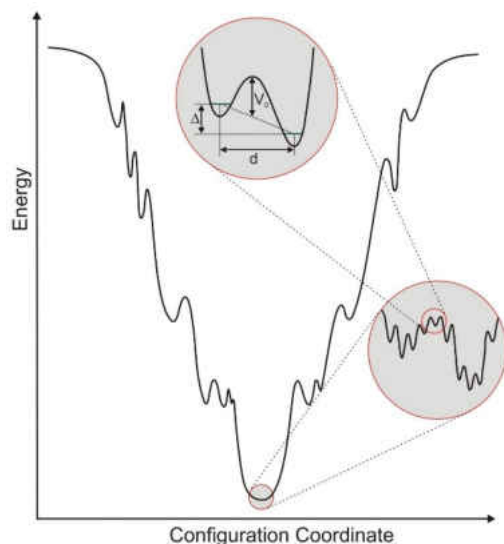
Figure 3.2. Two-level system illustrated for a protein. Local minima are seen in three successive enlarged regions of the potential energy funnel. Zooming in to the smooth minimum of the global well, surface roughness is revealed at better resolution. Then, two wells of the rough surface is then selected as the two-level system to be illustrated. Isolating one of these barriers between two small local minima defines three parameters: Δ = energy asymmetry parameter, d = distance between minima (along a configuration coordinate), and V_0 = barrier height.[3]

3.4 Model heme proteins

In this work, we consider myoglobin as a model heme protein, being both well-studied and simple (the "hydrogen atom of biology") as well as generating ongoing inquiry (a "paradigm of complexity").[24] As a point of contrast and comparison, we will also describe in this section two other heme proteins, hemoglobin and nitrophorin, as candidates for internal electric field calculations and experiment. In any description of a protein's tertiary and/or quaternary arrangement, it is implied the we are discussing its *electrostatic structure*, i.e., the placement of residues with respect to the heme active site has significance as described by Coulomb's law and not necessarily for steric effects. This gives specific structure a longer-range influence than might be initially recognized. The native protein structure will be described followed by alterations that are expected to measurably change electrostatic structure.

3.4.1   Myoglobin

The well-known myoglobin is a monomeric, globular protein that is known primarily for its

dioxygen (O_2) storage capabilities in mammalian muscle tissue. It was the first protein to have its

three-dimensional structure elucidated by x-ray crystallography and published by John Kendrew

in 1958.[25] It is known to have a tertiary structure comprised of eight alpha helices, labeled A

through H, that contain 153 amino acid residues and almost 2500 atoms (including hydrogens).

It is approximately 30 Å in diameter. It has five conserved cavities through which ligands can

traverse: four "xenon holes" in the protein interior and one heme cavity (also known as the "distal

cavity"). The seven residues that are typically considered to "line" the heme cavity are LEU 29,

PHE 43, LYS 45, PHE 46, HIS 64, VAL 68, HIS 93 (using the number from the sequence in Table 3.2

below; amino acid abbreviations are given in Appendix C). The entrance to the protein interior is

unclear. See illustration in Fig. 3.3.
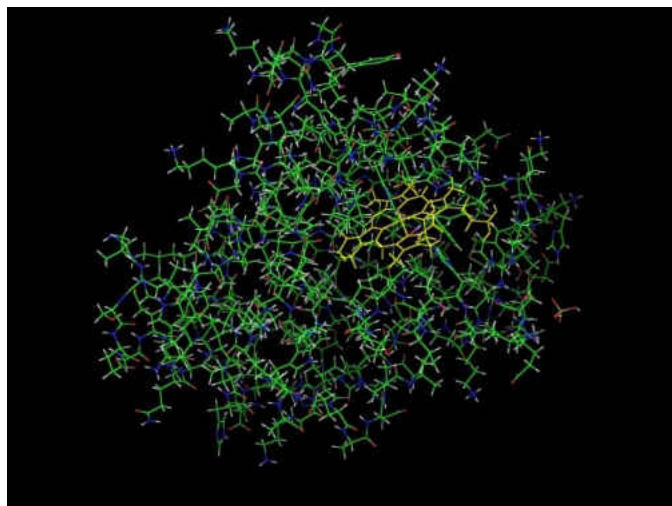
Figure 3.3. Stick structure of myoglobin (PDB code 1mbo) with heme highlighted in yellow, nitrogen in blue, carbon in green, oxygen in red, and hydrogen in gray.

The active site is the iron-centered heme macrocycle, which sits in the heme cavity

approximately 10 Å across with an edge that is solvent-exposed. It is at the iron that diatomic

ligands covalently bind: O_2, carbon monoxide (CO), nitric oxide (NO), and cyanide (CN^-).

Obviously, it is the dioxygen ligand that mammals prefer, but, based on binding affinities, the heme actually has a preference for binding CO. Interestingly, a heme molecule in solution has a binding affinity for CO that is 30 000 – 100 000 times greater than O_2. However, when the heme is embedded in myoglobin as a cofactor, that same ratio drops to *30*![26]

Clearly, the protein matrix assists in the ligand selectivity but the mechanism by which it does so is not entirely clear. Factors likely include the binding geometry of the ligand (perpendicular to the heme plane for CO and angled for O_2),[27] steric effects of the protein matrix[28] and distal histidine[29] (which can hydrogen bond to a bound ligand), the redox potential of the iron,[30] and the slight non-planarity of the heme macrocycle.[31] In this work, we postulate that the directional net electric field, either static or dynamic, generated by the protein matrix may have some bearing on function via ligand specificity. Two suggestions for this mechanism are: 1) influencing the electron distribution over the heme ring, especially at the iron center, to alter bond strength and binding kinetics; 2) directing the rotation of a diatomic ligand in the protein interior cavities to enable or hinder a pathway of the small molecule to the binding site at the iron center. Two interesting perturbations to myoglobin, pH and mutations, should alter the protein's electrostatic structure and hence the internal electric field. They are described below in a way that should inform future Stark spectroscopy experiments to attempt to sense the electrostatic changes and may serve as a way to calibrate sensitivity of the technique.

3.4.1.1 pH perturbation

As suggested by the discussion of two-level systems, myoglobin has a number of conformational substates. They can be induced by the presence or absence of a ligand. It is likely affected by solvent, as active forms of the protein have a one to two layer "hydration shell" with

a particular arrangement of water molecules that is unique compared to the solvent continuum.

An interesting finding is that myoglobin, as opposed to a long-held belief, is allosteric with two defined functions (and possibly more) that are associated with three to four distinct conformational substates. These substates can be induced with adjustments to pH.[32] Two substates have distinct roles: A_1 performs the well-known oxygen storage role while A_0 appears to convert NO to nitrate (NO_3^-). A_1 can be found at high pH values, which seems to keep the distal histidine inside of the heme pocket, and A_0 can be induced by low pH values where the distal histidine leaves the protein interior and rotates out into the solvent. From the electrostatic structure point of view, these substates should generate distinct electric field vectors at the active site due the major shift in the position of an amino acid in close proximity to the heme iron.

For myoglobin, with an isoelectric point (pI) of 7.2, "high pH" and "low pH" are with reference to this point. Experiments performed at pH ≈ 4 would induce a population of myoglobin primarily in the A_0 substate. Experiments performed at pH ≈ 6 would induce a population mixed equally between A_1 and A_3, a substate which is characterized by the protonation of another heme pocket residue, HIS 97, that lies at approximately the same distance from the heme iron as HIS 64. Table 3.1 lists the types and quantities of the ionizable residues in the human myoglobin (PDB code: 1mbo) along with typical pK_a value for the residue type. This was used to build the Coulomb's law model described below and can inform future models.

Table 3.1. Ionizable residues in myoglobin.

Quantity	Type of Amino Acid	pK$_a$	Charge at pH 7.2
4	ARG	12.48	+
6	ASP	3.90	-
14	GLU	4.07	-
12	HIS	6.04	neutral
19	LYS	10.54	+
3	TYR	10.46	neutral

3.4.1.1.1 Titration of glycerol–buffer solution

Because myoglobin experiments would be performed in a glassy glycerol–water mixture, another investigation that was performed to prepare for pH experiments was a titration of a glycerol–buffer mixture to see if that system behaved in a substantially different manner than a completely aqueous pH 7.0 phosphate buffer. As an alcohol with three ionizable protons, glycerol ($C_3H_8O_3$) has the potential to contribute to the solution pH. There is typically only a single pK$_a \approx$ 14 listed for glycerol,[33] which should not be of concern for neutral-range solutions, but the effect of mixing with dilute aqueous buffer was unknown. Phosphate buffer (0.1 M) was created as described in Chapter 2, Materials and Methods, and mixed with glycerol in a 3:1 volume ratio overnight to assure the complete mixing of the aqueous component with the viscous glycerol component.

First, a titration of the aqueous phosphate buffer (pH = 7.0) was performed by adding 0.1 M NaOH dropwise to pH 12 and again, starting at neutral, by adding 0.1 M HCl dropwise to pH 2. The phosphate buffer system should contain three pK values. However, the experimental pH range showed two sigmoidal jumps when "pH vs. volume of acid or base added" was plotted, and only one "half-titration point," which corresponded to pK_2 ($H_2PO_4^- \rightarrow H^+ + HPO_4^{2-}$). The sigmoidal curve was then differentiated using the Origin 7.0 program (Analysis $\rightarrow$ Calculus $\rightarrow$ Differentiate) to determine inflection points. Then the pH at the volume exactly between the inflection points was taken as pK_2.

The same experiment was repeated using the glycerol–buffer mixture, taking extra care to allow each addition to equilibrate. Because the solution contained organic solvent (glycerol), the pK_a is not a strict measurement because, by definition, pH is a measurement of an aqueous system. However, we will use the term "apparent pK_a" to apply to the mixture. At T $\approx$ 20° C, aqueous phosphate buffer (pH = 7.0) measured a pK_2=6.703 and for a 3:1 glycerol:buffer mixture, the apparent pK_2 = 7.054. This compared well to a report for such a mixture at a much lower glycerol concentration of pK_2=6.74.[34]

3.4.1.2 Mutations

Another major structural alteration that also informs both function and electrostatics is mutation. A number of mutations to the active site that can dramatically affect binding affinity and kinetics has been tabulated in Ref. [26]. Because Coulomb's law is based on two variables, distance from the point of measurement and magnitude of charge, interesting mutants will vary: 1) distance from the heme iron, especially at a position where the $1/r^2$ influence is strong; 2) the charge on the mutated residue (positive, negative, or neutral).

Table 3.2. Wild-type human myoglobin secondary sequence:

Sequence Label	Amino Acid Sequence
1	GLSDGEWQQV LNVWGKVEAD IAGHGQEVLI RLFTGHPETL EKFDKFKHLK
51	TEAEMKASED LKK<mark>H</mark>GTVVLT ALGGILKKKG HHEAELKPLA QSHATKHKIP
101	IKYLEFISDA IIHVLHSKHP GDFGADAQGA MTKALELFRN DIAAKYKELG
151	FQG

Efforts were made to research which myoglobin mutants might yield interesting results for Stark spectroscopy experiments. Referencing Springer (1994), values for ratios of equilibrium constants for CO to O_2 binding in myoglobin (K_{CO} / K_{O2}) are tabulated in Table 1 in Ref. [26] and provided guidance. However, only human, sperm whale, and pig myoglobin species were provided in that report while our experiments would use equine myoglobin to compare to previous experiments. A sequence comparison with the online tool NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi, accessed 25 July 2014) revealed that equine myoglobin is 88% similar to Homo sapien (human) and Physeter catadon (sperm whale) myoglobin. The high degree of alignment suggested that the tabulated values likely still apply to equine myoglobin. Then, a consideration of which mutants would provide a reasonable alteration to the electrostatic structure, based on change in position and charge per Coulomb's law, was made. As a result, three mutants to the distal histidine, HIS 64, are suggested for Stark spectroscopy experiments.

First, a H64V mutation is suggested. The mutation from HIS to VAL in sperm whale myoglobin shows in increase of K_{CO} / K_{O2} from 25 to 33 after the mutation. When visualized with the WinCOOT program (http://www.ysbl.york.ac.uk/~lohkamp/coot/wincoot-download.html), it was clear that valine will remove electron density from the vicinity of the heme ring but is not too small so that structural collapse might be prevented. In addition, the removal of HIS 64 seemed to allow LYS 45 (a positively charged residue) more access to the active site. The H64V mutant has been successfully made for equine myoglobin and the crystal structure has been deposited in the Brookhaven Protein Data Bank (PDB; http://www.rcsb.org/) under the PDB ID number 3HC9.[29c]

Second, the H64Y mutant is also suggested. A mutation from HIS to TYR allows the mutated residue to sit at about the same distance from the heme ring and also with an overall neutral charge.[35] However, a WinCOOT visualization revealed that the TYR version would have more electron density near the active site than HIS. More convincingly, this mutation had a huge 100-fold increase in preference for CO over the wild-type protein! The H64Y equine myoglobin structure has also been deposited in the PBD under the code 1YMA.[36]

Finally, the H64E mutant may be interesting. There is no published equilibrium constant data nor a crystal structure for the mutation from HIS to GLU. However, visualization with WinCOOT showed that the GLU residue was about the same size as HIS but would be negatively charged.

In preparation for future mutant production, sequences and primers were determined for the three mutants above and obtained along with E. Coli cells and purification columns. Vectors were ordered from Genscript (Piscataway, NJ) with the sequence given in Table 3.2 but with the

highlighted HIS in position 64 replaced with V, Y, and E for the three respective pUC57 vectors. A stop codon was included and the following restriction sites were avoided: XBA1, BSA1, NDE1, BAMH1, XHO1. For ligation of the specified vectors to pE-SUMO or pET-15b plasmids in a polymerase chain reaction (PCR), the following primer sequences were generated with Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and obtained from Integrated DNA Technologies (IDT, Coralville, IA):

- **Forward primer in SUMO**: GGT CTC AAG GTG CTC AGT CAC ACG CTA CCA A

- **Reverse primer in SUMO**: GCT CTA GAT CAA ACC CTG AAA GCC CAG TTC T

- **Forward primer in pET-15b**: GGA TTC CAT ATG GAA AAC CTG TAT TTT CAG GGT GCT CAG TCA CAC GCT ACC AA

- **Reverse primer in pET-15b**: CGG GAT CCA ACC CTG AAA GCC CAG TTC T

Finally, competent E. Coli cells were purchased from New England Biolabs (NEB; Ipswich, MA) and a nickel-containing HIS-trap column for purification was purchased from GE Healthcare (Uppsala, Sweden).

3.4.2 Hemoglobin

Hemoglobin, like myoglobin, also represents a classic, well-studied heme protein that continues to generate new research and ideas about its mechanism of allostery.[37] It is known even colloquially for its role in carrying O_2 in the bloodstream of mammals and delivering it to cells in respiration. The quaternary structure is made up of four subunits, two *alpha*- and two *beta*-subunits with individual sequences, to form a "dimer of dimers." Each subunit's tertiary structure resembles myoglobin in that each is globular with a heme active site where diatomic

ligands bind, including CO and NO. This gives a ratio of four heme binding sites per hemoglobin molecule. See Fig. 3.4.
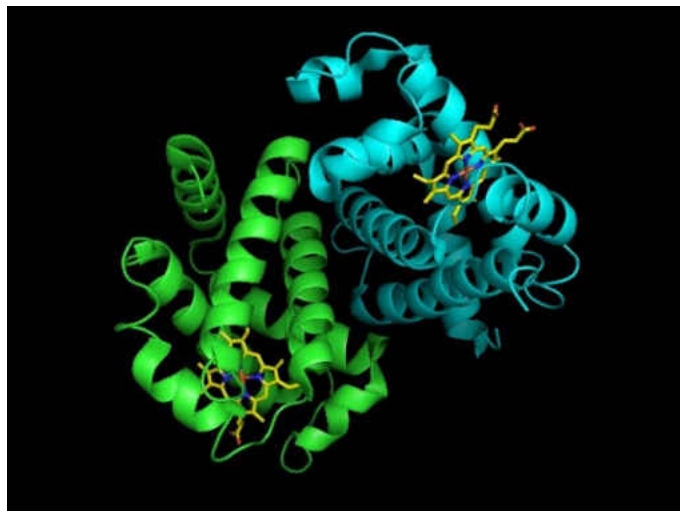
Figure 3.4. A dimer of α- and β-subunits of hemoglobin with heme highlighted in yellow (PDB code 1HHO) illustrated with PyMOL.

The cooperativity displayed in oxygen binding and release has a number of competing theories with the Monod–Wyman–Changeux (MCW) / Perutz model being the best known.[38] This model postulates that a global conformational change from the T (tense) state to the R (relaxed) state is induced by binding the first of four oxygen molecules. The conformational change increases the affinity of the other three binding sites for hemoglobin. The notion of conformation change being associated with a change in function suggests that the electrostatic structure change could be sensed and Stark experiments to measure internal electric field at individual subunits would be revealing. A description of how hemoglobin samples might be generated for such an experiment is given in Chapter 4.

3.4.3 Nitrophorin

A heme protein with distinct electrostatic structure and function from the related globins described above would make a good comparison for Stark spectroscopy experiments.

Nitrophorin is such a candidate. The protein is produced in the saliva of blood-sucking insects such as the kissing bug.[39] While in the lower pH saliva environment (pH 5-6) of the insect, the heme binds NO. However, once the enzyme has been injected into the bloodstream of a mammalian victim, the slightly higher pH ≈7 environment causes a conformational change that induces NO release. The destination for the NO ligand is soluble guanylate cyclase (sGC) where it begins the process of vasodilation in the victim, ultimately increasing the flow of blood for the feeding insect. In addition, nitrophorin also binds histamine to the heme once NO has been released, preventing an immune response in the victim and further enlarging the insect's meal volume.[40]

In terms of electrostatic structure, it has a tertiary structure that contains a single heme molecule inside a beta-barrel motif. The cavity is overall negatively charged, which would make a good comparison to the non-polar myoglobin heme cavity. In addition, the heme is ligated to the protein by a single covalent bond to a proximal histidine.[41]
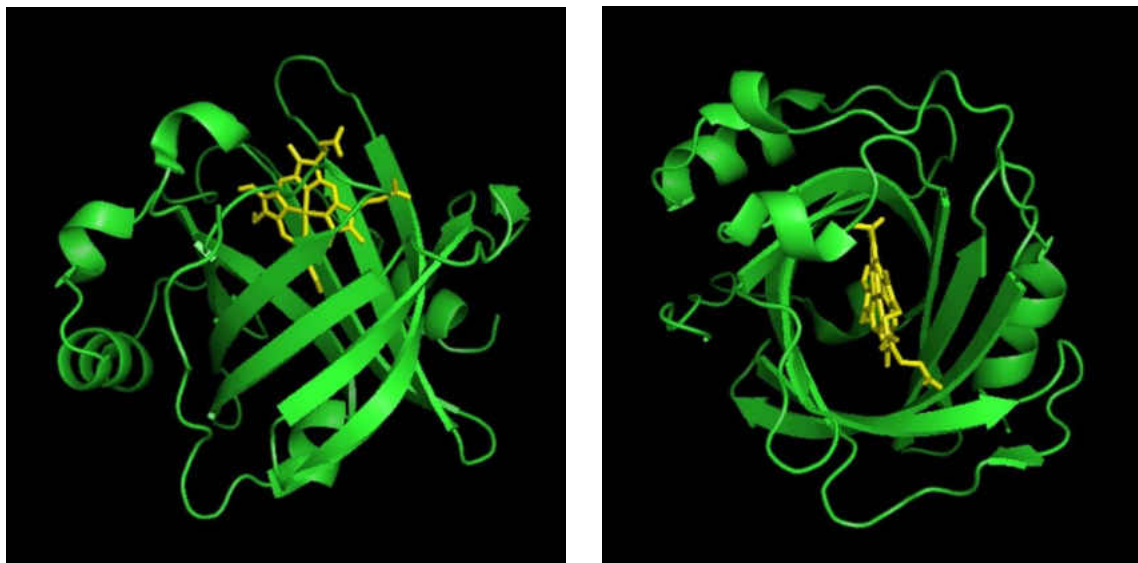
Figure 3.5. Nitrophorin's beta barrel structure with heme highlighted in yellow (PDB code 1ERX) illustrated with PyMOL.

3.5 Theoretical calculations of internal electric field in myoglobin

Using a point charge model, internal electric field for myoglobin was calculated. An in-house program (*proteinfield.c*) that evaluates Coulomb's law by plugging in values for the two variables for the program for each atom in the protein matrix used. The two variables according to Eq. 3.2 are *r,* distance from the point of measurement, and *q*, magnitude of charge. We will discuss each of these variables in turn.

3.5.1   Determining position of protein matrix atoms

Crystal structures provided an excellent resource for determining *r* in this type of calculation. An oxygenated human myoglobin was chosen and located in the PDB with code 1mbo. To prepare the file for use in a Coulomb's law calculation, the following steps were taken: 1) any non-polypeptide molecules were removed (discrete water molecules and inorganic

artifacts of crystallization) except for the cofactor (heme); 2) hydrogen atoms were added using the PyMOL program; 3) the structure was translated and rotated to position the origin at the heme iron and the x–y plane in the heme plane, using nitrogen pyrrole atoms as references of the x- and y-axes (see Appendix D for instructions on this step using the GaussView 5 program); 4) the heme cofactor was removed to leave only the 2496 protein matrix atoms. The internal electric field was then calculated at the origin of the coordinate system, which is where the heme iron was virtually positioned.

Examining the 1mbo file, it was found that four residues actually had two distinct positions with equal probability (i.e., two sets of coordinates with an occupancy of 0.50): VAL 13, LEU 86, LEU 89, and GLN 128. See visualization in Fig. 3.6. Individual myoglobin structures were generated to account for each of two alternate positions of these four residues ($4^2 = 16$ structures) and the internal field calculation was performed for each structure. In addition, mutants were created with the PyMOL program at two residue positions that are part of the heme cavity and deemed influential in terms of ligand discrimination based on Table 1 in Ref. [26]: HIS 64 (distal histidine) and LEU 29.
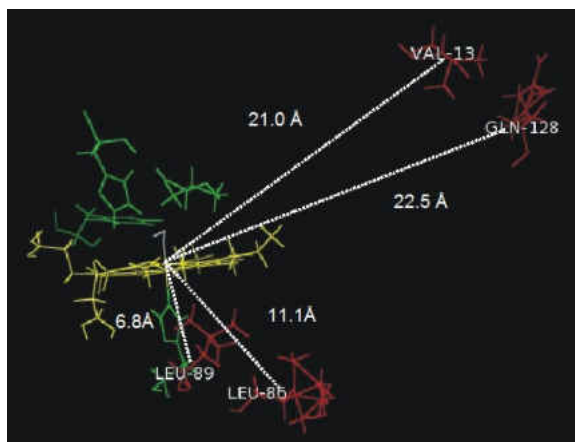
Figure 3.6. Four residues with alternate positions according to the 1mbo PDB file are highlighted in red. Distance to the heme iron is labeled for each. Heme is highlighted in yellow. Green residues are other heme cavity residues that are considered influential per the distance to the heme iron.

Because Coulomb's law states that point charges lose influence with distance scaling as $1/r^2$, an analysis was performed to determine whether there is a "cutoff radius" in the myoglobin matrix where it can be expected that most of the internal electric field is generated at the heme iron. This information is useful for determining where the most impactful mutations might lie as well as for methods like QM/MM (quantum-mechanics/molecular mechanics; a.k.a. "ONIOM")[42] in which proteins are modeled explicitly near a point of interest (e.g., active site) and less explicitly farther away. This exercise assumed equal charge of every atom, which, of course, is not physical but allows for this general partition into "likely to be influential" and "not likely to be influential." Using the rotated and protonated 1mbo coordinates, the distance to the origin, r, was calculated for each atom using the Pythagorean theorem, $r = \sqrt{(x^2 + y^2 + z^2)}$. Then the $1/r^2$ term for each atom was calculated. Finally, each distance was normalized with respect to the closest atom (a nitrogen atom in HIS 93, the proximal histidine, that bonds covalently to the heme iron) to allow sorting by distance and to assign the percentage of influence. Conveniently it was found that atoms that lie within 10 Å of the heme iron likely make up about 95% of the internal electric field. See Fig. 3.7. In addition, the positions of atoms in each individual residue were averaged and a list of the top twenty closest residues were determined and are given in Table 3.3 below.
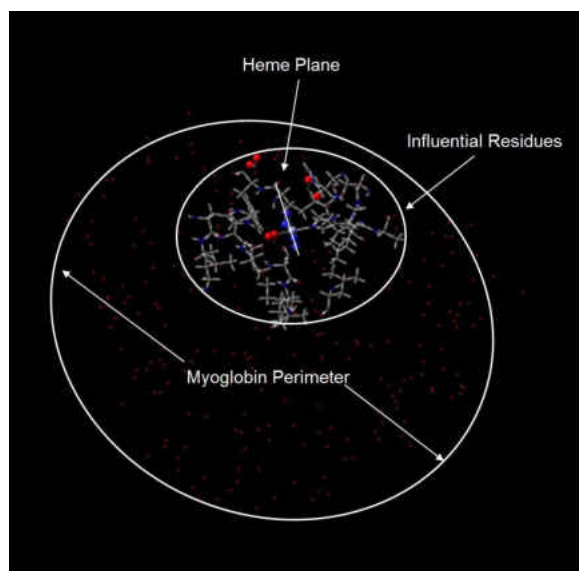
Figure. 3.7. A visualization of the volume of myoglobin that contains the most influential residues as given by Table 3.3 with respect to the heme active site as compared to the full myoglobin perimeter.

Table 3.3. Top 17 closest residues to the heme iron in 1mbo. Highlighted residues indicate those considered as heme cavity residues.

Residue	Proximity to Fe (Å)	Residue	Proximity to Fe (Å)	Residue	Proximity to Fe (Å)
HIS 93	5.259	PHE 43	7.435	ALA 71	8.514
VAL 68	5.956	ILE 107	7.840	LYS 42	8.937
HIS 64	6.696	SER 92	7.893	GLY 65	9.239
HIS 97	6.768	THR 67	8.023	ALA 90	9.277
ILE 99	6.967	LEU 104	8.062	LEU 72	9.407
LEU 89	7.150	LEU 29	8.399		

3.5.2   Determining partial atomic charge in myoglobin

As mentioned in Section 3.2.1, q is determined in this model by assigning a partial atomic charge for each atom in the protein matrix, which will always be an ultimately arbitrary

44

assignment but can have logical constraints to the scheme. In general, for amino acids, the magnitude of charge is kept to ±1 for neutral atoms. Known ions in the protein can take on full charge. Also, the sum of partial atomic charges must add up to the overall charge on the protein. In the case of myoglobin, the net charge should sum to zero. Besides the method for determining distribution of charge, one can also take one of two approaches in terms of resolution: 1) calculate partial atomic charge for each atom *type*, e.g., terminal carbons, carbonyl oxygens, etc., that recur in various amino acids; 2) calculate partial atomic charge for each *individual* atom in the protein based on its unique surroundings, both bonded and in proximity in the tertiary structure. Both approaches were taken and compared in this exercise.

The first approach has been tabulated in several reports.[12c, 43] The method here is to use an empirical force field, (CNDO/2, AMBER, and CHARMM, respectively) to generate molecular parameters for each of the common amino acids and the polypeptide backbone, including bond lengths, bond angles as well as partial atomic charge. The results are reported structurally with each atom given an atom type that can be found across all of the residues and backbone, which can usually be found to have the same partial atomic charge value. For use in Coulomb's law, an initial step is to list all atom types available for the method (which can vary) and then to associate a partial atomic charge value. The CNDO/2 ("complete neglect of differential overlap") results, which were used in this work, are obtained with a self-consistent field method (SCF) that treats all valence electrons but also electron interactions. The method also was carried out in an atom-centered manner, i.e., all charge was collapsed onto the nucleus and not in a lone-pair or bond location, which makes it especially appropriate for this implementation. Interestingly, the authors also introduced an effective dielectric constant of about 4 to move the resulting

structures away from gas phase and into a protein crystal, assuming there is a constant value for any protein.

A second approach was to calculate values for each atom in the protein individually. This calculation was able to be performed with an *ab initio* method using a low-level basis set. (More details about *ab initio* calculations will be discussed in Chapter 8.) Using the Gaussian09 package, a Hartree-Fock (HF) calculation at both the STO-3G and 3-21G levels also performs a Mulliken population analysis[44] by default, which returns partial atomic charges for each atom in the starting structure. Based on the molecular orbital coefficients that the HF method generates, a density matrix is formed, which, in combination with the overlap matrix, gives a population matrix. The population matrix allows for calculation of electron density as a function of space. The partial atomic charge, then, is the difference between the nuclear charge and the electron density assigned to it. For Mulliken population analysis, the electron density is assigned to an atom by cutting a bond in half and partitioning electron density equally between the two atoms. The partitioning step is one of the aspects that differentiates one partial atomic charge scheme from another when molecular orbitals are used. (Other partial atomic charge schemes available in G09 include Natural Bond Orbital Analysis [NBO], CHelpG, Merz-Kollman-Singh [MKS], Atoms-in-Molecules [AIM][45].) The rotated 1mbo structure with added hydrogens described in the previous section was used for the calculation *without* additional optimization, which is justified in this case because experimental coordinates in a rigid structure were used and no additional solvation was included (i.e., the calculation was performed *in vacuo*).

3.5.3   Results

3.5.3.1 Alternate positions using the CNDO/2 scheme

Results for potential and the three spatial components of internal electric field at the heme iron are given in Table 3.4. for the rotated and protonated 1mbo with partial atom charges assigned based on the CNDO/2 scheme given in Ref. [12c]. Each structural version is defined in Table 3.5. The perpendicular component of the field (z) is the most variable with alternate positions. The position of LEU 29, which was determined to be the sixth closest residue to the heme iron, seems to be the most influential.

Table 3.4. Electrostatic potential and components of the internal electric field for 1mbo as calculated with Coulomb's law.

Version	Potential (V)	Electric Field x (V/cm)	Electric Field y (V/cm)	Electric Field z (V/cm)
A	2.3318	-5.0054E+07	-7.3946E+06	5.5097E+05
B	2.3324	-5.0099E+07	-7.4286E+06	4.1846E+05
C	2.3311	-5.0059E+07	-7.3937E+06	5.5853E+05
D	2.3332	-5.0084E+07	-7.4327E+06	4.1698E+05
E	2.3320	-5.0062E+07	-7.3903E+06	5.4221E+05
F	2.3317	-5.0057E+07	-7.3957E+06	5.5365E+05
G	2.3318	-5.0065E+07	-7.3914E+06	5.4488E+05
H	2.3330	-5.0087E+07	-7.4337E+06	4.1966E+05
I	2.3309	-5.0062E+07	-7.3948E+06	5.6121E+05
J	2.3333	-5.0091E+07	-7.4284E+06	4.0822E+05
K	2.3312	-5.0067E+07	-7.3895E+06	5.4977E+05
L	2.3325	-5.0089E+07	-7.4318E+06	4.2454E+05
M	2.3331	-5.0094E+07	-7.4295E+06	4.1090E+05
N	2.3326	-5.0097E+07	-7.4275E+06	4.1578E+05
O	2.3323	-5.0092E+07	-7.4329E+06	4.2722E+05
P	2.3310	-5.0070E+07	-7.3905E+06	5.5244E+05
% Deviation	±0.051456%	±0.028956%	±0.29818%	±15.781%
Average	2.3321	-5.0077E+07	-7.4116E+06	4.8471E+05
Range high	0.0012	1.4812E+04	2.2100E+04	7.6496E+04
Range low	0.0012	1.4188E+04	2.2100E+04	7.6494E+04

Table 3.5. Sixteen unique combinations of the two alternate positions, *a* and *b*, of four residues (visualized in Fig. 3.6) noted in the crystal structure of 1mbo defines sixteen distinct structures of myoglobin.

Version	VAL13	LEU86	LEU89	GLN128
A	a	a	a	a
B	b	b	b	b
C	a	a	a	b
D	a	a	b	a
E	a	b	a	a
F	b	a	a	a
G	b	b	a	a
H	b	a	b	a
I	b	a	a	b
J	a	b	b	a
K	a	b	a	b
L	a	a	b	b
M	b	b	b	a
N	a	b	b	b
O	b	a	b	b
P	b	b	a	b

3.5.3.2 Mutants

The results for potential and internal electric field in Table 3.6 should be compared to the results for Version A in Table 3.4. Again, the z component appears to be the most variable, which is not unexpected based on the positions of each of these residues hovering above the heme

49

ring. See visualizations in Fig. 3.8a–i of the mutated residues with respect to the heme ring; carbon is shown in green, oxygen in red, nitrogen in blue, hydrogen in white, and iron in orange.

Table 3.6. Potential and internal electric field for seven mutants of 1mbo using Version A structure as defined above. The absolute change in average position of each residue is also noted.

Version	Potential (V)	Electric Field x (V/cm)	Electric Field y (V/cm)	Electric Field z (V/cm)	Change in Distance (Å)
H64G	2.2028	-4.4758E+07	-8.4309E+06	-1.2985E+06	4.2
H64L	2.2029	-4.4758E+07	-8.4240E+06	-1.1629E+06	3.6
H64A	2.2064	-4.4859E+07	-8.3929E+06	-1.1546E+06	3.7
H64I	2.2054	-4.4836E+07	-8.3799E+06	-1.1478E+06	2.1
H64V	2.2082	-4.4911E+07	-8.3993E+06	-1.1712E+06	2.2
L29I	2.1354	-4.2665E+07	-8.6580E+06	-1.0937E+06	0.9 - 1.1
L29A	2.1356	-4.2660E+07	-8.6351E+06	-1.0671E+06	2.4
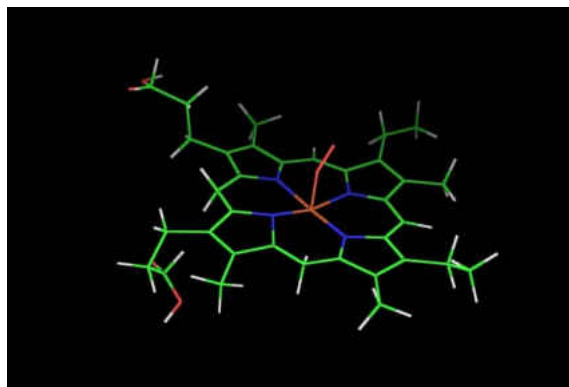
Figure 3.8a. Oxygenated heme.
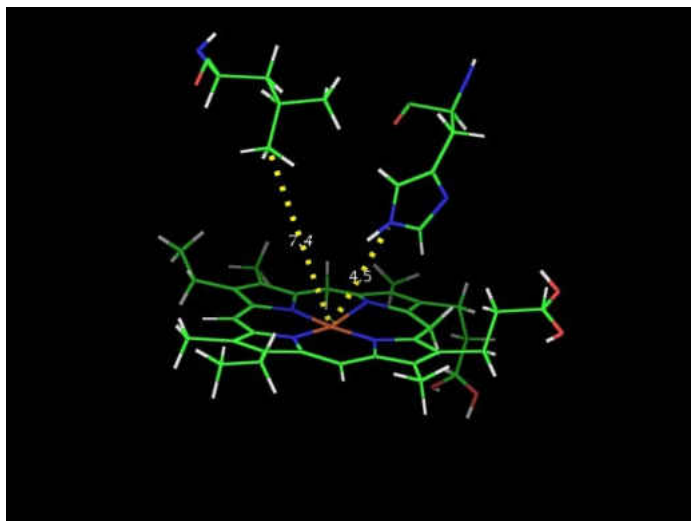
Figure 3.8b. Wild type 1mbo with HIS64 (ring structure at 4.5 Ẵ from heme iron) and LEU29 (branched structure at 7.4 Ẵ from heme iron).
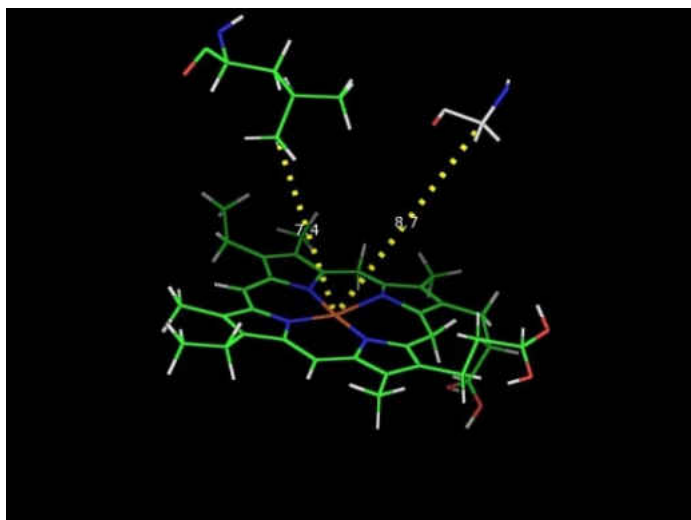
Figure 3.8c. Mutant H64G.
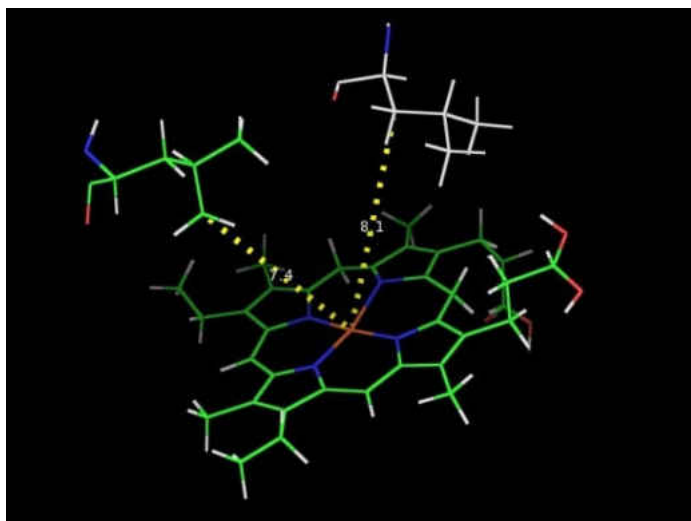
Figure 3.8d. Mutant H64L.
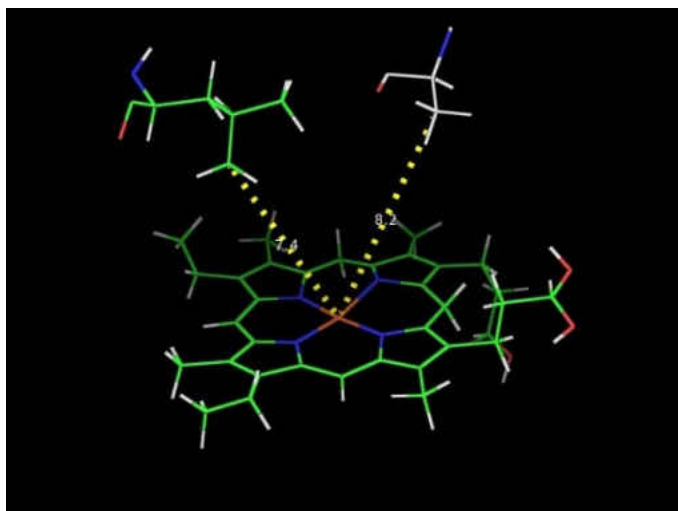
Figure 3.8e. Mutant H64A.
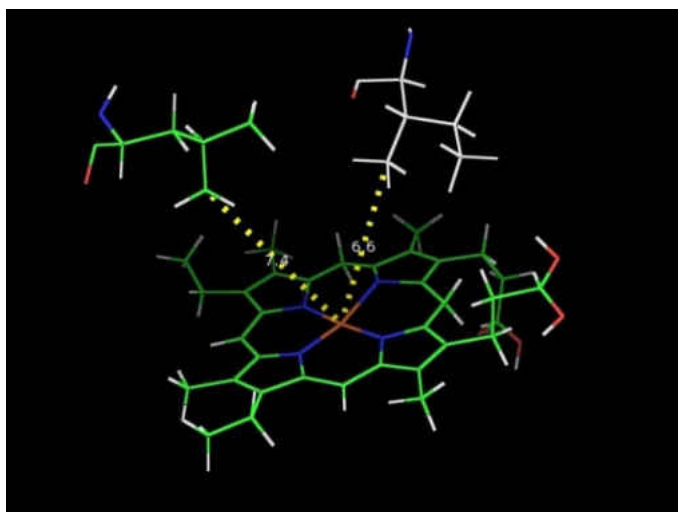
Figure 3.8f. Mutant H64I.
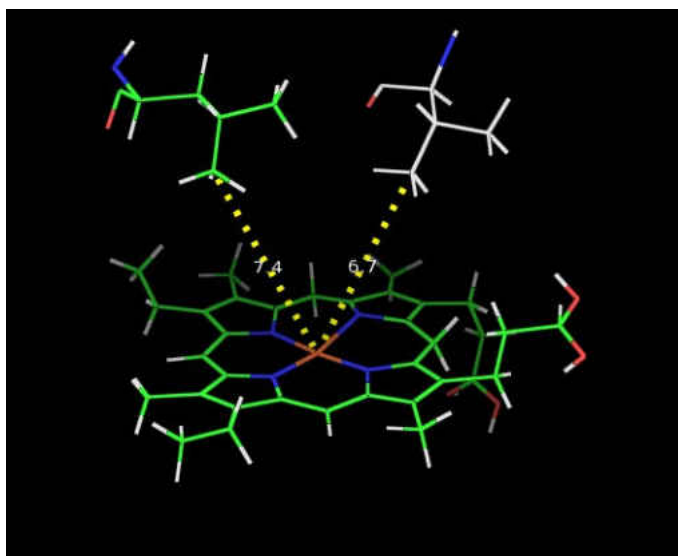
Figure 3.8g. Mutant H64V.
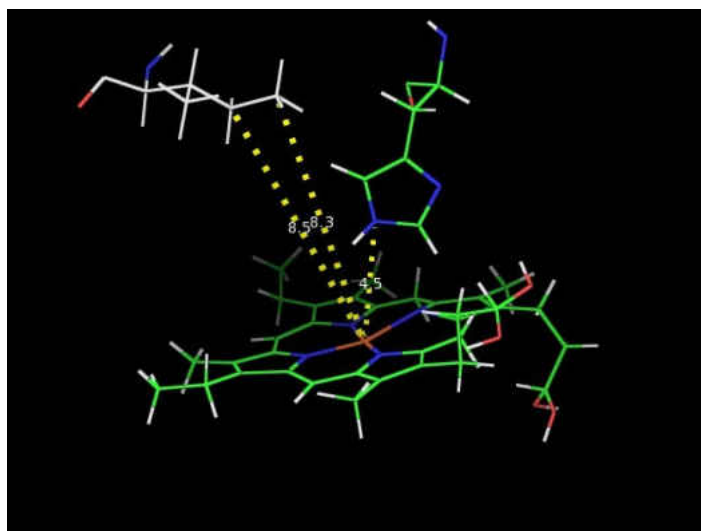
Figure 3.8h. Mutant L29I.
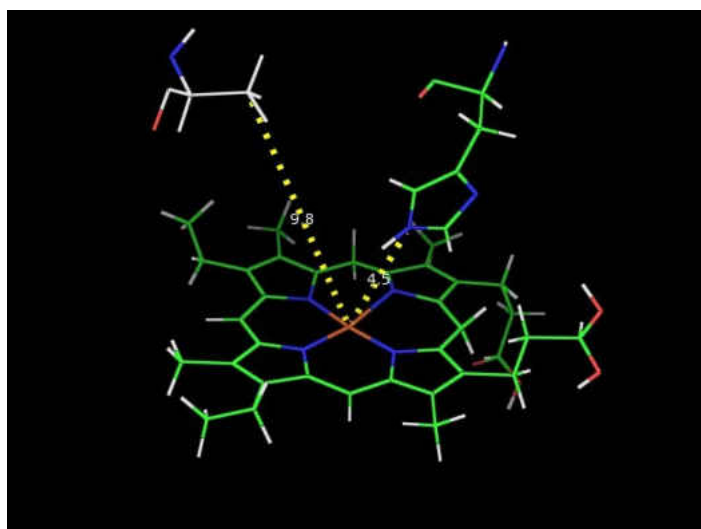
Figure 3.8i. Mutant L29A.

3.5.3.3 Results using *ab initio* partial atomic charge scheme

Presented in Table 3.7 are results for internal electric field components compared for rotated and protonated 1mbo (Version A) comparing partial atomic charge schemes calculated with the Hartree-Fock *ab initio* method (two basis sets, STO-3G and 3-21G) to the Table 3.4 Version A. The differences are large, one to two orders of magnitude. This is reflected in Table 3.8, which compares the partial atomic charge values generated by each method. The HF results are often ten times larger for the valine example used than the CNDO/2 values. CNDO/2 scheme

treated all VAL residues as identical, regardless of placement in the protein, which, in general,

assigns partial atomic charge values that are much, much lower than the *ab initio* values.

Gaussian09/HF values are Mulliken population analysis values calculated on the rotated and

protonated 1mbo structure for 8 individual valine residues. Note that VAL 1 is the protein's N-

terminus for which partial atomic charges are, at times, quite distinct from interior residues. Also

note that VAL 68 is considered a heme cavity residue and the second-closest residue to the heme

iron. A structure of valine with atom types is given in Fig. 3.9.

Table. 3.7. Internal electric field in myoglobin calculated with three different partial atomic

charge schemes.

Method	Electric Field x (V/cm)	Electric Field y (V/cm)	Electric Field z (V/cm)
HF/STO-3G	9.2458E+07	1.6118E+09	1.5125E+09
HF/3-21G	3.1568E+08	5.8422E+09	3.9284E+09
CNDO/2[12c]	1.8089E+07	1.7725E+07	4.3579E+07
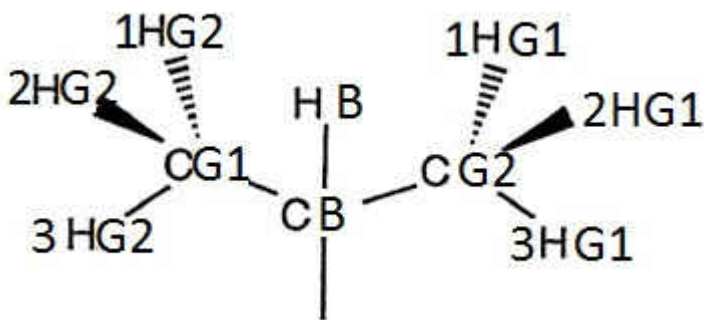
Figure **3.9**. Valine with atom types labeled as used in Table 3.8.

Table 3.8. Comparison of partial atomic charge values for valine per atom type and atom label between CNDO/2 and HF/3-21G

results. Atom labels are illustrated in Fig. 3.9; the remaining atom label that do not appear in the illustration are peptide backbone

residues.

Atom Type	Atom Label	CNDO/2 (1975)	VAL 1 – N-terminus (Gaussian09, HF)	VAL 10 (Gaussian09, HF)	VAL 13 (Gaussian09, HF)	VAL 17 (Gaussian09, HF)	VAL 21 (Gaussian09, HF)	VAL 66 (Gaussian09, HF)	VAL 68 (Gaussian09, HF)	VAL 114 (Gaussian09, HF)	Average for HF
C	CG2	-0.072	-0.572	-0.594	-0.581	-0.586	-0.561	-0.591	-0.589	-0.584	-0.582
C	CG1	-0.072	-0.558	-0.582	-0.603	-0.575	-0.579	-0.561	-0.586	-0.586	-0.579
C	CB	0.008	-0.360	-0.342	-0.366	-0.367	-0.346	-0.357	-0.367	-0.326	-0.354
C	CA	0.064	-0.158	-0.065	-0.044	-0.035	-0.063	-0.036	-0.072	-0.060	-0.067
C	C	0.450	0.936	0.959	0.951	0.928	1.007	0.963	0.985	0.945	0.959
H	HB	0.016	0.310	0.248	0.285	0.291	0.203	0.260	0.218	0.250	0.258
H	HA	0.020	0.324	0.277	0.239	0.278	0.310	0.285	0.305	0.257	0.284
H	3HG2	0.025	0.266	0.246	0.185	0.218	0.228	0.252	0.221	0.220	0.229
H	3HG1	0.025	0.149	0.233	0.233	0.205	0.230	0.213	0.278	0.228	0.221
H	2HG2	0.025	0.145	0.210	0.233	0.208	0.226	0.192	0.274	0.181	0.209
H	2HG1	0.025	0.238	0.199	0.246	0.192	0.208	0.213	0.189	0.204	0.211
H	1HG2	0.025	0.272	0.206	0.198	0.203	0.216	0.188	0.199	0.226	0.213
H	1HG1	0.025	0.273	0.206	0.228	0.216	0.224	0.195	0.207	0.228	0.222

H	1H	0.176	0.452	0.420	0.417	0.430	0.403	0.419	0.451	0.442	0.429
H	2H	0.176	0.439								
H	3H	0.176	0.459								
N	N	-0.356	-0.850	-0.938	-0.921	-0.924	-0.910	-0.931	-0.946	-0.926	-0.918
O	O	-0.384	-0.655	-0.693	-0.710	-0.725	-0.699	-0.741	-0.709	-0.714	-0.706

3.6 Conclusions

Evolving a viewpoint from individual electrostatic residues, surfaces, or smaller areas to view a protein matrix as the source of a global field has options for quantification (Coulomb's law vs. a bulk dielectric property). The effect of different environments on the electronic structure of porphyrin is briefly explored in Chapter 8, and the key to connecting net electric field to function relies on such a viewpoint. Along with this global viewpoint of the protein matrix is that of a glassy solvent with a constantly dynamic structure. The effect of global changes due to pH or temperature on net electric field may also be a useful viewpoint in beginning to model the changes in function due such perturbations and, depending on the sensitivity of the method, should be able to be sensed with the Stark spectroscopy method for the measurement of internal electric field discussed in Chapter 6.

Quantification of net electric field in proteins with Coulomb's law is a simplistic approach for determining the value but yielded some interesting insights. Because each scheme for calculating partial atomic charge produces widely varied results when summed for internal electric field values, the usefulness of the exercise is probably not in the absolute value it produces but in the relative values. For example, the insight that slight position changes of individual residues can alter the field values so dramatically (as in Table 3.4) is a set of relative values that provides some real physical conclusions. Or using the calculation to predict mutations that would be highly impactful on the net electric field of the protein matrix would be another way to avoid the variability of the partial atomic charge scheme in this application

3.7 References

1.   (a) Honig, B.; Nicholls, A., Classical electrostatics in biology and chemistry. *Science* **1995,** *268* (5214), 1144-1149; (b) Matthew, J. B., Electrostatic Effects in Proteins. *Annu. Rev. Biophys. Biophys. Chem.* **1985,** *14*, 387-417; (c) Warshel, A.; Russell, S. T., Calculations of electrostatic interactions in biological systems and in solutions. *Quarterly Reviews of Biophysics* **1984,** *17* (03), 283-422; (d) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M., Electrostatic Basis for Enzyme Catalysis. *Chemical Reviews* **2006,** *106* (8), 3210-3235.

2.   DeLano, W. L., *The PyMOL Molecular Graphics System*. http://www.pymol.org ed.; DeLano Scientific: San Carlos, CA, USA, 2002.

3.   Wagie, H. E.; Geissinger, P., Hole-Burning Spectroscopy as a Probe of Nano-environments and Processes in Biomolecules: A Review. *Applied Spectroscopy* **2012,** *66* (6), 609-627.

4.   Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W., Modeling electrostatic effects in proteins. *Biochimica et Biophysica Acta (BBA) - Proteins &; Proteomics* **2006,** *1764* (11), 1647-1676.

5.   Tanford, C.; Kirkwood, J. G., Theory of Protein Titration Curves. I. General Equations for Impenetrable Spheres. *J. Am. Chem. Soc.* **1957,** *79* (20), 5333-5339.

6.   Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **1976,** *103* (2), 227-249.

7.   (a) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences* **2001,** *98* (18), 10037-10041; (b) APBS & PDB2PQR download. (accessed 20 July 2015).

8.   (a) Nicholls, A.; Honig, B., A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *Journal of Computational Chemistry* **1991,** *12* (4), 435-445; (b) Li, L.; Li, C.; Sarkar, S.; Zhang, J.; Witham, S.; Zhang, Z.; Wang, L.; Smith, N.; Petukh, M.; Alexov, E., DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophysics* **2012,** *5* (9); (c) DelPhi Software download. (accessed 20 July 2015).

9.   Laberge, M., Intrinsic Protein Electric Fields: Basic Non-Covalent Interactions and Relationship to Protein-Induced Stark Effects. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1998,** *1386* (2), 305-330.

10.   Simonson, T.; Archontis, G.; Karplus, M., Continuum Treatment of Long-Range Interactions in Free Energy Calculations. Application to Protein Ligand Binding. *The Journal of Physical Chemistry B* **1997,** *101* (41), 8349-8362.

11.   Geissinger, P., Kohler, Bryan E., Woehl, Jorg C., Electric Field and Structure in the Myoglobin Heme Pocket. *J. Phys. Chem.* **1995,** *99*, 16527-16529.

12.   (a) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Kenneth M. Merz, J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995,** *117*, 5179-5197; (b) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P., A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* **2003,** *24* (16), 1999-2012; (c) Momany, F. A.; McGuire, R. F.; Burgess, R. W.; Scheraga, H. A., Energy Parameters in Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids. *Journal of Physical Chemistry* **1975,** *79* (22), 2361-2381.

13.   Johnson, B. G.; Gill, P. M. W.; Pople, J. A.; Fox, D. J., Computing molecular electrostatic potentials with the PRISM algorithm. *Chemical Physics Letters* **1993,** *206* (1-4), 239-246.

14.   (a) Laberge, M.; Köhler, M.; Vanderkooi, J. M.; Friedrich, J., Sampling Field Heterogeneity at the Heme of c-Type Cytochromes by Spectral Hole Burning Spectroscopy and Electrostatic Calculations. *Biophys. J.* **1999,** *77* (6), 3293-3304; (b) Manas, E. S.; Vanderkooi, J. M.; Sharp, K. A., The Effects of Protein Environment on the Low Temperature Electronic Spectroscopy of Cytochrome c and Microperoxidase-11. *The Journal of Physical Chemistry B* **1999,** *103* (30), 6334-6348; (c) Manas, E. S.; Wright, W. W.; Sharp, K. A.; Friedrich, J.; Vanderkooi, J. M., The Influence of Protein Environment on the Low Temperature Electronic Spectroscopy of Zn-Substituted Cytochrome c. *The Journal of Physical Chemistry B* **2000,** *104* (29), 6932-6941.

15.   (a) Scorciapino, M. A.; Robertazzi, A.; Casu, M.; Ruggerone, P.; Ceccarelli, M., Breathing Motions of a Respiratory Protein Revealed by Molecular Dynamics Simulations. *Journal of the American Chemical Society* **2009,** *131* (33), 11825-11832; (b) Tomita, A.; Sato, T.; Ichiyanagi, K.; Nozawa, S.; Ichikawa, H.; Chollet, M.; Kawai, F.; Park, S.-Y.; Tsuduki, T.; Yamato, T.; Koshihara, S.-y.; Adachi, S.-i., Visualizing breathing motion of internal cavities in concert with ligand migration in myoglobin. *Proceedings of the National Academy of Sciences* **2009,** *106* (8), 2612-2616.

16.   (a) Davis, M. E.; McCammon, J. A., Electrostatics in biomolecular structure and dynamics. *Chemical Reviews* **1990,** *90* (3), 509-521; (b) King, G.; Lee, F. S.; Warshel, A.,

Microscopic Simulations of Macroscopic Dielectric Constants of Solvated Proteins. *J. Chem. Phys.* **1991,** *95* (6), 4366-4377.

17.   Schutz, C. N.; Warshel, A., What Are the Dielectric "Constants" of Proteins and How to Validate Electrostatic Models? *Proteins: Struct., Funct., Bioinf.* **2001,** *44* (4), 400-417.

18.   (a) Gafert, J.; Friedrich, J.; Vanderkooi, J. M.; Fidy, J., Correlation between Protein Conformation and Prosthetic Group Configuration as Tested by pH Effects: A Hole-Burning Study on Mesoporphyrin-IX-Substituted Horseradish Peroxidase. *J. Phys. Chem.* **1994,** *98* (8), 2210-2214; (b) Gafert, J.; Pschierer, H.; Friedrich, J., Proteins and Glasses: A Relaxation Study in the Millikelvin Range. *Phys. Rev. Lett.* **1995,** *74* (18), 3704-3707; (c) Gafert, J.; Friedrich, J.; Vanderkooi, J. M.; Fidy, J., Structural Changes and Internal Fields in Proteins: A Hole-Burning Stark Effect Study of Horseradish Peroxidase. *J. Phys. Chem.* **1995,** *99* (15), 5223-5227.

19.   (a) Ambrose, W. P.; Moerner, W. E., Fluorescence spectroscopy and spectral diffusion of single impurity molecules in a crystal. *Nature* **1991,** *349* (6306), 225-227; (b) Tamarat, P.; Maali, A.; Lounis, B.; Orrit, M., Ten Years of Single-Molecule Spectroscopy. *J. Phys. Chem. A* **1999,** *104* (1), 1-16; (c) Moerner, W. E., Examining Nanoenvironments in Solids on the Scale of a Single, Isolated Impurity Molecule. *Science* **1994,** *265* (5168), 46-53; (d) Moerner, W. E., A Dozen Years of Single-Molecule Spectroscopy in Physics, Chemistry, and Biophysics. *J. Phys. Chem. B* **2002,** *106* (5), 910-927.

20.   Ponkratov, V. V.; Friedrich, J.; Vanderkooi, J. M., Hole Burning Experiments with Proteins: Relaxations, Fluctuations and Glass-Like Features. *J. Non-Cryst. Solids* **2006,** *352* (42–49), 4379-4386.

21.   Anderson, P. W.; Halperin, B. I.; Varma, C. M., Anomalous Low-Temperature Thermal Properties of Glasses and Spin Glasses. *Philosophical Magazine* **1972,** *25* (1), 1-9.

22.   Friedrich, J.; Haarer, D., Photochemical Hole Burning: A Spectroscopic Study of Relaxation Processes in Polymers and Glasses. *Angew. Chem., Int. Ed. Engl.* **1984,** *23* (2), 113-140.

23.   Storkel, U.; Creemers, T. M. H.; den Hartog, F. T. H.; Volker, S., Glass Versus Protein Dynamics at Low Temperature Studied by Time-Resolved Spectral Hole Burning. *J. Lumin.* **1998,** *76-77* (0), 327-330.

24.   Frauenfelder, H.; McMahon, B. H.; Fenimore, P. W., Myoglobin: The hydrogen atom of biology and a paradigm of complexity. *Proceedings of the National Academy of Sciences* **2003,** *100* (15), 8615-8617.

25.   Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958,** *181* (4610), 662-666.

26.   Springer, B. A.; Sligar, S. G.; Olson, J. S.; Phillips, J., George N., Mechanisms of Ligand Recognition in Myoglobin. *Chemical Reviews* **1994,** *94*, 699-714.

27.   Phillips, G. N.; Teodoro, M. L.; Li, T.; Smith, B.; Olson, J. S., Bound CO Is A Molecular Probe of Electrostatic Potential in the Distal Pocket of Myoglobin. *The Journal of Physical Chemistry B* **1999,** *103* (42), 8817-8829.

28.   Plattner, N.; Meuwly, M., Quantifying the Importance of Protein Conformation on Ligand Migration in Myoglobin. *Biophysical journal* **2012,** *102* (2), 333-341.

29.   (a) Mansy, S. S.; Olson, J. S.; Gonzalez, G.; Gilles-Gonzalez, M. A., Imidazole Is a Sensitive Probe of Steric Hindrance in the Distal Pockets of Oxygen-Binding Heme Proteins. *Biochemistry* **1998,** *37* (36), 12452-12457; (b) Quillin, M. L.; Arduini, R. M.; Olson, J. S.; Phillips Jr, G. N., High-Resolution Crystal Structures of Distal Histidine Mutants of Sperm Whale Myoglobin. *Journal of Molecular Biology* **1993,** *234* (1), 140-155; (c) Yi, J.; Heinecke, J.; Tan, H.; Ford, P. C.; Richter-Addo, G. B., The Distal Pocket Histidine Residue in Horse Heart Myoglobin Directs the O-Binding Mode of Nitrite to the Heme Iron. *Journal of the American Chemical Society* **2009,** *131* (50), 18119-18128.

30.   (a) Gunner, M. R.; Alexov, E.; Torres, E.; Lipovaca, S., The importance of the protein in controlling the electrochemistry of heme metalloproteins: methods of calculation and analysis. *Journal of Biological Inorganic Chemistry* **1997,** *2* (1), 126-134; (b) Kassner, R. J., Effects of Nonpolar Environments on the Redox Potentials of Heme Complexes. *Proceedings of the National Academy of Sciences* **1972,** *69* (8), 2263-2267; (c) Swartz, P. D.; Ichiye, T., Protein contributions to redox potentials of homologous rubredoxins: an energy minimization study. *Biophysical journal* **1997,** *73* (5), 2733-41.

31.   A. Shelnutt, J.; Song, X.-Z.; Ma, J.-G.; Jia, S.-L.; Jentzen, W.; J. Medforth, C., Nonplanar porphyrins and their significance in proteins. *Chemical Society Reviews* **1998,** *27* (1), 31-42.

32.   Muller, J. D.; McMahon, B. H.; Chien, E. Y. T.; Sligar, S. G.; Ulrich Nienhaus, G., Connection between the Taxonomic Substates and Protonation of Histidines 64 and 97 in Carbonmonoxy Myoglobin. *Biophysical Journal* **1999,** *77* (2), 1036-1051.

33.   Bhattacharyya, L.; Rohrer, J. S., Appendix 2. In *Applications of Ion Chromatography for Pharmaceutical and Biological Products*, John Wiley & Sons, Inc.: 2012.

34.   Chuy, S.; Bell, L. N., Buffer pH and pKa values as affected by added glycerol and sucrose. *Food Research International* **2006,** *39* (3), 342-348.

35.   Hildebrand, D. P.; Burk, D. L.; Maurus, R.; Ferrer, J. C.; Brayer, G. D.; Mauk, A. G., The proximal ligand variant His93Tyr of horse heart myoglobin. *Biochemistry* **1995,** *34* (6), 1997-2005.

36.   Maurus, R.; Bogumil, R.; Luo, Y.; Tang, H. L.; Smith, M.; Mauk, A. G.; Brayer, G. D., Structural characterization of heme ligation in the His64-->Tyr variant of myoglobin. *Journal of Biological Chemistry* **1994,** *269* (17), 12606-10.

37.   (a) Yuan, Y.; Tam, M. F.; Simplaceanu, V.; Ho, C., New Look at Hemoglobin Allostery. *Chemical Reviews* **2015,** *115* (4), 1702-1724; (b) Eaton, W. A.; Henry, E. R.; Hofrichter, J.; Mozzarelli, A., Is cooperative oxygen binding by hemoglobin really understood? *Nature Structural Biology* **1999,** *6* (4), 351.

38.   (a) Monod, J.; Wyman, J.; Changeux, J.-P., On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* **1965,** *12* (1), 88-118; (b) Perutz, M. F., Stereochemistry of Cooperative Effects in Haemoglobin: Haem-Haem Interaction and the Problem of Allostery. *Nature* **1970,** *228* (5273), 726-734.

39.   (a) Knipp, M.; He, C., Nitrophorins: Nitrite disproportionation reaction and other novel functionalities of insect heme-based nitric oxide transport proteins. *IUBMB Life 63* (5), 304-312; (b) Walker, F. A., Nitric oxide interaction with insect nitrophorins and thoughts on the electron configuration of the {FeNO}6 complex. *Journal of Inorganic Biochemistry* **2005,** *99* (1), 216-236.

40.   Ribeiro, J. M.; Walker, F. A., High affinity histamine-binding and antihistaminic activity of the salivary nitric oxide-carrying heme protein (nitrophorin) of Rhodnius prolixus. *The Journal of Experimental Medicine* **1994,** *180* (6), 2251-2257.

41.   (a) Andersen, J. F.; Weichsel, A.; Balfour, C. A.; Champagne, D. E.; Montfort, W. R., The crystal structure of nitrophorin 4 at 1.5 Ã resolution: transport of nitric oxide by a lipocalin-based heme protein. *Structure* **1998,** *6* (10), 1315-1327; (b) Ding, X. D.; Weichsel, A.; Andersen, J. F.; Shokhireva, T. K.; Balfour, C.; Pierik, A. J.; Averill, B. A.; Montfort, W. R.; Walker, F. A., Nitric Oxide Binding to the Ferri- and Ferroheme States of Nitrophorin 1, a Reversible NO-Binding Heme Protein from the Saliva of the Blood-Sucking Insect, Rhodnius prolixus. *Journal of the American Chemical Society* **1998,** *121* (1), 128-138; (c) Andersen, J. F.; Ding, X. D.; Balfour, C.; Shokhireva, T. K.; Champagne, D. E.; Walker, F. A.; Montfort, W. R., Kinetics and Equilibria in Ligand Binding by Nitrophorins 1-4: Evidence for Stabilization of a Nitric Oxide-Ferriheme Complex through a Ligand-Induced Conformational Trap. *Biochemistry* **2000,** *39* (33), 10118-10131; (d) Amoia, A. M.; Montfort, W. R., Apo-nitrophorin 4 at atomic resolution. *Protein Science* **2007,** *16* (9), 2076-2081.

42.   (a) Clemente, F. R.; Vreven, T.; Frisch, M. J., Getting the Most out of ONIOM: Guidelines and Pitfalls. In *Quantum Biochemistry*, Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 61-83; (b) Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K., The ONIOM Method and Its Applications. *Chemical Reviews* **2015,** *115* (12), 5678-5796.

43.   (a) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Salvatore Profeta, J.; Weiner, P., A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *Journal of the American Chemical Society* **1984,** *106*, 765-784; (b) A. D. MacKerell, J.; Bashford, D.; Bellott, M.; R. L. Dunbrack, J.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; W. E. Reiher, I.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M., Supplement to All-Atom Empirical Potential for Molecular Modeling & Dynamics Studies of Proteins. *Journal of Physical Chemistry B* **1998,** *102*, 3380-3386.

44.   Mulliken, R. S., Overlap Integrals and Chemical Binding. *Journal of the American Chemical Society* **1950,** *72* (10), 4493-4503.

45.   Bader, R. F. W., A quantum theory of molecular structure and its applications. *Chemical Reviews* **1991,** *91* (5), 893-928.

Chapter 4

Porphyrin Probes

4.1 Introduction: "molecular probes"

Porphyrins have a number of properties that recommend their usage as molecular probes in heme proteins. First, however, let us define "molecular probe." In this context, "guest–host" system is an appropriate term, with our molecular probe as the "guest" surrounded by the protein "host." The experimental technique then monitors the guest probe as it senses changes in the host environment. Employing UV-VIS spectroscopy here, we are monitoring the effect of the protein matrix (i.e., the tertiary structure of the polypeptide) on the electronic properties of the porphyrin probe, which is necessarily a chromophore with a conjugated structure. The protein matrix has both steric and electrostatic influences that the probe will sense and, in turn, report spectroscopically.

A number of studies have employed molecular probes in biological systems, which are interesting to compare and contrast this idea of using porphyrins as a spectroscopic probe:

Nitrile probes: Since about 2000, nitrile (C–N-based) probes have been investigated.[1] Here the vibrational stretch frequency is monitored with infrared spectroscopy as it responds to electrostatic environment. As this group is relatively simple to incorporate into a variety of proteins using amino acid substitution with artificial residues, the method is can be used at very specific locations in a variety of proteins that accept such mutations but will, in turn, only have the capability of reporting the electrostatic conditions over the very small area of that two-atom

probe. Investigated proteins include myoglobin mutants,[2] human aldose reductase,[3] and bacterial dihydrofolate reductase.[4]

Heme coordinating ligands: This includes diatomic ligands, carbon monoxide (C–O) and nitric oxide (N–O); and nitrogen-based heterocycles such as imidazole, nicotine, and histamine. These provide a number of tools for probing heme proteins. The probes sit at the heme cofactor in these proteins, which is often the active site, but only indirectly interact with the protein environment as its major interaction is with the heme itself.[5]

Chlorophyll probes in chlorosome antennae: Related in structure to porphyrins and, in turn, acting similarly as chromophores, chlorophyll in chlorosome antennae represent the active sites in these supramolecular structures. However, in contrast to probing the protein matrix, the chlorophyll exist in aggregates and as a molecular probe, likely reports interactions with other chlorophyll molecules.[6]

Thus, the approach we take using porphyrin as a full molecular probe has the advantage of sitting precisely where the heme active site resides with the same spatial relationship to the protein that influences it (Fig. 4.1). In addition, this approach considers the full surface area of the cofactor as affected by the net electrostatic environment, which means the electronic properties of the cofactor as a whole can be considered (e.g., dipole moment, partial atomic charge of the metal as well as bonded atoms, etc.). This is in contrast to an atomic resolution method developed as an extension of this molecular resolution method, which certainly has utility for systems where heterogeneity of the field is relevant.[7]
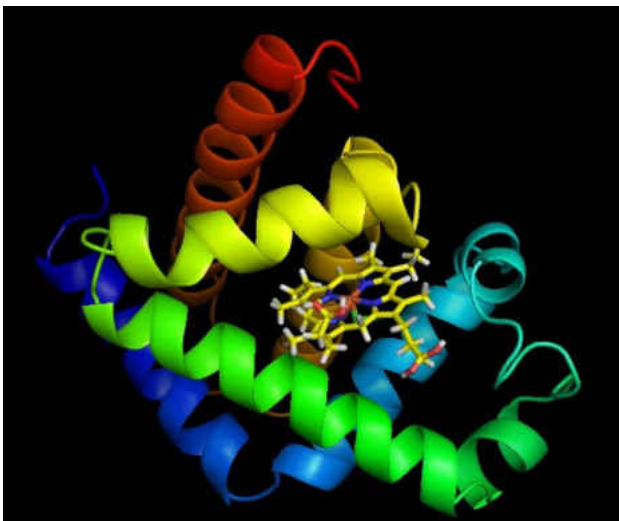
Figure 4.1. A ribbon structure of myoglobin highlighting the position of the heme cofactor in a stick structure.

What can we learn about the protein environment from molecular probes?[8] The locality of the probe is advantageous in these measurements. When placed in a position of importance, i.e., the active site, conditions in that functional space are described. One property is the physical free volume of the active site, as reflected by compressibility measurements around the probe. The variability of the probe's immediate environment is quantifiable by noting spectral diffusion and gives information about protein dynamics and the excursion across its multi-dimensional potential energy surface. Access to protein dynamics opens the door for insights into folding processes. The local electrostatic environment can be detected by changes to the molecular probe's electronic transition energies. In this work, it is the net electric field generated by the entire protein matrix that will be deduced from our spectral probe.

4.2 Molecular properties of porphyrin probes

Porphyrins are a diverse class of molecules that share a highly conjugated, nitrogen-containing heterocycle center, porphin. Species differ by type, number, and placement of substituents to the porphin center. In addition, porphyrins may be metallated, in which the four

pyrrole nitrogen atoms chelate a transition or rare-earth metal (Fig. 4.2a), or it may be "free-base", in which the pyrrole nitrogens individually bond hydrogen rather than a metal (Fig. 4.2b).
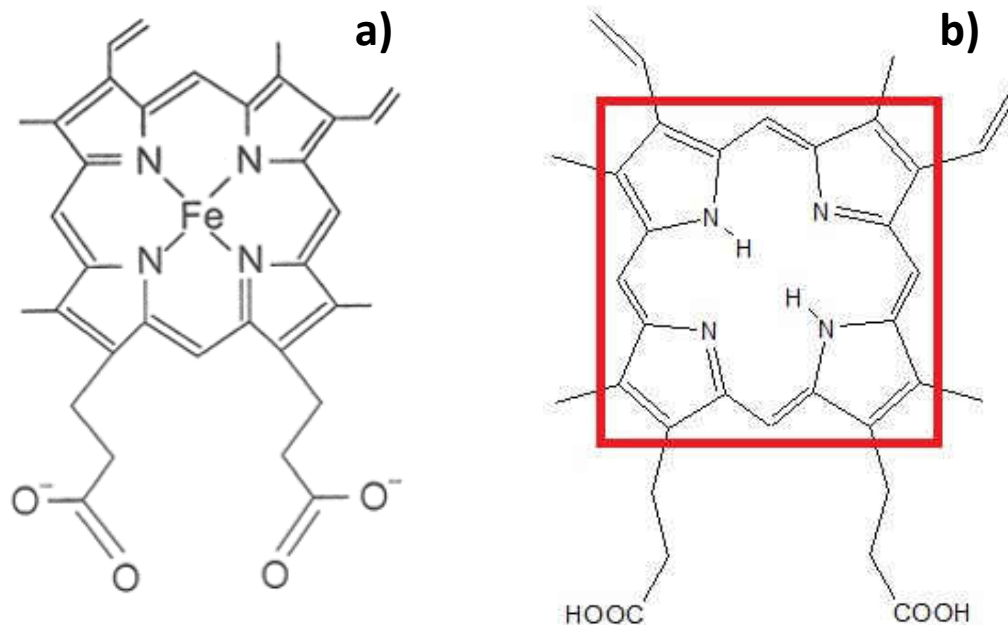
Figure 4.2. a) Iron-centered heme is the native cofactor in myoglobin and hemoglobin. The central iron atom may be replaced by various metals. b) Free-base protoporphyrin IX is a heme analog with no metal center. The center portion boxed in red is the structure of porphin.

This work considers free-base porphin and zinc porphin (containing no substituents, see portion of Fig. 4.2b boxed in red) as the *theoretical* molecular probes of interest. However, because we are interested in sensing heme protein environment, the *experimental* probes are heme analogs: free-base protoporphyrin IX and zinc protoporphyrin IX. In spectroscopic experiments, analog substitution (procedure described below) is necessary because iron-centered heme typically does not fluoresce, having a short lifetime because of efficient, radiationless decay. The analogs contain the same substituents: methyl, vinyl, and, most notably, two propionate groups. Thus, these substituents are considered part of the environment influencing the porphin probe. In both metallated and free-base protoporphyrin IX (PPIX), the

propionate groups have a pK$_a$ of 5.7, which causes each them to be deprotonated at physiological pH (pH ≈ 7) and each to carry a -1 charge. The two carboxylate ion oxygen atoms share the charge as it is delocalized across the resonance bonds, but crystal structures and optimized theoretical structures show each three-carbon chain to curl in opposite directions along the perpendicular axis to the planar porphin ring (Fig. 4.3a), placing the center of charge approximately 6–8 angstroms from the center of the ring. The close proximity of these highly charged groups makes them strong contributors to the net electric field sensed by the conjugated porphin probe. The electrostatic influence of these propionate groups likely has physiological significance,[9] but the aim of this work is to probe the protein matrix itself. To retain the important spatial features of heme-analog porphyrins but to remove any additional source of electric field, protoporphyrin IX dimethyl ester (PPIX DME) was also characterized. PPIX DME is structurally similar to PPIX but the charged carboxylate ion is instead capped with a methyl ester group (Fig. 4.3b).
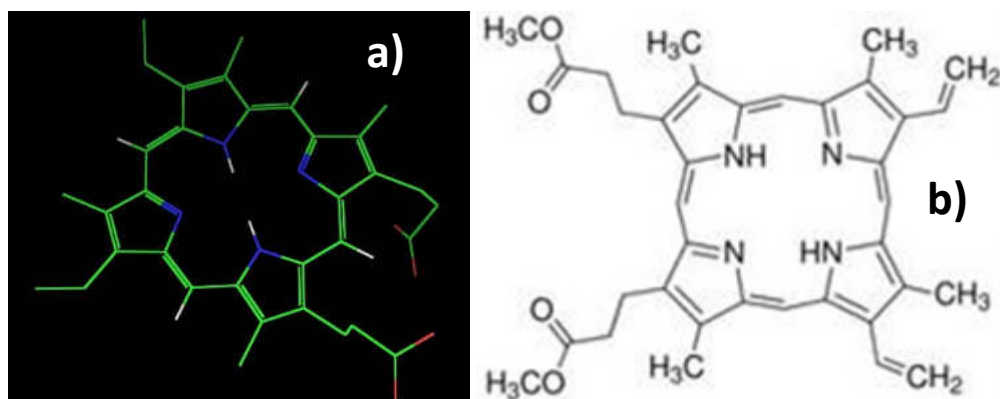
Figure 4.3. a) Another view of protoporphyrin IX with non-planar propionate groups above and below the ring plane. b) Protoporphyrin IX dimethyl ester has propionate groups capped with methyl groups.

For free-base PPIX, the inner hydrogens also have individual pK$_a$ values. Using the nomenclature of Falk,[10] the nitrogens that are effectively always protonated (pK$_1$ = pK$_2$ ≈ 16) are referred to as "pyrrole

nitrogens" and are arranged in a *trans* position from one another in porphin's most stable form. (The *cis* tautomer – and rapid interconversion between the two - has been explored for some porphyrins[11] but the *trans* form is generally accepted as the lowest energy[12] and will be used as the model here.) The other two nitrogens are referred to as "pyrrolenine nitrogens" and have $pK_3 \approx pK_4$ as high as 7.2 when the propionate side groups are charged.

Although the choice of which two nitrogen atoms are protonated "pyrrole" atoms in free-base porphin is arbitrary based on its high symmetry, however, in PPIX, closer examination reveals that the positions of the vinyl substituents renders this porphyrin non-symmetric (observe Fig. 4.2). This non-equivalent tautomerization of inner hydrogen atoms is the mechanism for spectral hole burning (discussed in Chapter 6) with free-base porphyrins. Spectroscopically, however, both free-base (D_{2h} symmetry) and metalloporphyrins (higher D_{4h} symmetry) have inversion-symmetry characteristics that are retained by the central porphin despite symmetry-breaking substituents.[13] In the context of a specific protein environment, it is precisely this symmetry that gives porphyrins their sensitivity as molecular probes. "Symmetry breaking," imposed on the porphyrin by steric or electrostatic factors inside of a protein,[14] gives rise to properties that are not, in a neutral environment, possessed by centrosymmetric molecules. For example, a centrosymmetric molecule theoretically has a difference dipole moment, $\Delta\vec{\mu}$, equal to zero. This quantity is the change in permanent dipole moment between the ground and first electronic excited state, which can be measured using Stark spectroscopy (discussed in Chapter 6).[15] Symmetry concepts will be used throughout the following spectroscopy discussions.

4.3 Methods

4.3.1 Preparation of apo-globins

In order to implement the porphyrins described above as molecular probes in heme proteins, a method to remove heme must first be employed. Two major, well-established methods are currently used to remove heme from myoglobin and hemoglobin: 1) the acid-acetone method[16] and 2) Teale's 2-butanone (methyl ethyl ketone) method.[16b, 17] Both methods were performed so that the proteins were removed of the heme cofactor as completely as possible from sample, but as heme is non-radiative, it is not necessarily a contaminant for use in fluorescence experiments.

Extracted proteins in lyophilized powder form were purchased from Sigma–Aldrich, Milwaukee, WI at a purity of: 1) myoglobin from equine skeletal muscle, 95-100% purity, "essentially salt free"; 2) equine hemoglobin.

Equine myoglobin is a single chain globular protein of 153 residues containing a single heme cofactor. Its molecular weight is approximately 17,500 Da. Solubility for myoglobin is about 20 mg/mL.

Equine hemoglobin is tetrameric protein, each subunit resembling myoglobin's globular structure containing a single heme for a total of four heme molecules per hemoglobin. Its molecular weight is approximately 64,500 Da. Solubility for hemoglobin is also about 20 mg/mL.

4.3.2 Acid–acetone method of heme removal

Typically, a $0.2 - 0.6$ mM aqueous solution of protein was prepared in cold ultrapure water. First, the heme protein is oxidized to assure a met-globin state (Fe^{3+}) and, consequently,

no bound ligands throughout the sample. This is accomplished with potassium ferricyanide ($K_3[Fe(CN)_6]$) added to an aqueous solution of protein in a ratio of two oxidizers per heme and allowed to stir in the cold and dark for several hours. This step is followed by exhaustive dialysis in cold (4 °C) ultrapure water to remove any trace of oxidizer. Acid–acetone was prepared by as 20%–2 M hydrochloric acid (HCl) in acetone[18] and cooled to −20 °C. Then, the acid–acetone solution was added to the protein solution in a 15-20 volume excess and vortexed for 5-10 min. The protein was then allowed to precipitate at -20 °C for 2-3 hours. The cloudy brown solution was then centrifuged in several microcentrifuge tubes at 4 °C at 13 200 rpm for 15 min. The brown supernatant (containing heme) was decanted from each tube and discarded, leaving a grayish pellet. Each pellet was washed twice by re-suspending in pure, - 20 °C acetone and spun down again in identical fashion. The resulting clean apo-protein pellet was then re-dissolved in cold ultrapure water and placed in cold ultrapure water for exhaustive dialysis.

4.3.3   Methyl ethyl ketone method

As with the acid–acetone method, the heme protein was first oxidized with $K_3[Fe(CN)_6]$. After dialysis, the aqueous protein solution was transferred to a beaker on ice with stirring. Cold 1-M HCl was added dropwise to the protein solution to a pH < 1.5. In a 4 °C walk-in cooler, the acidic protein solution was transferred to a glass separatory funnel along with cold 2-butanone (also known commonly as methyl ethyl ketone) in a 2:1 volume ratio. Then funnel was then capped and shaken vigorously and the layers allowed to separate (about 10 min). Once fully separated (which was simple to verify visually), the tan aqueous layer (containing protein) was dispensed through the stopcock into one container, while the dark brown organic layer was decanted into a waste container. This separation was repeated three more times but with a 1:1

volume ratio of organic to aqueous solvent. (Typically, the second extraction took much longer than the others (about 30 min) as it tended to form a large, emulsified, intermediate layer.) The final aqueous solution contained the apo-protein; apo-myoglobin was typically a bright yellow tint and apo-hemoglobin typically had a greenish tint.

The apo-protein solution was then transferred to a cold 1 L - 0.6 mM sodium bicarbonate ($NaHCO_3$) dialysis bath and repeated once. This step is intended to precipitate any remaining holo-protein[17b] but appeared to precipitate aggregate as well. A cold ultrapure water dialysis bath followed. Next, the apo-protein was placed in a 0.1 mM sodium ethylene diamine tetra-acetic acid (Na_2EDTA) solution to remove any heavy metals, which can destabilize the apo-protein over time. Finally, the pH was adjusted with a pH 7.3 sodium phosphate buffer dialysis bath.

The neutral apo-protein solution absorption spectrum was then taken (scanning λ = 250 – 700 nm) to 1) determine the completeness of the heme removal and 2) determine the concentration of the protein in solution. The presence of heme could be noted with a Soret peak (the most intense peak characteristic of porphyrins) around 410 nm. Concentration of protein was determined with Beer's law ($c = A/(\varepsilon \cdot l)$), where c = concentration in M, A = absorbance units, l = pathlength of cuvette, and ε = molar absorptivity coefficient in $mM^{-1} \cdot cm^{-1}$. For myoglobin, the molar absorptivity coefficient at 280 nm (ε_{280} = 15.8 $mM^{-1} \cdot cm^{-1}$). For hemoglobin, the average molar absorptivity coefficient for individual subunits (averaging alpha and beta) is 12.7 $mM^{-1} \cdot cm^{-1}$.

Apo-protein solutions were either immediately substituted or stored in apo-form for later substitution. If prepared for storage, the apo-protein solution was concentrated using a SpinX

centrifuge filter tube and then transferred to polypropylene microcentrifuge tubes in aliquots of < 1 mL for storage in a liquid nitrogen storage dewar.

4.3.4   Comparison of heme removal methods

In general, the methyl ethyl ketone extraction method proved the most convenient for myoglobin and hemoglobin heme removal. The visual confirmation of removal is clearer and an extra extraction step is simple to enact if the deep brown heme seems to linger in the aqueous layer. In addition, the method does not rely on precipitation so the yield of recovered protein appears to be higher, despite literature comments that the methods have equivalent yields. Finally, this method is said to be appropriate for unstable proteins (especially apo-globins), avoiding denaturation that accompanies precipitation.[16b]

4.4 Reconstitution of apo-globins with porphyrin probes

The essentially planar heme, once removed from globular proteins, can be replaced with a variety of planar, conjugated molecules, most importantly other porphyrins.[19] The goal of these studies is to preserve the protein matrix structure while improving the luminescence properties of the porphyrin probe, so closely related heme analogs were chosen for substitution.

For these high-resolution spectroscopic studies, the most disruptive contaminants are porphyrins that are not correctly substituted into a protein host. As a result, steps were taken to prevent and remove any porphyrins in free solution as well as porphyrins entangled in aggregated apo-protein. At the present time, the Brookhaven Protein Data Bank (www.rcsb.org) does not contain any myoglobin or hemoglobin substituted with free-base protoporphyrin IX, so a

"correct" substituted structure is not known, especially without a metal to ligate covalently to the distal histidine.

4.4.1 Substitution with free-base and zinc protoporphyrin IX[17b, 20]

Both free-base (>97% purity, MW = 562.66) and zinc PPIX (MW = 626.05) (Frontier Scientific, formerly Porphyrin Products, Logan, UT) were dissolved in 0.1 M NaOH solution (pH > 12) to assure that the molecules' propionate groups were negatively charged (and were consequently in their most soluble form). These porphyrin solutions were made up in 50 – 100 mL volumes at a concentration of 0.16 mM and kept wrapped in foil to avoid photobleaching as they were stored. For substitution, porphyrin and apo-protein were mixed in a ratio of one porphyrin to two heme binding sites to avoid unsubstituted porphyrin. Apo-protein was dissolved in a dilute buffered solution and the basic porphyrin solution added to it in a much smaller volume to avoid major pH changes. The combined solutions were allowed to stir in the cold and dark overnight. Porphyrin is spontaneously taken up by myoglobin and hemoglobin. Subsequent purification and concentration steps are described below.

4.4.2 Substitution with hydrophobic protoporphyrin IX dimethyl ester

For substitution with the electrostatically neutral PPIX DME, the porphyrin, not being water soluble, required a modified substitution procedure.[21] A concentrated, rose-pink 10-µM solution of iridescent, violet PPIX DME crystals (Frontier Scientific, Logan, UT, >97% purity, MW = 590.7113) in 1% v/v pyridine in methanol was created, heating briefly to avoid solvent loss. (Solubility is 12 µM in methanol.)[22] The substitution procedure was similar to the water-soluble porphyrins but the porphyrin–protein mixture needed to be kept <10% methanol by volume. As a result, additional concentration steps were required as described below.

74

4.4.3 Metallation of free-base porphyrins

To broaden options for proteins substituted with both metallo- and free-base porphyrins, a procedure to metallate metal-free porphyrins was explored.[10, 23] In particular, the neutral protoporphyrin IX dimethyl ester species did not have a commercially available metalloporphyrin analog. The following procedure successfully produced a zinc analog of PPIX DME, which was clearly identified spectroscopically by the collapse of four Q-bands characteristic of free-base porphyrins into two Q-bands characteristics of metalloporphyrins (see Chapter 5 for theoretical details and a spectrum). The procedure is generally useful with manganese, iron, cobalt, nickel, and copper derivatives reported along with zinc.

About 60 mg of PPIX DME was dissolved in 2 mL chloroform and refluxed for about 15 min. Meanwhile, 0.8819 g of zinc acetate was weighed out into a 25-mL roundbottom flask and dissolved in 10 mL glacial acetic acid. The zinc solution was stirred and heated under a condenser, but only partial dissolution was achieved (solubility of zinc acetate is 43 g / 100 mL) as the white inorganic powder was apparent. A 2.5-mL portion of the zinc slurry was added to the refluxing PPIX DME solution and allowed to continue reflux for two hours. No solid material was left at the end of this step. Then the reddish-purple metal and porphyrin solution was transferred to a separatory funnel and an equal amount of ultrapure water was added to wash. This wash step was repeated four times, adding chloroform to rinse the walls of porphyrin. The washed solution was transferred to a 50-mL roundbottom flask and rotoevaporated at 45 °C to dryness. The reddish-purple crystals were re-dissolved in a small amount of chloroform and poured onto a large watchglass for a final rinse with a 1:1 mixture of chloroform:methanol. Finally, the

metallated porphyrin was allowed to air dry. Subsequent use in protein substitution would benefit from further purification on a separation column.

4.4.4 Purification procedure

Original procedures for the preparation of heme-free myoglobin and hemoglobin and subsequent reconstitution with a porphyrin called for lengthy column separations. Following the final EDTA dialysis of apo-protein, anion-exchange chromatography at pH 8.3 (Tris-HCl buffer) was indicated to remove free heme and any remaining met-globin. After reconstitution, a gel filtration (size exclusion) column was used to remove any porphyrin not taken up by the apo-protein.

Both of these separation types are lengthy and require access to a high-performance liquid chromatography system outside of the immediate lab. Thus, exploration of alternative purification procedures was undertaken. First, thoughtful consideration of what constituted a "contaminant" was considered. The samples prepared for hole-burning Stark spectroscopy experiments required that the molecular probe reported a single environment type; therefore free solution (unsubstituted) porphyrin was to be avoided as well as incorrectly substituted porphyrin. Incorrect substitution might take the form of either misfolded, aggregated target protein or might be a protein other than the target protein that occurs as an impurity in the purchased native sample. Finally, salt may be a contaminant and can affect protein stability. Because photoexcited heme relaxes non-radiatively, it is not considered a contaminant for the purposes of luminescence experiments; however, enough heme must be removed to provide for a reasonable high concentration of substituted porphyrin. Based on this criteria, only a size-based

separation is necessary to ensure a "pure sample" of substituted proteins for the experiments described in Chapter 6.

If the initial native protein purity is questionable, an HPLC size-exclusion (e.g., HiPrep 26/60 Sephacryl S100 from GE Healthcare Bio-Sciences) column should be used before or after the substitution to assure a single protein type. In addition, soluble aggregate may also be addressed this way (though a typical preparation of substituted myoglobin did not appear to contain any appreciable amount of aggregate as Chapter 5).

To address the issue of salt or free solution porphyrin, a simpler approach can be taken because of the disparate sizes of protein and contaminant. Three types of columns were explored to substitute for the lengthy columns previously indicated: 1) solid-phase extraction (SPE) tubes (vacuum), SpinX centrifugal concentrator tubes (Corning, Inc., Corning, NY) and 3) Pierce polyacrylamide desalting spin columns (Thermo Scientific, Waltham, MA). The methods above were compared spectroscopically. The presence of protein was monitored with absorbance at 280 nm; the presence of porphyrin was monitored around 400 nm. Each phase, retained fraction vs. runoff, was compared per method to determine the effectiveness of the separation.

Three types of SPE tubes were investigated: 1) strong anion exchange (SAX), 2) strong cation exchange (SCX), and 3) retention of nonpolar organic molecules (C18). The tubes were attached to a vacuum flask to flow through the column with the goal of trapping contaminating free porphyrin and elute the target protein. Working with myoglobin, the protein's isoelectric point of pH ≈ 7.2 was the basis for setting up a pH gradient using phosphate buffer for the ion exchange tubes. The propionate pK_a of free protoporphyrin IX, 5.7, provided the boundary

between negatively charged and neutral forms of the molecule. The C18 tube was determined to not be effective for retaining the neutral form of protoporphyrin IX. Both ion-exchange tubes failed to separate the protein and porphyrin, except at the extreme ends of their respective pH ranges. It was concluded that the SPE columns did not meet the criteria of simplifying separation.

For a second comparison, a 20-mL SpinX centrifugal concentrator tube with a 5000 molecular-weight cutoff (MWCO) for its polyethersulfone membrane was used to retain protein with MW > 5000 Da while smaller molecules passed through the filter into the runoff. (Porphyrins used in this study have molecular weights less than 650 g / mol). The SpinX tubes were used with a bucket rotor at a maximum speed of 4.2 rpm. The retained portion contained evidence of substituted myoglobin (both peaks at 280 nm and at 400 nm), but the runoff did not show a spectral signature at all

Finally, a 0.7 mL desalting spin column with a 7000 MWCO, in which small molecules are retained in the resin bed while larger molecules pass through, was investigated. The column was loaded and spun at 6000 rpm in a microcentrifuge with a fixed angle rotor. Additional phosphate buffer and spin time eluted any retained molecules from the size-exclusion bed. The initial protein fraction revealed a substituted myoglobin, and the subsequent fractions revealed a smaller amount of protein. The final fractions revealed free porphyrin with no remaining protein, showing no peak at 280 nm but a peak at 400 nm.

It was concluded that the desalting column was an easy-to-use substitute for a size-exclusion column for the step of separating free solution porphyrin from substituted protein. A

gravity desalting column gave similar results with a larger, 10-mL volume (Pierce polyacrylamide desalting columns, 6000 MWCO, Thermo Scientific, Waltham, MA).

4.4.5 Kinetics of reconstitution

To understand when experimental samples would be ready at an equilibrium state, the kinetics of myoglobin reconstitution were examined. A similar experiment using a single subunit of hemoglobin was reported by Vasudevan and McDonald with cyanide-coordinated protohemin (an iron-centered porphyrin).[24] This previous experiment revealed that the mechanism of porphyrin uptake into a heme protein is more complex than a simple bimolecular event and rate constants for each of four steps were reported. Here, tin (IV) protoporphyrin IX dimethyl ester (SnPPIX) was used to reconstitute apo-myoglobin as an example of a metalloporphyrin which likely forms a covalent bond with the protein's proximal histidine. In addition, this porphyrin has an intense Soret peak that red-shifted appreciably as it was taken up by the protein.

The heme-free proteins prepared in Section 4.3.1 certainly have a tertiary structure that is different than the holo-protein. However, there is evidence that it retains a looser but distinct "molten globule" structure with more than one intermediate form and measurable conformational entropy.[25] In addition to the initial insertion step, porphyrin substitution then requires side-chain rearrangement, which may occur in phases.

Here, the kinetics of SnPPIX substitution into apo-myoglobin was studied with absorption spectroscopy. The porphyrin was dissolved in a 0.1 M NaOH solution and added to an aqueous solution of apo-myoglobin in overwhelming excess (10 : 1 ratio) at room temperature to assure the reaction proceeded readily. First, a photodiode array took broadband spectra every 30

seconds for five hours. This monitored the shift in the easy-to-see Soret band of SnPPIX as the substitution proceeded (Fig. 4.4a). In addition, a scanning spectrophotometer monitored absorbance at 421 nm (the visible wavelength of greatest intensity increase) over four hours in another trial of the same substitution (Fig. 4.4b).
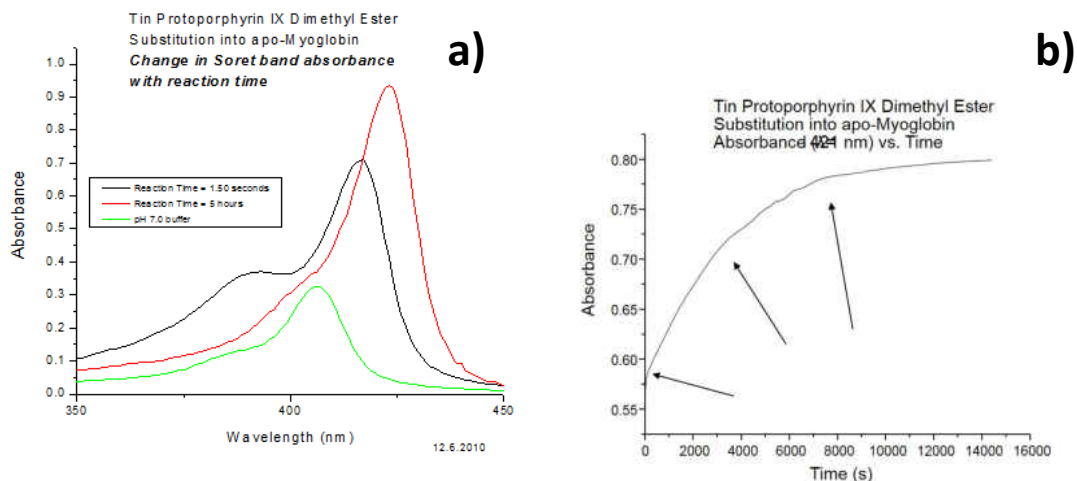
Figure 4.4. a) Red-shift and intensification of Soret band as SnPPIX is substituted into myoglobin from pH 7.0 buffer (green) immediately after mixing with protein (black) to 5 hours after reaction time (red). b) Intensity of Soret peak over 5 hours reveals a number of steps with individual rates. The change in slope for the plot is indicated by arrows.

The results showed a red-shift of the Soret band maximum approximately 10 nm and increasing absorption intensity, which characteristics of a porphyrin uptake into apo-myoglobin.[26] Like the Vasudevan and McDonald results, a plot of absorbance vs. time over four hours indeed seems to show a four-phase reaction, each with a characteristic rate. Figure 4.4b captures the spectrum at the apparent phase changes. The first phase shows a rapid jump in Soret peak intensity (from A = 0.3 in solution to A = 0.7 after 1.50 seconds reaction time) as well as a 10 nm shift to lower energy. The second phase continues the intensity increase at a slower

rate. The third and fourth phase are characterized by a slow drift to the red, but intensity seems to be constant at those stages. The total reaction time is greater than several hours at room temperature. Thus, the reaction seems to be comprised of an initial bimolecular, fast step followed by three unimolecular, slow steps of comparable magnitude

For the purposes of preparing substituted myoglobin and hemoglobin for internal electric field measurements, the results of the above kinetics study indicates that any newly prepared sample should be allowed to equilibrate for several hours, or longer if kept at cold temperatures, before use or storage in liquid nitrogen.

4.5 Hemoglobin subunit isolation

Because hemoglobin's native quaternary structure is comprised two distinct subunits, alpha and beta, each with its own heme active site,[27] it was desirable for the purpose of these experiments to have the ability to substitute only one of the two subunit types in a given quaternary structure. This, in effect, would allow only either the alpha or beta subunit environment to be reported spectroscopically while the other, still containing the native heme, would be invisible in a fluorescence experiment. There are a number of hemoglobin subunit separation methods summarized by Bucci[28], some of which are species specific, and with some appropriate for the holo-protein and others for the apo-protein. The following procedure was explored to separate native equine hemoglobin subunits and does so after the heme has been removed using one of the heme removal methods detailed earlier. Once subunits are separated, subsequent substitution can take place one at a time and then mixed with excess beta subunit, where the quaternary structure apparently reassembles spontaneously.[26]

4.5.1 Mercaptoethanol / urea method for apo-subunits[28-29]

Using either the acid-acetone or methyl ethyl ketone method for heme removal as described above, apo-hemoglobin was produced from a native sample as completely as possible. The sample was then separated into subunits on a cation-exchange column (5-mL HiTrap SP Sepharose FF Column, GE Healthcare Bio-Sciences, Uppsala, Sweden). The sample was first dialyzed in a 50-fold greater volume of the starting buffer of 8 M urea (CH_4N_2O) and 0.05 M 2-mercaptoethanol (C_2H_6OS) in pH 6.7 phosphate buffer. Once loaded onto the column, a two-column volume of starting buffer removed impurities. Then, the elution buffer, which was the starting buffer plus 1 M NaCl, introduced a sodium ion gradient over a period of 30 min or more at a flow rate of 1 mL / min. The two subunits are to elute in distinct peaks, the beta subunit first followed by the alpha subunit. This was not achieved in the course of this work, however, both better control of pH as well as a decrease in the sodium ion gradient might improve results.

4.6 Conclusions

Molecular probes are well established as sensors of electrostatic field in proteins. The use of porphyrins as molecular probes to quantitatively report net internal electric field in heme proteins is being developed. The well-known structural properties of porphyrins assist this development and indicates careful control of sample pH. Advantages of a relatively large molecular probe over a diatomic or atomic-resolution method include the ability to sense the *net* environment at an active site as opposed to microenvironment at that does not describe the cofactor's electrostatic experience. In addition, as heme analogs, the effect of the net environment on electronic properties, such as partial atomic charge on individual atoms and induced dipole moment, of the porphyrin probes can be extended to the native heme.

Efforts to create myoglobin and hemoglobin samples containing porphyrin probes of high purity for use with such experiments has been described. Purification criteria has been established and a number of tools have been assessed for purification procedures. A look at the kinetics of porphyrin probe substitution revealed a long equilibrium time, which should also be taken into account when preparing a high-quality sample.

Options for porphyrin probes in future experiments include both free-base and zinc metalloporphyrins in addition to charged and capped (neutral) propionate substituents. A metallation procedure has been reviewed and provides access to classes of porphyrins that may not be commercially available in with zinc metal. Hemoglobin samples have the potential for mixtures of probes using a subunit separation step.

4.7 References

1.   Andrews, S. S.; Boxer, S. G., Vibrational Stark Effects of Nitriles I. Methods and Experimental Results. *The Journal of Physical Chemistry A* **2000,** *104* (51), 11853-11863.

2.   Park, E. S.; Andrews, S. S.; Hu, R. B.; Boxer, S. G., Vibrational Stark Spectroscopy in Proteins: A Probe and Calibration for Electrostatic Fields. *The Journal of Physical Chemistry B* **1999,** *103* (45), 9813-9817.

3.   Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G., Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment with Theory. *Science* **2006,** *313* (5784), 200-204.
4.   Liu, C. T.; Layfield, J. P.; Stewart, R. J.; French, J. B.; Hanoian, P.; Asbury, J. B.; Hammes-Schiffer, S.; Benkovic, S. J., Probing the Electrostatics of Active Site Microenvironments along the Catalytic Cycle for Escherichia coli Dihydrofolate Reductase. *Journal of the American Chemical Society* **2014**.

5.   Anderson, J. L. R.; Chapman, S. K., Ligand probes for heme proteins. *Dalton Transactions* **2005,** (1), 13-24.

6.   Wu, H. M.; Ratsep, M.; Young, C. S.; Jankowiak, R.; Blankenship, R. E.; Small, G. J., High-Pressure and Stark Hole-Burning Studies of Chlorosome Antennas from Chlorobium tepidum. *Biophys. J.* **2000,** *79* (3), 1561-1572.

7.   (a) Kohler, B. E.; Woehl, J. C., Measuring internal electric fields with atomic resolution. *Journal of Chemical Physics* **1995,** *102* (20), 7773; (b) Woehl, J. C. Measuring Internal Electrostatic Fields and Potentials at Molecular and Atomic Resolution using Hole-Burning Spectroscopy. University of California, Riverside, Riverside, CA, 1996.

8.   (a) Renge, I.; van Grondelle, R.; Dekker, J. P., Pigment Spectra and Intermolecular Interaction Potentials in Glasses and Proteins. *Biophys. J.* **2007,** *93* (7), 2491-2503; (b) Berlin, Y.; Burin, A.; Friedrich, J.; Kohler, J., Spectroscopy of Proteins at Low Temperature. Part I: Experiments with Molecular Ensembles. *Physics of Life Reviews* **2006,** *3* (4), 262-292.

9.   (a) Lim, A. R.; Sishta, B. P.; Grant Mauk, A., Contribution of the heme propionate groups to the electron transfer and electrostatic properties of myoglobin. *Journal of Inorganic Biochemistry* **2006,** *100* (12), 2017-2023; (b) De Petris, A.; Chiavarino, B.; Crestoni, M. E.; Coletti, C.; Re, N.; Fornarini, S., Exploring the Conformational Variability in the Heme b Propionic Acid Side Chains through the Effect of a Biological Probe: A Study on the Isolated Ions. *The Journal of Physical Chemistry B* **2015**.

10.   Falk, J. E., *Porphyrins and Metalloporphyrins*. Elsevier Publishing Company: Amsterdam, 1964; Vol. 2.

11.   (a) Storm, C. B.; Teklu, Y.; Sokoloski, E. A., N-H TAUTOMERISM IN PORPHYRINS AND CHLORINS*. *Annals of the New York Academy of Sciences* **1973,** *206* (1), 631-640; (b) Abraham, R. J.; Hawkes, G. E.; Smith, K. M., N-H tautomerism in porphyrins: an NMR study. *Tetrahedron Letters* **1974,** *15* (16), 1483-1486.

12.   (a) Fidy, J.; Vanderkooi, J. M.; Zollfrank, J.; Friedrich, J., More Than Two Pyrrole Tautomers of Mesoporphyrin Stabilized by a Protein. High Resolution Optical Spectroscopic Study. *Biophys. J.* **1992,** *61* (2), 381-391; (b) Gouterman, M., Spectra of porphyrins. *Journal of Molecular Spectroscopy* **1961,** *6* (0), 138-163.

13.   Gafert, J.; Friedrich, J., Looking into Frozen Proteins with Hole Burning. *J. Lumin.* **1995,** *64* (1), 45-50.

14.   Kohler, M.; Friedrich, J.; Balog, E.; Fidy, J., Symmetry Breaking by the Heme Pocket in Horseradish Peroxidase as Revealed by Stark Spectroscopy. *Chem. Phys. Lett.* **1997,** *277* (5), 417-422.

15.   Boxer, S. G., Stark Realities. *Journal of Physical Chemistry* **2009,** *113* (2972-2983).

16.   (a) Rossi Fanelli, A.; Antonini, E.; Caputo, A., Studies on the structure of hemoglobin I. Physicochemical properties of human globin. *Biochimica et Biophysica Acta* **1958,** *30* (3), 608-615; (b) Ascoli, F.; Rossi Fanelli, M. R.; Antonini, E., Preparation and Properties of Apohemoglobin and Reconstituted Hemoglobins. *Methods in Enzymology* **1981,** *76*, 72-87.

17.   (a) Teale, F. W. J., Cleavage of the haem-protein link by acid methylethylketone. *Biochim. Biopys. Acta* **1959,** *35*, 543; (b) Breslow, E., Changes in Side Chain Reactivity Accompanying the Binding of Heme to Sperm Whale Myoglobin. *The Journal of Biological Chemistry* **1964,** *239* (2), 486-496.

18.   O'Malley, S. M.; McDonald, M., Steady State Fluorescence Energy Transfer Measurements of Human Alpha Apohemoglobin Structure. *Biochemical and Biophysical Research Communications* **1994,** *200* (1), 384-388.

19.   Fruk, L.; Kuo, C.-H.; Torres, E.; Niemeyer, C. M., Apoenzyme Reconstitution as a Chemical Tool for Structural Enzymology and Biotechnology. *Angewandte Chemie International Edition* **2009,** *48* (9), 1550-1574.

20.   Winterhalter, K. H.; Huehns, E. R., Preparation, Properties, and Specific Recombination of alpha & beta-Globin Subunits. *J. Biol. Chem.* **1964,** *239* (11), 3699-3705.

21.   Tamura, M.; Woodrow Iii, G. V.; Yonetani, T., Heme-modification studies of myoglobin: II. Ligand binding characteristics of ferric and ferrous myoglobins containing unnatural hemes. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1973,** *317* (1), 34-49.

22.   Sono, M.; Asakura, T., Separation and properties of spirographis and isospirographis porphyrin dimethyl esters. *Biochemistry* **1974,** *13* (21), 4386-4394.

23.   Atassi, M. Z., Immunochemistry of Sperm-Whale Myoglobins Prepared with Various Modified Porphyrins and Metalloporphyrins. *Biochemical Journal* **1967,** *103* (2), 29-35.

24.   Vasudevan, G.; McDonald, M. J., Spectral Demonstration of Semihemoglobin Formation during CN-Hemin Incorporation into Human Apohemoglobins. *Journal of Biological Chemistry* **1997,** *272* (1), 517-524.

25.   (a) Jennings, P.; Wright, P., Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993,** *262* (5135), 892-896; (b) Loh, S. N.; Kay, M. S.; Baldwin, R. L., Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proceedings of the National Academy of Sciences* **1995,** *92* (12), 5446-5450; (c) Stadler, A. M.; Koza, M. M.; Fitter, J., Determination of Conformational Entropy of Fully and Partially Folded Conformations of Holo- and Apomyoglobin. *The Journal of Physical Chemistry B* **2014**.

26.   Leonard, J. J.; Yonetani, T.; Callis, J. B., Fluorescence study of hybrid hemoglobins containing free base and zinc protoporphyrin IX. *Biochemistry* **1974,** *13* (7), 1460-1464.

27.   Voet, D., Voet, Judith G., Hemoglobin: Protein Function in Microcosm (Chapter 9). In *Biochemistry, Ed. 2*.

28.   Bucci, E., Preparation of Isolated Chains of Human Hemoglobin. *Methods in Enzymology* **1981,** *76*, 97-106.

29.   (a) Wilson, S.; Smith, D. B., Separation of the Valyl-Leucyl- and the Valyl-Glutamyl-Polypeptide Chains of Horse Globin by Fractional Precipitation and Column Chromatography. *Canadian Journal of Biochemistry and :Physiology* **1959,** *37*, 405; (b) Clegg, J. B.; Naughton, M. A.; Weatherall, D. J., Abnormal human haemoglobins: Separation and characterization of the Î± and Î² chains by chromatography, and the determination of two new variants, Hb chesapeake and Hb J (Bangkok). *Journal of Molecular Biology* **1966,** *19* (1), 91-IN2.

Chapter 5

Spectral Characterization of Porphyrins in Heme Proteins

5.1 Introduction

The porphyrin probes described in Chapter 4 were characterized spectrally with absorbance and emission spectra. The porphyrins of interest are treated in solution and substituted into protein as well as in a number of solvents and at room and low (77 K) temperature. These characterizations were performed in preparation for use with hole-burning Stark spectroscopy (described in Chapter 6).

First, the well-developed porphyrin theory known as the four-orbital model is discussed. Then, inhomogeneous broadening as a factor in spectra is also described theoretically to gain a better understanding of some interesting "splitting" revealed during spectral collection (described in Section 5.3.3.2). Experimental results follow with tabulated maxima for porphyrins in different environments, which demonstrates the sensitivity of the probe towards its surroundings. Much of the experimental section is devoted to developing a method to collect spectra of porphyrin solutions submerged in liquid nitrogen followed by results that show the unexpected splitting in the spectra. The remainder of the chapter attempts to deduce the cause of spectral splitting.

5.2 Theory of porphyrin UV–VIS spectra

5.2.1 Gouterman's "four-orbital model"

As established in the previous chapter, porphyrin's symmetry properties lends predictability to its electronic spectrum. A theory was developed by Gouterman in the late 1950s to explain the consistent pattern of peaks in the UV–VIS spectrum of porphyrins. [1-2] It became known as the "four-orbital model," using only the two highest-occupied and two lowest-unoccupied molecular orbitals in a simplified picture. It is based on Hückel's LCAO-MO (linear combination of atomic orbitals – molecular orbitals) theory to establish the energies of the transitions in the UV–VIS region but also importantly brings in configuration interaction (CI) to explain the distinct differences in intensity between transitions seen specifically in porphyrins.

Experimentally, a porphyrin's absorption spectrum typically consists of a high-intensity peak near the UV region (~350-450 nm), known as the Soret (or B) band, as well as peak of much lower intensity in the visible region ~(550-650 nm), known as the Q ("quasi-allowed") band. The difference in intensity between these two bands is 100 fold. The Q band is accompanied by a resolved vibronic band of similar intensity at slightly higher energy (~500 nm), which may be denoted as Q_{elec} and Q_{vib}, respectively. Further, in free-base porphyrins, where symmetry is reduced, the Q band splits into a lower-energy Q_x peak and higher-energy Q_y peak, each with accompanying vibronic band (see Fig. 5.1).
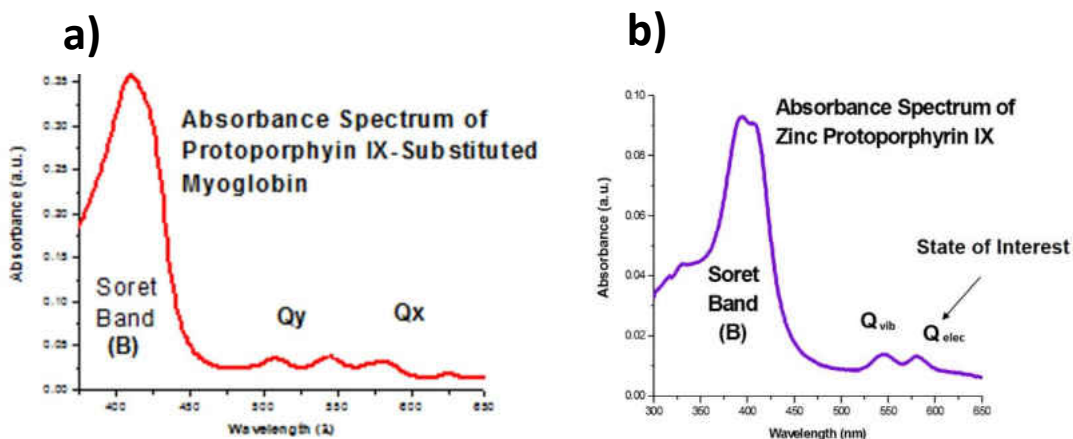
Figure 5.1. A typical UV-VIS absorption spectrum for **a)** free-base porphyrin and **b)** metalloporphyrin. (Author's data.)

This typical spectral pattern holds under a remarkable number of perturbations. However, the energetic position and relative intensities of the peaks can be affected by varying any of the following:[3] type of metal center (or removal of the metal), electron-donating or –withdrawing substituents, symmetry of substituent placement around the ring, solvent, dimerization or aggregation, and protein substitution. It is important to understand the factors influencing porphyrin spectra because the method described in this work of measuring internal electric fields in heme proteins treats the surrounding electrostatic environment as a perturbation to the theoretical probe, free-base or metallic porphin. Thus we begin with a presentation of this basic four-orbital model theory of porphyrin spectra.

a) Protoporphyrin IX Dimethyl Ester
Chloroform solvent

(506, 0.19747)
(542, 0.17592)
(576, 0.12607)
(631, 0.10804)

IV
III
II
I

*Q-band region

b) Zinc Protoporphyrin IX Dimethyl Ester
Incomplete metal chelation
Chloroform solvent

(541, 0.23466)
(578, 0.25563)
(506, 0.16818)
(631, 0.07897)

IV
III
II
I

*Q-band region

c) Zinc Protoporphyrin IX Dimethyl Ester
Complete chelation
Chloroform solvent

(540, 0.0659)
(577, 0.09186)

vibrational band
electronic band

*Q-band region

Figure 5.2. Chelation of zinc by protoporphyrin IX dimethyl ester as described in Ch. 4 illustrates the evolution of the Q-band region of free-base to metalloporphyrin.
(Author's data.) **a)** Before chelation, **b)** incomplete chelation, **c)** complete chelation. In a) and b), Roman numerals tag individual peaks in the Q-band region from low to high energy.

90

First, spatial orientation and nomenclature must be established (see Fig. 2). The porphyrin ring lies in the x–y plane with in-plane axes cutting through the pyrrole nitrogens. In free-base porphyrin, the x-axis is specified as the one that cuts through the two inner hydrogens and the y-axis cuts through the pyrrolenine nitrogens, which each have a lone pair. As a result, the in-plane axes in free-base porphin are not equivalent, but for metallic porphin, they are degenerate. The outer pyrrole carbons are numbered 1–8 clockwise, starting at the positive y-axis; the methine bridge carbons are labeled clockwise α–δ. Pyrrole rings are also labeled clockwise with Roman numerals, I–IV.
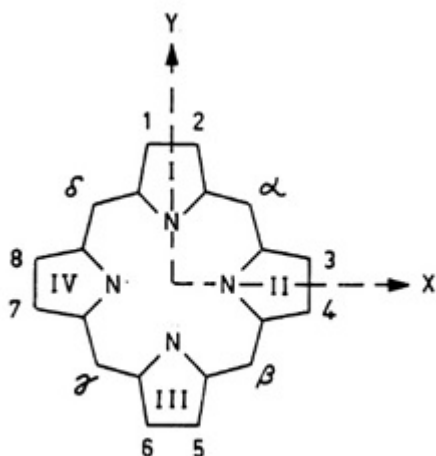
Figure 5.3. Nomenclature and spatial orientation of porphyrin ring skeleton. (Figure from Ref. [1].)

Visible transitions in porphin are π–π transitions. The four orbitals alluded to are the HOMO and HOMO-1 with symmetry a_{2u} and a_{1u}, and the LUMO and LUMO+1, both with symmetry e_g. However, the relative positions each HOMO orbital and each LUMO with respect to one another are influenced by the perturbations mentioned above. Because the four-orbital model can still be applied to porphyrins that do not retain strict symmetry, the HOMO orbitals are labeled b_1 and b_2 and the LUMO orbitals are labeled c_1 and c_2 (see Fig. 5.3). Molecular orbital calculations in

porphyrins give b_1 a higher absolute energy than b_2. As a result, the two transitions originating

from the b_1 orbital (b_1c_1 and b_1c_2) are assigned as Q transitions while the two originating from

the b_2 orbital (b_2c_1 and b_2c_2) are assigned as B transitions. However, one of the important insights

contributed by Gouterman based on CI is that this situation should result in Q and B bands of

similar intensity due to orbital mixing, i.e., intensity "borrowing" from the highly allowed B-band

by the technically forbidden Q-band. For this theory, then an assumption was made that the two

HOMO levels, b_1 and b_2, are "accidentally degenerate."

In metallic porphin, with D_{4h} symmetry, the identical symmetry LUMO orbitals are

energetically degenerate. In free-base porphin, however, the c_1 LUMO molecular orbital has

density on the pyrrolenine nitrogens (along the y-axis) while the c_2 has density on the pyrrole

nitrogens, which are protonated along the x-axis. It is assumed the hydrogens have a stabilizing

effect on the c_2 orbital, so, for free-base porphyrin, c_2 is assigned to have lower energy than c_1.
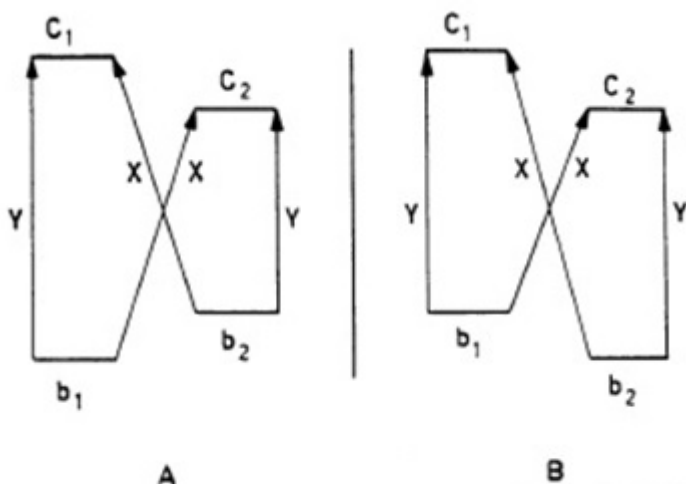
Figure 5.4. The four orbitals of porphin (two HOMO and two LUMO) with possible energetic arrangements. Possibility A is shown theoretically to be the likely arrangement. (Figure from Ref. [1].)

For free-base porphyrins, it has been shown by fluorescence polarization experiments[4] that both the Q-band and B-band each contain two transitions, as predicted, that are polarized perpendicularly to one another along the x- and y-axes. For metalloporphyrins, the degenerate transitions make further differentiation impossible, however, for free-base porphyrins, the distinct splitting in the now non-degenerate Q-region makes assignment of the higher (~550 nm) and lower (~650 nm) energy peaks as either x-polarized or y-polarized possible. Further arguments from the four-orbital model, especially the CI insight, and observations from experiment assign the lowest-energy Q-peak as Q_x and the higher-energy Q-peak as Q_y in free-base porphin, as explained in the following.

Experiments designed to intentionally raise or lower the HOMO energies have predictable effects on the Q-band intensity, recalling that the proximity of the b_1 and b_2 states dictate the

extent of mixing according to Eq. 5.1 and 5.2, noting that only the same polarization is available

for mixing:

$$Q_x = Q_x^0 + \lambda_x B_x^0 \tag{5.1a}$$

$$Q_y = Q_y^0 + \lambda_y B_y^0 \tag{5.1b}$$

$$\lambda_x = \{[\varepsilon(c_2) - \varepsilon(b_1)] - [\varepsilon(c_1) - \varepsilon(b_2)]\}/2\,\Delta \tag{5.2a}$$

$$\lambda_y = \{[\varepsilon(c_1) - \varepsilon(b_1)] - [\varepsilon(c_2) - \varepsilon(b_2)]\}/2\,\Delta \tag{5.2b}$$

where Δ is the initial energy gap between the Q-band and the B-band; $\varepsilon(x)$ is the energy of x

orbital; and Q^0 and B^0 are the transition intensities pre-mixing. From the equations, the conditions

that produce a larger amount of mixing (i.e., largest λ value) are those in which the pairs of

transitions with the same polarization have the largest difference in transition energy. Therefore,

experiments that can adjust the energies of the four orbitals individually might reveal the identity

of each with respect to the porphin structure by noting the effect on band intensity.

$c_1 (e_g)$ $c_2 (e_g)$

$b_1 (a_{2u})$ $b_2 (a_{1u})$
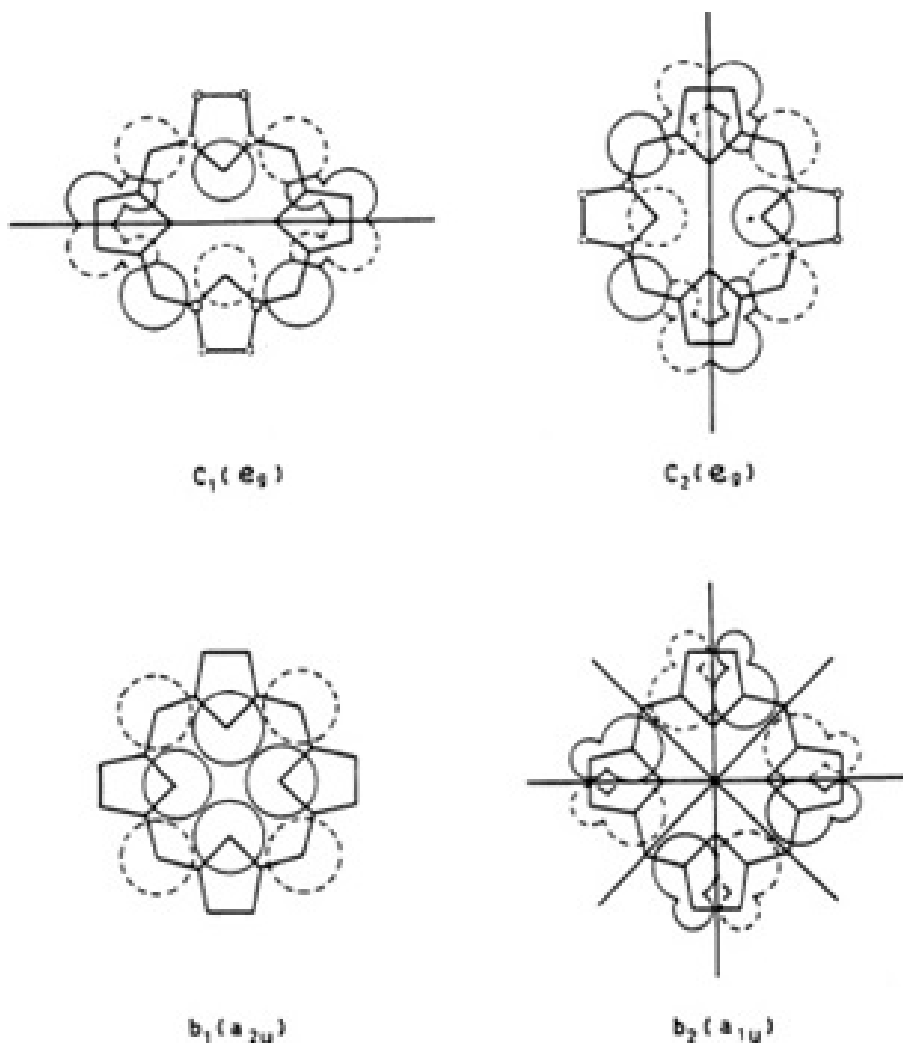
Figure 5.5. Plots of MO coefficients for porphyrin. The inner hydrogens of porphin specifically have density in the c_2 orbital. Also note where density lies for positions 1–8 to interpret the experiments that follow. (Figure from Ref. [1].)

Such experimental observations came from the addition of symmetric octa-alkyl substituents to the free-base porphin ring at positions 1–8. Looking at MO plots of porphin (see Fig. 5.5) it can be seen that substituents in those positions would affect orbitals, b_2, c_1, and c_2 but not b_1. As the substituents become larger and more electron donating, the energy of these

orbitals can expect to rise. The spectral effect observed was that the intensity of the Q-bands increases at the same rate. This would occur if the λ term also increased, indicating that the difference in energy between Q- and B-transitions of *both* polarizations became more exaggerated. This is only possible if b_2, being raised in energy by the substituents, was already higher than $b_{1'}$ which remains stationary in this substitution, and moved farther apart. The same substituents, now applied along one axis or the other as 1,2,5,6-tetra-alkylporphin (*y*-axis, affecting b_2 and c_1) and 3,4,7,8-tetra-alkylporphin (*x*-axis, affecting b_2 and c_2), have the effect of lessening the intensity between the Q-bands and increasing them, respectively.

5.2.1 Inhomogeneous broadening of spectra

In addition to the environment discretely shifting the transition energy of the UV-VIS spectrum (in addition to rotational and vibrational transitions), like all chromophores in solid environments, porphyrins are also subject to inhomogeneous broadening of spectral bands.[5] Each guest chromophore in the ensemble will have an individual *homogeneous* linewidth whose frequency and width is influenced by the variety of steric and energetic microenvironments provided by the host, for example, the two-level systems in proteins as discussed in Chapter 3. Then, the summation of all of the individual homogeneous linewidths is the *inhomogeneously* broadened band as seen in an experimental spectrum. The results are bands that are spread out, reducing intensity and covering a wider energetic range. This often causes spectral details to be obscured. For measurement of internal electric fields by Stark spectroscopy, this poses a special challenge because the expected Stark shifts for porphyrins in heme protein are far smaller than the width of inhomogeneously broadened bands. The theory of inhomogeneous broadening presented here will help to understand the subtle changes that broadband spectra display in

different environments presented in this chapter; then, overcoming inhomogeneous broadening to resolve Stark shifts by spectral hole burning will be discussed in Chapter 6.

First, the homogeneous linewidth produced by a single guest molecule (or several with accidentally degenerate energetic environments) in an ensemble in a solid host is established by two quantities: electronic excited lifetime (τ_1) and phonon scattering, which also has an associated lifetime (τ_2). Each process its own temperature dependence, both leading to line broadening as temperature increases. Thus, in this work, many of the spectra are collected at low temperature (T < 77K), necessitating our treatment of homogeneous linewidths in a solid host matrix.

Two distinct parts comprise the shape actual homogeneous spectrum, each related to the lifetime quantities. In the limit of low temperature (T = 0 K), the spectrum is only made up of the zero-phonon line (ZPL), which is Lorentzian in shape and whose width given by:

$$\Gamma_{hom}(0) = \frac{1}{2\pi \cdot c \cdot \tau_1(0)} \ cm^{-1} \qquad (5.3)$$

which reflects that a long excited-state lifetime, and the radiative pathway is more probable. As temperature begins to rise, vibrations in the solid host matrix (i.e., phonons) and the result is a shorter excited-state lifetime as well as the introduction of an additional guest-electron–host-phonon interaction lifetime, τ_2, which acts through dephasing the wavefunction of the guest molecule, temporarily "resetting" it. This τ_2 lifetime also contributes to the width of the ZPL. At temperatures above 0 K, the homogeneous linewidth is:

$$\Gamma_{hom}(T > 0 \ K) = \frac{1}{2\pi}\left[\frac{1}{\tau_1(T)} + \frac{2}{\tau_2(T)}\right] \qquad (5.4)$$

As phonons arise in the solid host matrix, an additional phonon side band (PSB) also appears independently of the ZPL. Aside from the width, the combination of ZPL and PSB functions into a homogeneous *lineshape* function is also of interest:

$$HLF(v) = \alpha \cdot ZPL(v) + (1 - \alpha) \cdot PSB(v) \qquad (5.5)$$

where α is the Debye–Waller factor, the ratio of the ZPL intensity over the total homogeneous line intensity (see Fig. 5.5a):

$$\alpha(T) = \frac{I_{ZPL}}{I_{ZPL} + I_{PSB}} \qquad (5.6)$$

Thus, temperature contributes to overall linewidth by affecting the guest molecule lifetimes and in turn the *homogeneous* linewidth. In some systems, the temperature dependence can be very large, on the order of T^7.

The superposition of individual homogeneous linewidths in a statistical distribution, on the other hand, are what comprise an inhomogeneously broadened band (see Fig. 5.5b). The energetic position of the ZPL is established by not only the identity of guest molecule but also by the type and intensity of influence of the host environment. An environmental influence might be steric factors (or strain fields in solid hosts) that alter the guest molecule's bond lengths or bond angles. This can often be the largest contributor to inhomogeneous broadening.[6] Other factors are necessarily electrostatic: internal electric fields generated by charged groups in the host and dispersive interactions, often associated with solvent shifts.[7]

a)

Absorption or Emission Intensity

Wavelength or Frequency
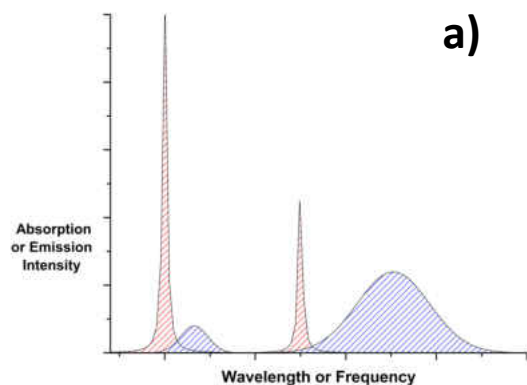
Figure 5.6. **a)** Two examples of a homogeneous line. Left: high Debye–Waller factor sharpens the ZPL. Right: lower Debye–Waller factor moves intensity from the ZPL to the PSB. Total area under each curve remains constant. **b)** A distribution of homogeneous lines in an inhomogeneously broadened band.

b)

Absorption or Emission Intensity

Wavelength or Frequency
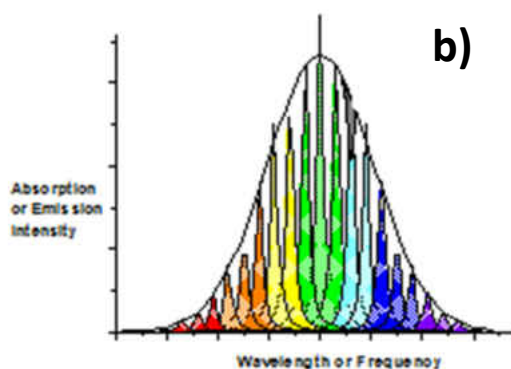
The resulting broadened band tends to be Gaussian-shaped in general. However, theoretically, point dipole perturbations result in a Lorentzian distribution but the inhomogeneous broadening is much larger than the Lorentzian distribution, which results in the overall Gaussian shape familiar in absorption spectra.[6] The extent of broadening in PPIX-substituted myoglobin has been measured at 4×10^4 times the homogeneous linewidth at T = 1.5 K.[8]

Two general types of solid hosts characterize the extreme limits of broadening environments (see Fig. 5.6). At the narrow limit, crystalline hosts, by definition, form periodic matrices with little variation for a guest molecule to sense. In fact, a perfect crystal without defect may, in theory, reveal the true homogeneous linewidth. At the broadest limit, amorphous hosts are characterized by random orientations where guest molecules may sample the full distribution of microenvironments that result. In this work, analysis relies on the existence of a completely random spatial distribution of chromophores. Thus, both protein-substituted and non-substituted porphyrins have been characterized in glycerol solutions when analyzed at low temperature. In addition, in Chapter 3, we have already characterized the protein matrix itself as a glassy system like frozen glycerol, i.e., a solid amorphous host, where structural relaxations continue to contribute to overall entropy even at very low temperature.

a) Planar reporter chromophore inside a spherical host, e.g., heme in myoglobin

b) Planar reporter chromophore inside disordered host
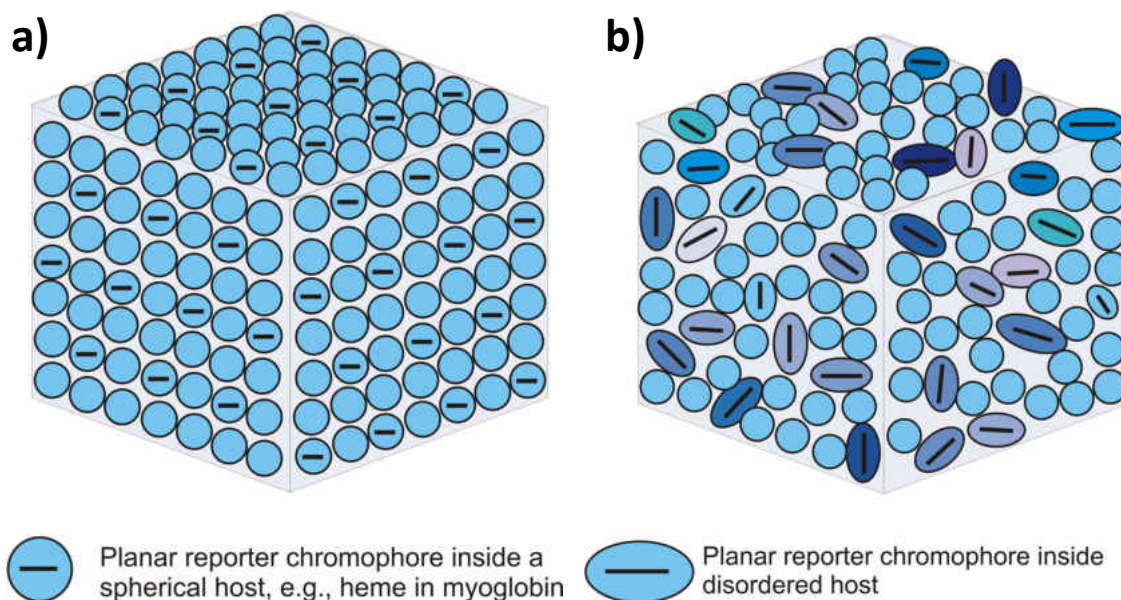
Figure 5.7. a) A crystalline host matrix in which each planar guest chromophore, e.g., porphyrin, experiences approximately the same environment. **b)** An amorphous host matrix in which each individual guest experiences slightly different microenvironments.

Finally, broadened spectra also may be due to not only a *distribution* of environments but in fact may conceal a mixture of distinct environments. One example that concerns this work is tendency of porphyrins to form dimers and larger aggregates when in solution because of the tendency of these planar molecule to undergo *pi*-stacking.[9] Another example is a sample containing both correctly protein-substituted and unsubstituted porphyrin, which we have described as being contaminants. We will see that the energetic distance between these two porphyrin environments can be greater than 10 nm in an absorption spectrum. The success of these experiments relies on being able to assume the measurement of the average of a distributed but *sole* environment.

5.3 Experimental porphyrin spectra

Broadband absorption and emission spectra were collected for porphyrins in preparation for Stark spectroscopy analysis, where the spectral window is on the order of a thousandth of a nanometer and peak maxima are targeted. The porphyrins of interest were analyzed with a commercial absorption spectrophotometer and fluorimeter to locate the absorption maxima (see Chapter 2, Materials & Methods). The porphyrins studied were protoporphyrin IX, zinc protoporphyrin IX, protoporphyrin IX dimethyl ester, zinc protoporphyrin IX dimethyl ester, and tin protoporphyrin IX dimethyl ester. A number of environments were studied: aqueous (or organic in the case of dimethyl ester porphyrins) solution vs. glycerol solution; myoglobin substituted vs. solution; room temperature vs. low temperature (77 K). The maxima for each UV–VIS band are tabulated here.

5.3.1 Tabulated room-temperature spectra

See pages following.

5.3.1.1 Protoporphyrin IX

Table 5.1. Peak maxima (wavelength in nm) in absorbance spectra for protoporphyrin IX (PPIX) in a number of local environments at room temperature. Molar absorptivity coefficient for PPIX in aqueous solution at $\varepsilon_{407\ nm}$ = 4.4 x 10^4 M^{-1} cm^{-1}.[10] Molar absorptivity coefficient for PPIX-myoglobin at ε_{405nm}= 8 ×10^4 M^{-1} cm^{-1}.[11]

Absorbance Measurement for Protoporphyrin IX							
Protein	Solvent	Notes	Soret (nm)	$Q_{y,vib}$ (nm)	$Q_{y,elec}$ (nm)	$Q_{x,vib}$ (nm)	$Q_{x,elec}$ (nm)
None	Aqueous	pH 7.0 phosphate buffer	374	484	536	589	643
None	3:1 glycerol:water		397	506	540	565	643
None	DMSO		407	505	541	574	629
Myoglobin	Aqueous	pH 7.0 phosphate buffer	410	507	544	583	626
Myoglobin	3:1 glycerol:water		419	481	543	584	625

Table 5.2. Peak maxima (wavelength in nm) in emission spectra for protoporphyrin IX (PPIX) in a number of local environments at room temperature. Excitation region and wavelength in nm are given in parentheses following emission peak position.

Emission Measurement for Protoporphyrin IX							
Protein	**Solvent**	**Notes**	**Peak 1 (Excitation)**	**Peak 2 (Excitation)**	**Peak 3 (Excitation)**	**Peak 4 (Excitation)**	**Peak 5 (Excitation)**
None	Aqueous	pH 7.0 phosphate buffer	618 (*Soret,397*)	676 (*Soret,397*)			
None	3:1 glycerol:water		623 (*Soret,378*)	624 (*$Q_{x,vib}$,482*)	623 (*$Q_{x, elec}$,541*)	646 (*$Q_{y,vib}$,593*)	652 (*$Q_{x, elec}$,644*)
Myoglobin	Aqueous	pH 7.0 phosphate buffer	653 (*Soret,400*)				
Myoglobin	3:1 glycerol:water		434 (*Soret,410*)	630 (*Soret,410*)	672 (*Soret,410*)		

5.3.1.2   Zinc protoporphyrin IX

Table 5.3. Peak maxima (wavelength in nm) in absorbance spectra for zinc protoporphyrin IX (ZnPPIX) in a number of local environments at room temperature. Molar absorptivity coefficient for ZnPPIX-hemoglobin subunits is at ε_{423nm}=122 mM^{-1} cm^{-1}, which was used as a comparable value for ZnPPIX-myoglobin.[12]

Absorbance Measurement for Zinc Protoporphyrin IX					
Protein	**Solvent**	**Notes**	**Soret (nm)**	**Q_{vib} (nm)**	**Q_{elec} (nm)**
None	Aqueous	pH 7.0 phosphate buffer	405	544	580
None	3:1 glycerol:water		411	544	580
None	DMSO		422	548	586
Myoglobin	Aqueous	pH 7.0 phosphate buffer	423	550	584
Myoglobin	3:1 glycerol:water		423	551	584

Table 5.4. Peak maxima (wavelength in nm) in emission spectra for zinc protoporphyrin IX (ZnPPIX) in a number of local environments at room temperature. Excitation region and wavelength in nm are given in parentheses following emission peak position.

Emission Measurement for Zinc Protoporphyrin IX								
Protein	**Solvent**	**Notes**	**Peak 1 (Excitation)**	**Peak 2 (Excitation)**	**Peak 3 (Excitation)**	**Peak 4 (Excitation)**	**Peak 5 (Excitation)**	**Peak 6 (Excitation)**
None	Aqueous	0.1 M NaOH	587 (*Soret,410*)	635 (*Soret,410*)				
None	3:1 glycerol:water		582 (*Soret,410*)	582 ($Q_{elec},578$)	585 ($Q_{vib},543$)	627 (*Soret,410*)	630 ($Q_{elec},578$)	632 ($Q_{vib},543$)
Myoglobin	Aqueous	pH 7.0 phosphate buffer						
Myoglobin	3:1 glycerol:water	pH 7.0 phosphate buffer	465 (*Soret,410*)	594 ($Q_{vib},543$)	595 ($Q_{elec},410$)	624 (*Soret,410*)	624 ($Q_{elec},578$)	

5.3.1.3 Protoporphyrin IX dimethyl ester

Table 5.5. Peak maxima (wavelength in nm) in absorbance spectra for protoporphyrin IX dimethyl ester (PPIX DME) in a number of local environments at room temperature. Molar absorptivity coefficient for PPIX DME-myoglobin is ε_{407nm}=145 mM^{-1} cm^{-1}.[13]

Absorbance Measurement for Protoporphyrin IX Dimethyl Ester							
Protein	**Solvent**	**Notes**	**Soret (nm)**	**$Q_{y,vib}$ (nm)**	**$Q_{y,elec}$ (nm)**	**$Q_{x,vib}$ (nm)**	**$Q_{x,elec}$ (nm)**
None	Chloroform		424	506	542	576	631
None	Methanol / 1% pyridine		401	502	538	573	628
None	3:1 glycerol: Methanol / 1% pyridine		402	486	542 (shoulder at 566)	595	645
Myoglobin	Aqueous	pH 7.3 phosphate buffer	427	Unobserved	553	597	610
Myoglobin	3:1 glycerol:water	Summed fluorescence excitation	428	506	553	Unobserved	Unobserved

Table 5.6. Peak maxima (wavelength in nm) in emission spectra for protoporphyrin IX dimethyl ester (PPIX DME) in a number of local environments at room temperature. Excitation region and wavelength in nm are given in parentheses following emission peak position.

Emission Measurement for Protoporphyrin IX Dimethyl Ester								
Protein	**Solvent**	**Notes**	**Peak 1 (Excitation)**	**Peak 2 (Excitation)**	**Peak 3 (Excitation)**	**Peak 4 (Excitation)**	**Peak 5 (Excitation)**	**Peak 6 (Excitation)**
None	Methanol / 1% pyridine		675 (*Soret,401*)	740 (*Soret,401*)	675 ($Q_{y,vib}$,*502*)	740 ($Q_{y,vib}$,*502*)	633 ($Q_{y,elec}$,*538*)	675 ($Q_{y,elec}$,*538*)
None	Methanol / 1% pyridine	Cont'd	741 ($Q_{y,elec}$,*538*)	633 ($Q_{x,vib}$,*573*)	675 ($Q_{x,vib}$,*573*)	739 ($Q_{x,vib}$,*573*)	740 ($Q_{x,elec}$,*628*)	
None	3:1 glycerol: Methanol / 1% pyridine		487 (*Soret,402*)	708 (*Soret,401*)	569 ($Q_{y,vib}$,*486*)	711 ($Q_{y,vib}$,*486*)	672 ($Q_{y,elec}$,*542*)	714 ($Q_{y,elec}$,*542*)
None	3:1 glycerol: Methanol / 1% pyridine	Cont'd	715 ($Q_{x,vib}$,*595*)	700 ($Q_{x,elec}$,*643*)				

5.3.1.4 Zinc protoporphyrin IX dimethyl ester

Table 5.7. Peak maxima (wavelength in nm) in absorbance spectra for zinc protoporphyrin IX dimethyl ester (ZnPPIX DME) in a number of local environments at room temperature.

Absorbance Measurement for Zinc Protoporphyrin IX Dimethyl Ester					
Protein	**Solvent**	**Notes**	**Soret (nm)**	**Q_{vib} (nm)**	**Q_{elec} (nm)**
None	Chloroform		409	540	577
None	Ethanol	Absolute grade	416	546	583
None	50:50 ethanol:glycerol		414	543	579

Table 5.8. Peak maxima (wavelength in nm) in emission spectra for zinc protoporphyrin IX dimethyl ester (ZnPPIX DME) in a single local environment at room temperature. Excitation region wavelength in nm are given in parentheses following emission peak position.

Emission Measurement for Zinc Protoporphyrin IX Dimethyl Ester								
Protein	Solvent	Notes	Peak 1 (Excitation)	Peak 2 (Excitation)	Peak 3 (Excitation)	Peak 4 (Excitation)	Peak 5 (Excitation)	Peak 6 (Excitation)
None	Chloroform		582 ($Soret, 409$)	631 ($Soret, 409$)	581 ($Q_{vib}, 541$)	630 ($Q_{vib}, 541$)	582 ($Q_{elec}, 577$)	630 ($Q_{elec}, 577$)
None	50:50 ethanol: glycerol	No emission features with Q_{elec} excitation	471 ($Soret, 414$)	580 ($Soret, 414$)	632 ($Soret, 414$)	667 ($Soret, 414$)	581 ($Q_{vib}, 543$)	666 ($Q_{vib}, 543$)

5.3.1.5 Tin protoporphyrin IX dimethyl ester (SnPPIX DME)

Table 5.9. Peak maxima in absorbance spectra for tin protoporphyrin IX dimethyl ester (SnPPIX DME) in a number of local

environments at room temperature.

Absorbance Measurement for Tin Protoporphyrin IX Dimethyl Ester					
Protein	Solvent	Notes	Soret (nm)	Q_{vib} (nm)	Q_{elec} (nm)
None	Aqueous	pH 7.0 phosphate buffer	406	541	580
None	3:1 glycerol:water		411	543	582
Myoglobin	Aqueous	pH 7.0 phosphate buffer	416	546	590

Table 5.10. Peak maximum (wavelength in nm) in emission spectra for tin protoporphyrin IX dimethyl ester (SnPPIX DME) in a single

local environment at room temperature. Excitation region wavelength in nm are given in parentheses following emission peak

position.

Emission Measurement for Tin Protoporphyrin IX Dimethyl Ester								
Protein	**Solvent**	**Notes**	**Peak 1 (Excitation)**	**Peak 2 (Excitation)**	**Peak 3 (Excitation)**	**Peak 4 (Excitation)**	**Peak 5 (Excitation)**	**Peak 6 (Excitation)**
None	3:1 glycerol: water		584 (*Soret,411*)	637 (*Soret,411*)	584 (Q_{vib},*541*)	637 (Q_{vib},*541*)	584 (Q_{elec},*582*)	637 (Q_{elec},*582*)

5.3.2   Trends in porphyrin spectra in various environments

In absorbance spectra, a few trends appear for the protoporphyrin species analyzed (see Fig. 5.5). First, between solution and protein substitution, there is a general contraction of the spread of energies, i.e., the Soret band tend to red-shift while the Q bands tend to blue shift. However, the influence on individual bands can vary considerably. Between different solvents, there is a general trend of the Soret band toward lower energies with decreasing dielectric constant:[14] aqueous buffer ($\varepsilon \approx 78$), 3:1 glycerol:water mixture ($\varepsilon \approx 50$),[15] dimethyl sulfoxide ($\varepsilon \approx 47$), methanol / pyridine mixture ($\varepsilon \approx 30$), and chloroform ($\varepsilon \approx 5$). Interestingly, protein's dielectric constant, as questionable as the bulk parameter's application to a macromolecule might be, is typically noted as $\varepsilon \approx 2-4$. The shift in Soret band to lower energies is consistent with the trends seen in the solvent series.

In emission spectra, the appearance of certain maxima is certainly subject to resolution and chromophore concentration, so the data above is not complete. However, it is interesting to note that strong emission maxima appear in the same position for most of the chromophores regardless of whether the excitation was made in the Soret region or a Q region. This may indicate that a typical emission pathway may include a non-radiative relaxation to the lower-energy Q region before radiative relaxation from that position. This idea is reinforced in analyzing low-temperature fluorescence excitation spectra below.

UV - VIS Absorption of Zinc Protoporphyrin IX in Various Environments
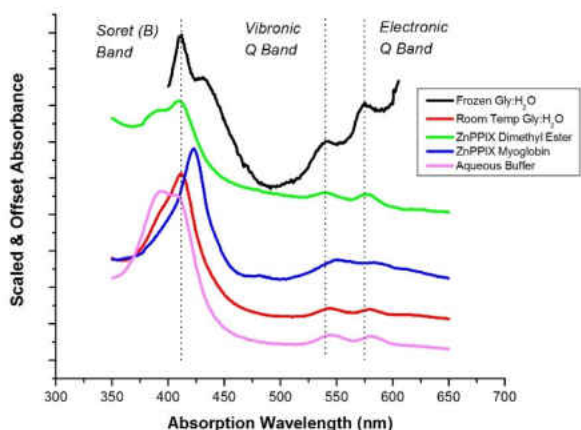
Figure 5.8. Absorbance spectra of zinc protoporphyrin IX in various environments (from top to bottom): frozen in glycerol:buffer, as a dimethyl ester derivative, substituted into myoglobin, room temperature in glycerol:buffer, and in aqueous buffered solution. (Spectra are scaled and offset for clarity to emphasize lineshapes.)

5.3.3 Low-temperature (77 K) spectra

Because Stark spectroscopy would be performed at liquid helium temperatures (>4 K), absorption maxima in a frozen glycerol–water sample were desirable. Spectroscopic measurements using cryogens have well-known difficulties,[16] and the following describes a general simplification of the process. A small (a few hundred milliliters), homemade optical cryostat placed in a commercial spectrophotometer was first employed. The sample was placed in a glass test-tube attached to a rod topped with a rubber stopper. The cryostat was filled with liquid nitrogen and the sample immersed in the cryogen where it solidified immediately, often cracking visibly. Major issues with this direct approach arose. First, the scattering losses were far too great for a transmission experiment. The incident beam of the spectrophotometer encountered the sample only after passing through an outer flat quartz window, an identical inner window, vigorously bubbling cryogen, and the cylindrical test tube wall. Exiting the sample, the light encountered the same optical obstacles as transmission was detected at 180°. In addition, the homemade cryostat, though it contained an evacuated jacket between the inner

and outer walls, failed to insulate effectively and the optical windows began to quickly frost over, especially in humid summer conditions. It was clear that another approach was necessary.

5.3.3.1 Summed fluorescence excitation spectra

To circumvent the overwhelming scattering losses using the direct absorbance measurement with an optical cryostat, it was thought to try to "hunt" for absorption maxima by measuring emission intensity and noting the excitation wavelength that gave the most intense emission. Emission spectra would not be subject to scattering losses that absorption would for two reasons: 1) the measurement does not rely on the detection of an incident beam that undergoes a loss to its intensity during the course of a measurement but rather the luminescence generated by the sample itself; 2) the emission measurement is not done at 180° but rather 90° ("right-angle collection") as typical or in "front-face collection" mode (22.5°), specifically designed for solid samples.

However, the problem of frosted windows remained. An initial solution was to direct a stream of dry-grade compressed air from a cylinder over the windows, which improved the signal enough for a "proof of concept" set of spectra. The homemade cryostat, which was suspended on a ring stand, had the additional issue of altering the position of its flat windows slightly with respect to the incident beam and detector as the experiment proceeded. This caused variations in intensity trial to trial and after more cryogen was adding mid-experiment. Fortunately, the commercial fluorometer had available a liquid-nitrogen dewar add-on sample holder (see Chapter 2, Materials & Methods). This accessory bolted directly to the existing sample stage to reduce variation in position, had a permanent vacuum jacket that prevented frosting, and sealed

around the sample tube well enough to give greater than 30 minutes experiment time before refilling cryogen.

In fact, "hunting" for maxima was not necessary because, as a fluorophore with high quantum yield (i.e., a large proportion of absorbed light is emitted as fluorescence), the fluorescence excitation spectra of porphyrins can be summed to yield an absorption profile and peak maxima can be determined. A fluorescence excitation spectrum is the collection of emission intensity at a specified wavelength while a range of excitation wavelengths is scanned. Then, emission intensity is collected at regular intervals over a range of emission wavelengths that has been determined as an active region by previous emission experiments (see tabulated data above). The summed fluorescence excitation spectrum is then produced by adding up all of the emission intensity produced by a single excitation wavelength and plotting total emission intensity vs. the scanned excitation region. The result is analogous to an absorption spectrum.

a)

Zinc Protoporphyrin IX-substituted Myoglobin
in 3:1 glycerol:phosphate buffer
immersed in liquid nitrogen
Fluorescence Excitation Spectrum; λ_{em} = 605 nm

b)

Zinc Protoporphyrin IX-substituted Myoglobin
in 3:1 glycerol:phosphate buffer
immersed in liquid nitrogen
Fluorescence Excitation Spectrum; λ_{em} = 610 nm

c)

Zinc Protoporphyrin IX-substituted Myoglobin
in 3:1 glycerol:phosphate buffer
immersed in liquid nitrogen
Fluorescence Excitation Spectrum; λ_{em} = 615 nm

Figure 5.9. Fluorescence excitation spectra of zinc protoporphyrin IX-substituted myoglobin immersed in liquid nitrogen evolves dramatically as the emission window moves approximately 10 nm.

a) Emission at 605 nm, **b)** emission at 610 nm, **c)** emission at 615 nm.

118
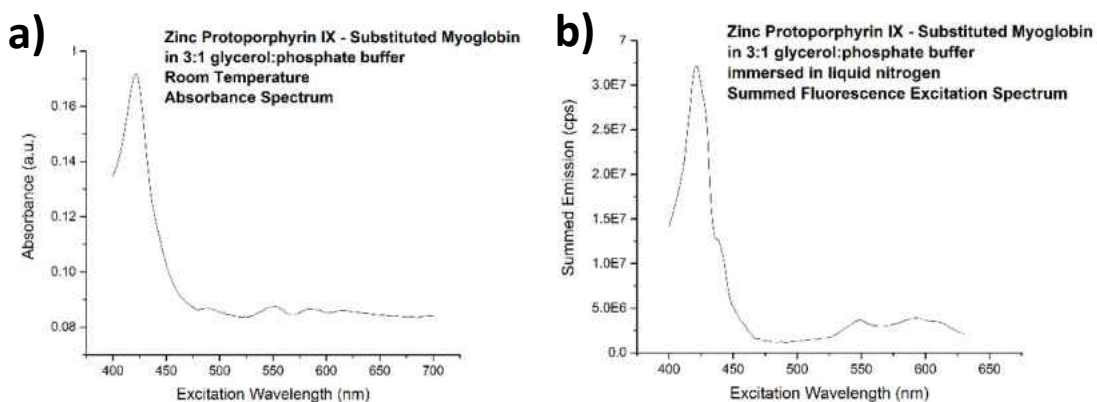
Figure 5.10. The summed fluorescence excitation method produces a comparable line shape to absorption spectroscopy. **a)** Absorbance spectrum of zinc protoporphyrin IX-substituted myoglobin at room temperature. **b)** Summed fluorescence excitation spectrum of the same species at low temperature.

For the initial experiments, emission was sampled at every 5 nm across the active emission region with an excitation scan at 1-nm increments. Later experiments sampled emission as well as excitation wavelengths in 1-nm increments. In this case, resolution, determined by the fluorometer's bandpass setting, needed additional consideration. The instrument contained two monochromators for excitation and two for emission (a "double/double monochromator"), each with a grating groove density of 1200 grooves / mm. This produced a linear dispersion of 2.1 nm spectral width per mm slit width. Bandpass is then given by:

Bandpass (nm) = slit width (mm) × dispersion (nm / mm)   (5.7)

Adjusting the slit width, then, allowed for consistent sampling. For a 2-nm bandpass, the slit-width was set to 0.95 mm for both emission and excitation. Using 1-nm increments, this produced overlap of 1 nm from one position to the next, essentially sampling each wavelength twice per trial and providing twice the intensity. At 5-nm increments, the resolution was poorer but

sufficient to provide the survey necessary to determine absorption maxima in low-temperature samples.

For seven species of interest for Stark spectroscopy experiments, the maximum position of the lowest-energy Q band while the sample was immersed in liquid nitrogen is given in Table 5.11.

Table 5.11. Peak maxima in summed fluorescence excitation spectra for seven porphyrin species provides the location of the lowest-energy Q band at low temperature.

Porphyrin Species	Maximum of Lowest-Energy Q Band (nm)	Solvent
PPIX	618.5	3:1 glycerol:aqueous buffer
PPIX DME	636.0	3:1 glycerol:ethanol
PPIX-myoglobin	620.0	3:1 glycerol:aqueous buffer
PPIX DME-myoglobin	624.0	3:1 glycerol:aqueous buffer
ZnPPIX	574.5	3:1 glycerol:aqueous buffer
ZnPPIX DME	580.5	3:1 glycerol:ethanol
ZnPPIX-myoglobin	581.0	3:1 glycerol:aqueous buffer

5.3.3.2 Band splitting in individual fluorescence excitation spectra

The individual fluorescence excitation spectra in Fig. 5.8 show some unexpected splitting in the Soret region that required further investigation. Splitting in the Soret and/or Q bands was present to various degrees in the initial fluorescence excitation spectra of each porphyrin species listed

in Table 5.11 when immersed in liquid nitrogen. As discussed in Section 5.1 and demonstrated in

Fig. 5.1, a typical porphyrin spectrum has a single Soret peak and a pattern of two Q peaks for

metalloporphyrins or four Q peaks for free-base porphyrins. Occasionally, a shoulder might

appear as part of the Soret band (see Fig. 5.9b), which has been attributed to a vibronic

sideband.[17] However, the complete resolution of two bands which independently grow and

shrink in intensity on the red and blue sides of the band depending on emission wavelength

seemed unusual, especially for well-studied protoporphyrin species. First, it was confirmed that

the special liquid nitrogen apparatus itself was not the source of the extra peak. Then, efforts to

establish precedence for this kind of spectral behavior in porphyrins were made. Finally, a

number of possible explanations were formulated.

Reports of splitting in the Soret peak as well as electronic Q peaks in spectra of porphyrins

certainly exist. The mechanisms were either further lifting the degeneracy of the transitions that

lay within a peak or reducing the inhomogeneous broadening that obscured details within a

band. The reports seem to be sorted into three general conditions where either the degeneracy

of the transitions in the peak were further lifted: 1) in dimers and aggregates,[18] 2) with large

substituents or ligands,[19] and 3) substituted into proteins.[20] Many examples also were also noted

under conditions of low temperature, some overlaying the spectra of many temperatures to note

exactly where splitting appeared. However, the access this "two-dimensional" spectral method

gave to the individual emission contributions of otherwise obscured transitions in the Soret

bands did not seem to be reported and was worth pursuing further.

A number of questions arose regarding the origin of the unusual spectral behavior. Did

the splitting arise due to low temperature or to the solid matrix it was confined in? Was it induced

by the glycerol–water solvent or might it occur in other amorphous solids? Was dimerization or aggregation causing splitting in the solution samples like some of the existing reports pointed to? Because that would indicate a mixture of species, could the protein-substituted samples contain contamination in the form of a mixture of substituted and unsubstituted proteins? Could pH induce two subpopulations with individual spectral signatures?

If the effect was independent of the contaminants and environments described above, which existing transitions were being resolved? Was it the nearly degenerate pair of B transitions in the Soret band? Were vibronic sidebands appearing out of the inhomogeneously broadened band?

5.3.3.3 Low-temperature fluorescence excitation spectra in various solvents

To test the solvent effect on the spectra, two amorphous-solid forming matrices were explored. First, a silica-based sol gel glass had been reported to encapsulate both porphyrins and proteins.[21] Sol gel glasses could provide an easily formed solid matrix at both room and low temperature to test whether the immobilization of the chromophore was the source of splitting, which might indicate coupling between them, or if low temperature was the necessary condition, which might indicate simply an increase in spectral resolution to reveal hidden bands. Second, dimethyl sulfoxide (DMSO) was known to both dissolve porphyrins readily at room temperature and to form an amorphous solid when frozen, so was also chosen as a solvent to compare results. For this comparison, free-base protoporphyrin IX and zinc protoporphyrin IX were the chromophores used.

5.3.3.3.1 Sol-gel glass matrix

Sol gel glass was prepared with the following procedure.[22] To a cell-culture tube, 7 mL of tetramethylorthosilicate ($Si(OCH_3)_4$, TMOS) was added along with 1.7 mL ultrapure water and 0.1 mL 0.04 M HCl. The tubes containing the TMOS mixture was placed in a test-tube rack and were sonicated in a room-temperature water bath for 20 min. During sonication, a pipette was used to mix the contents every 3 min. After sonication, the contents were transferred evenly into three, 1-cm, disposable methacrylate cuvettes. Then, 0.53 mL of a neutral solution containing the chromophore to be studied was added. The sol gel solution was left to solidify overnight at 4 °C. This procedure yielded a total of about 9 mL of solution, divided into 3 − 3-mL aliquots. The dielectric constant of the solidified sol gel glass is very low, likely less than 5.[23]

Fluorescence excitation spectra for this experiment did not show any splitting for PPIX or ZnPPIX in sol gel at room temperature. The room-temperature spectra were taken in the transparent methacrylate cuvettes. However, when both samples were crushed and placed in the liquid nitrogen dewar assembly, splitting was revealed when immersed in liquid nitrogen. From these experiments, it was concluded that low temperatures appears to resolve hidden spectral bands.

5.3.3.3.2 DMSO

Similar spectra were taken of PPIX and ZnPPIX dissolved in DMSO. Again, room temperature spectra did not show splitting when a sufficiently dilute concentration was used (10^{-6} M) but the same samples at low temperature showed splitting in B and/or Q bands. It was concluded that the effect was not solvent dependent and again reinforced the idea that low temperature was improving resolution.

More concentrated solutions were also mistakenly created and analyzed, producing spectra with peaks of intensity beyond the typical porphyrin pattern as well as extra peaks (see Fig. 5.11). This indicated that dimerization and aggregation might be a player in the peak splitting behavior and prompted the dilution experiments described next.
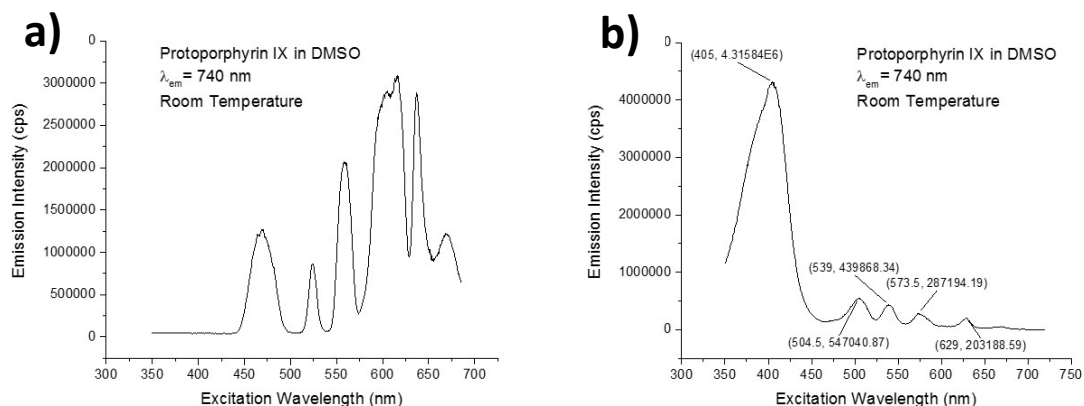
Figure 5.11. Illustration of concentration effect on fluorescence excitation spectra of PPIX in DMSO, even at room temperature. Both spectra were taken monitoring the same emission wavelength, 740 nm. **a)** M = 7.46 × 10^{-4}; **b)** M = 7.46 × 10^{-6}.

5.3.3.4 Low-temperature fluorescence excitation spectra at various concentrations

Interestingly, a literature search revealed a claim by White in "The Porphyrins" that porphyrins tend to aggregate only in aqueous solution, suggesting that the low dielectric constant of organic solvents prevents the process.[9] The source also noted that zinc porphyrins in particular resist aggregation. However, the experiments performed here seem to contradict that. To determine the concentrations at which PPIX and ZnPPIX tend to dimerize and/or aggregate, a spectral method was devised invoking the linearity predicted by Beer's Law, about which White references deviation as a sign of aggregation.

The spectra in these experiments were taken at room temperature. Both absorbance and fluorescence spectra were collected. The comparison of the two spectral methods further showed that fluorescence excitation can be a more sensitive tool, showing changes with concentration that a simple absorbance spectrum did not reveal. The spectral series also used two solvents of similar dielectric constant, 3:1 glycerol: water and DMSO, to see if the aggregation process was solvent dependent.

Solutions of PPIX and ZnPPIX in 3:1 glycerol:water and in DMSO at a known high concentration, ~7×10^{-4} M, were made precisely in a volumetric flask. Then, the original solution was serially diluted by a factor of ten until an extremely dilute concentration of 7×10^{-11} M was created, far past the point of visible color in the deeply colored porphyrin solutions. The micromolar range was then tested with more resolution, taking the 10^{-6} M solutions and diluting by a factor of two until reaching 10^{-7} M.

Quantitative comparison of the spectra's linearity was done by referencing the Soret peak intensity in each run. Then, the absorbance (a.u.) or emission (cps) intensity value of that peak was plotted against the porphyrin concentration. The points were fit using the linear regression tool in Microsoft Excel, adding points of increasing concentration to determine where deviation from linearity occurred. Based on these results, Table 5.12 recommends maximum working concentrations for porphyrin–solvent combinations to avoid apparent dimerization or aggregation. In general, approximately 1×10^{-6} M seems to be a concentration where the porphyrin population exists in monomer form.

Table 5.12. Recommended working concentrations of PPIX and ZnPPIX for spectral experiments to avoid dimerization or aggregation.

Chromophore	Solvent	Maximum Concentration (M)
PPIX	3:1 glycerol:buffer	3.75×10^{-6}
PPIX	DMSO	1.865×10^{-6}
ZnPPIX	3:1 glycerol:buffer	1.865×10^{-6}
ZnPPIX	DMSO	1.865×10^{-6}

Finally, a series of similar spectra were taken with PPIX-substituted myoglobin and ZnPPIX-substituted myoglobin in 3:1 glycerol:water. It was not expected that these species would be affected by aggregation as pi-stacking would not be possible like in solution but the possibility of some other interaction between individual proteins was to be explored. The results showed linearity of Soret peak maxima with concentration and concentrations of porphyrin up to 10^{-4} M showed no additional peaks, so it was deduced that proteins either did not aggregate at the tested concentrations or that any aggregation did not affect the porphyrin spectra.

5.3.3.5 Low-temperature fluorescence excitation spectra with contaminants

Having ruled out aggregation in porphyrin-substituted myoglobin, the possibility of contamination was explored. As defined earlier, "contamination" in the context of porphyrin-substituted proteins means porphyrins existing in solution outside of a properly substituted protein matrix. In Ch. 4, the substitution procedure used a ratio of 1 porphyrin : 2 apo-protein in

an effort to avoid and porphyrin remaining in solution. Here, the Soret band was made the focus spectral area.

5.3.3.5.1   Experiments with PPIX DME - myoglobin

The following experiments used PPIX DME–substituted myoglobin for two reasons: 1) the substituted protein is analyzed in aqueous solution but the chromophore itself is not water soluble, making any unsubstituted porphyrin easier to filter out in the substitution process; 2) the capped propionate groups in the dimethyl ester porphyrin took away the possibility that the pH conditions of the solvent would affect the chromophore. A series of fluorescence excitation spectra taken at 1-nm excitation and 1-nm emission increments was taken of PPIX DME–myoglobin in 3:1 glycerol:buffer at neutral pH immersed in liquid nitrogen. See Fig. 5.12 for a contour plot of the experiment; contour colors indicate emission intensity at that combination of excitation and emission energies. More apparent in Fig. 5.13a is that four excitation maxima are noted and individual emission scans at these four wavelengths in Fig. 5.13b can reveal possible sources.
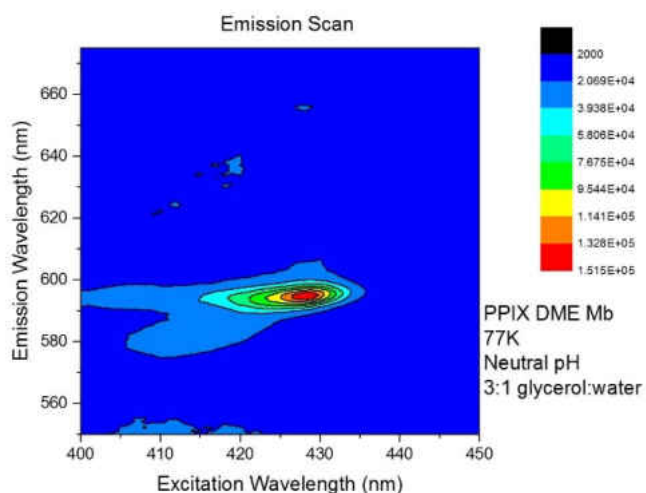
Figure 5.12. Compilation of emission scans taken at 1-nm excitation wavelength increments of PPIX DME–substituted myoglobin in 3:1 glycerol:buffer at neutral pH immersed in liquid nitrogen. Red indicates a peak maximum.

a) Protoporphyrin IX Dimethyl Ester-Subsituted Myoglobin 2D Fluorescence Excitation Plot at neutral pH

428 nm

415 nm

406 nm

438 nm

b) Protoporphyrin IX Dimethyl Ester in Myoglobin
77K
Neutral pH
3:1 Glycerol:Water

406
415
428
438
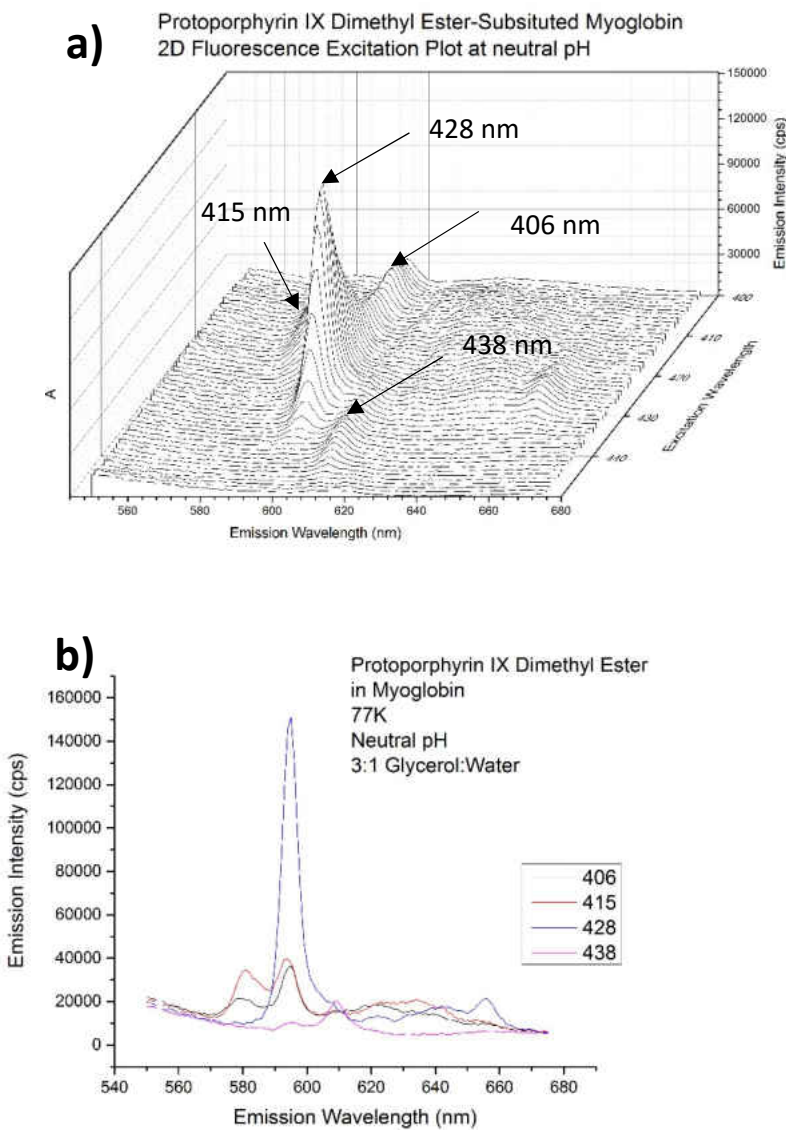
Figure 5.13. PPIX DME-myoglobin at neutral pH. **a)** Another view of fluorescence excitation with two frequency axes. Four excitation maxima are noted with arrows: 406 nm, 415 nm, 428 nm, and 438 nm. **b)** Emission spectra at the four excitation maxima with individual colors (legend lists wavelength in nm).

The first attempt to intentionally contaminate a sample of PPIX DME–myoglobin was to add water-insoluble PPIX DME to a solution of PPIX DME–myoglobin in aqueous buffer. The result was deep-purple particles suspended in solution; this mixture was further mixed into glycerol for a final 3:1 ratio. While particles remained suspended, the mixture was immersed in liquid nitrogen. A sample without any protein was created in a similar manner for comparison. Fig. 5.14 shows the two samples' compiled emission spectra; compared to Fig. 5.12, any unsubstituted PPIX DME, assuming it took on a particulate form during the long substitution process, has a quite

distinct spectral signature. The particulate porphyrin spectrum lacks the splitting features seen in the protein and shows a ridge constant across the Soret band at an emission wavelength of approximately 680 nm, well out of the maximum emission range of the substituted protein. It was concluded that unsubstituted PPIX DME was not present in typical preparations of PPIX DME–myoglobin and that the splitting noted in the original sample was not due to unsubstituted porphyrin.
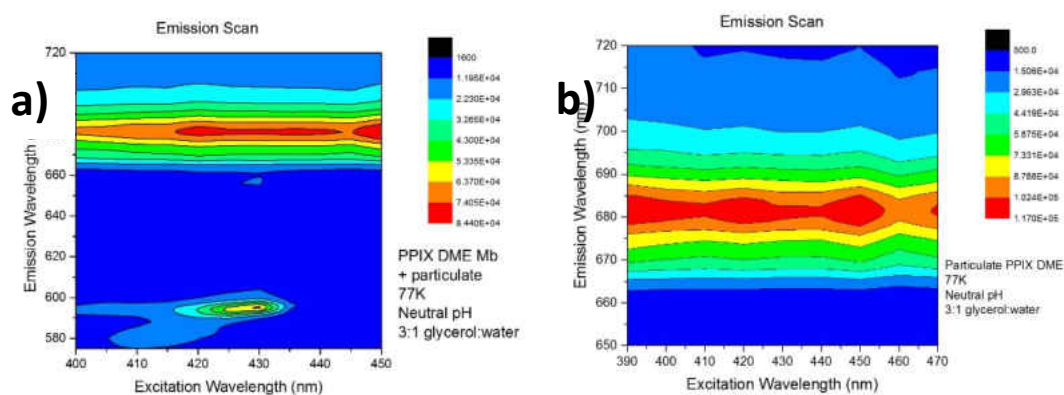
Figure 5.14. The spectral signature of particulate PPIX DME in aqueous solution. **a)** Combination of PPIX DME-substituted myoglobin (spectral signature on the lower half of the plot) and particulate PPIX DME (upper half of the plot); **b)** Particulate PPIX DME alone.

The second attempt to recreate a "contaminated" sample involved inducing denaturation of PPIX DME-substituted myoglobin so a sample of incorrectly substituted protein might be assessed. Denaturation was induced by lowering the protein sample to pH = 2 by adding concentrated 1 M HCl dropwise into the same PPIX DME-myoglobin in 3:1 glycerol:water. The sample was quickly immersed in liquid nitrogen. It was expected that a mixture of species would be created. The resulting emission scans compiled in Fig. 5.15 compared to the solution at neutral pH revealed that the sample is indeed a mixture.
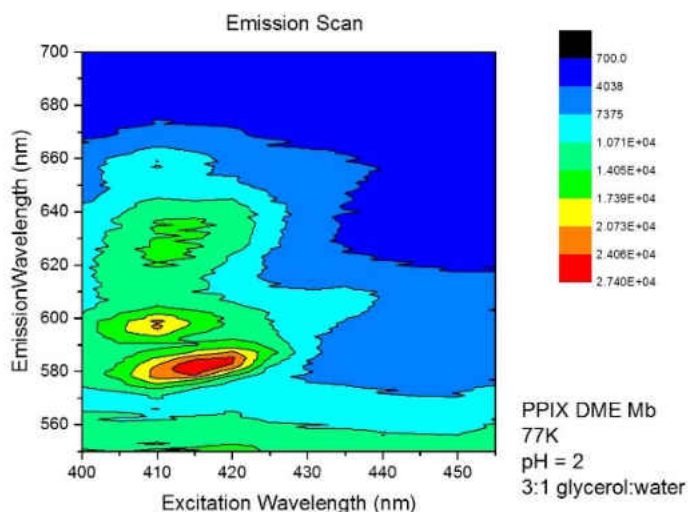
Figure 5.15. Contour plot of PPIX DME-myoglobin in 3:1 glycerol:buffer at pH 2.

The global emission maximum moves along both frequency axes to higher energy approximately 15 nm at $\lambda_{ex, max}$ = 415 nm and $\lambda_{em, max}$ = 580 nm along with a maximum of similar magnitude at $\lambda_{ex, max}$ = 410 nm and $\lambda_{em, max}$ = 595 nm. A much smaller local maximum does indeed appear in the low pH plot at approximately the position of the neutral pH maximum ($\lambda_{ex, max}$ = 428 nm, $\lambda_{em, max}$ = 595 nm), revealing that proteins with conformations as at neutral pH exist in solution, but incorrectly substituted/denatured proteins dominate the plot at low pH. It is interesting to note that these new maxima are unlike the spectra produced by the particulate plots, indicating that the solution created by denatured protein has an influence of its own on the porphyrins. A comparison with the Fig. 5.13b shows that the low-pH maxima correspond well with the peaks produced with excitation at 415 nm. This seems to indicate that the change in conformation induced by low pH also exists to some extent in typical preparation of PPIX DME-myoglobin. As a result further efforts were taken to separate different conformations of the substituted protein with a size-exclusion column, but the results of this single effort seemed to strip the porphyrin from the protein. More attempts at this purification step need to be made.

5.3.3.5.2 Experiments with PPIX-myoglobin

After the apparent instability of PPIX DME-myoglobin post-purification column, PPIX-myoglobin preparations were undertaken again. This time, to further explore the effect of unsubstituted porphyrin on spectra, they were done with varying ratios of porphyrin to apo-protein: A) 2 porphyrin : 1 apo-myoglobin; B) 1 porphyrin : 1 apo-protein; C) 1 porphyrin : 2 apo-protein. (Sample C represents a typical preparation of substituted porphyrin in this work.) The expectation was that excess, unsubstituted porphyrin would produce a clear spectral signature. Here, fluorescence excitation data in Fig. 5.16 is simply the Soret region overlaid with a single frequency axis.
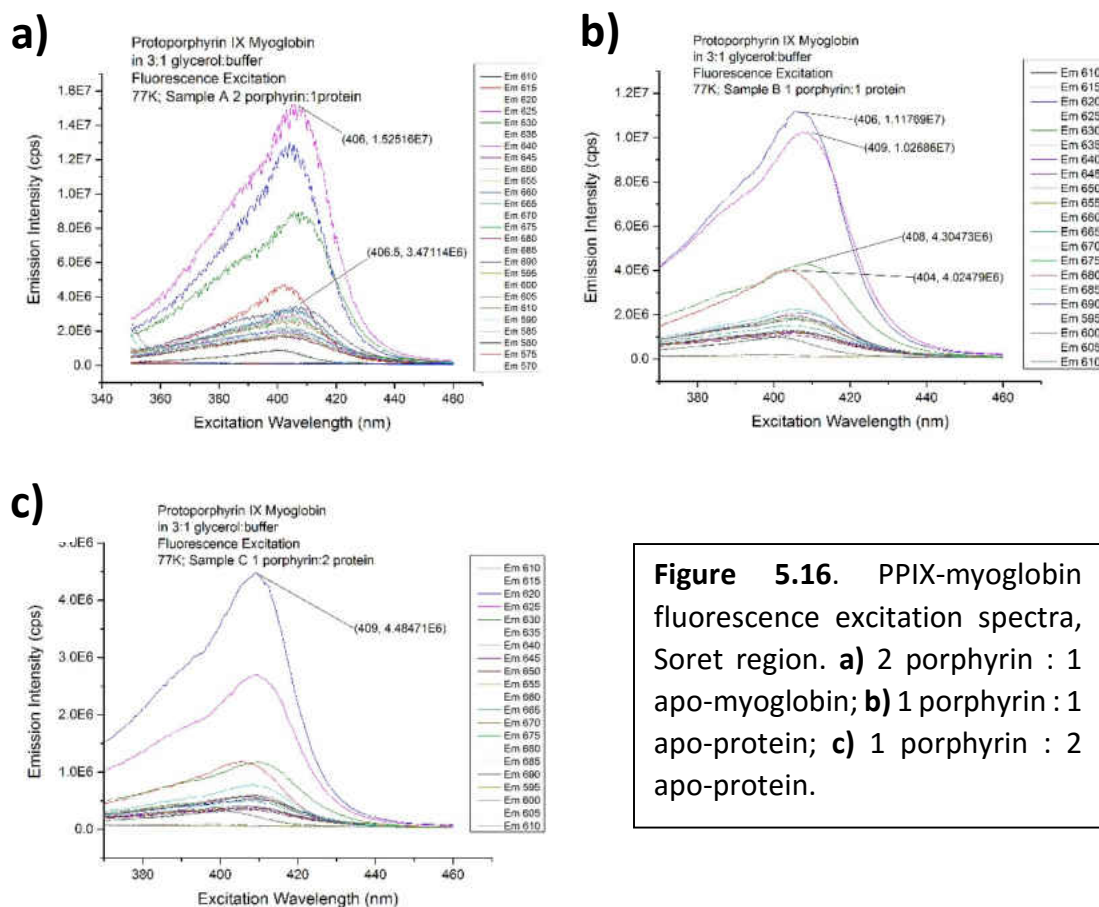
Figure 5.16. PPIX-myoglobin fluorescence excitation spectra, Soret region. **a)** 2 porphyrin : 1 apo-myoglobin; **b)** 1 porphyrin : 1 apo-protein; **c)** 1 porphyrin : 2 apo-protein.

First, no splitting is noticed in the Soret region with higher resolution data, which is consistent with earlier, lower resolution data that showed splitting only in the Q_x-band region for PPIX-myoglobin. This may indicate that PPIX-myoglobin has a very wide distribution of "correctly substituted" species, unlike PPIX DME-myoglobin, which seemed to have distinct peaks for "correct" and "incorrect" substitutions. Both of these substituted proteins would benefit from a crystal structure to clarify the way the porphyrin sits in the myoglobin heme cavity and if there is a distribution of conformations.

A surprising observation is that there is no additional band for Sample A, even though there should be an equal amount of substituted and unsubstituted porphyrin. Free-base protoporphyrin IX in glycerol–water solution typically has an absorption maximum at 397 nm, which, with a lower molar absorptivity coefficient than the substituted protein, may be hidden beneath the broadened Soret band. The unsubstituted porphyrin may also have a shifted maximum as the dissolved protein may alter the solution's dielectric constant enough to shift the Soret peak closer to that of the substituted protein.

Another surprising observation is the non-linearity of the Soret band intensity. With a mixture of substituted and unsubstituted porphyrin, Sample A is not expected to be linear in the intensification of its Soret band over Samples B and C. However, between Samples B and C, there should only be substituted porphyrins and twice the amount in B as in C; however, the Soret band intensifies approximately two and a half times between the two samples at its peak. It may be interesting to further investigate how apo-protein affects the solution characteristics over holo-protein.

5.3.3.5.3   Experiments with ZnPPIX-myoglobin

Finally the above experiment with PPIX-myoglobin was repeated with ZnPPIX –myoglobin with Samples A, B, and C prepared in the same manner (Fig. 5.17):
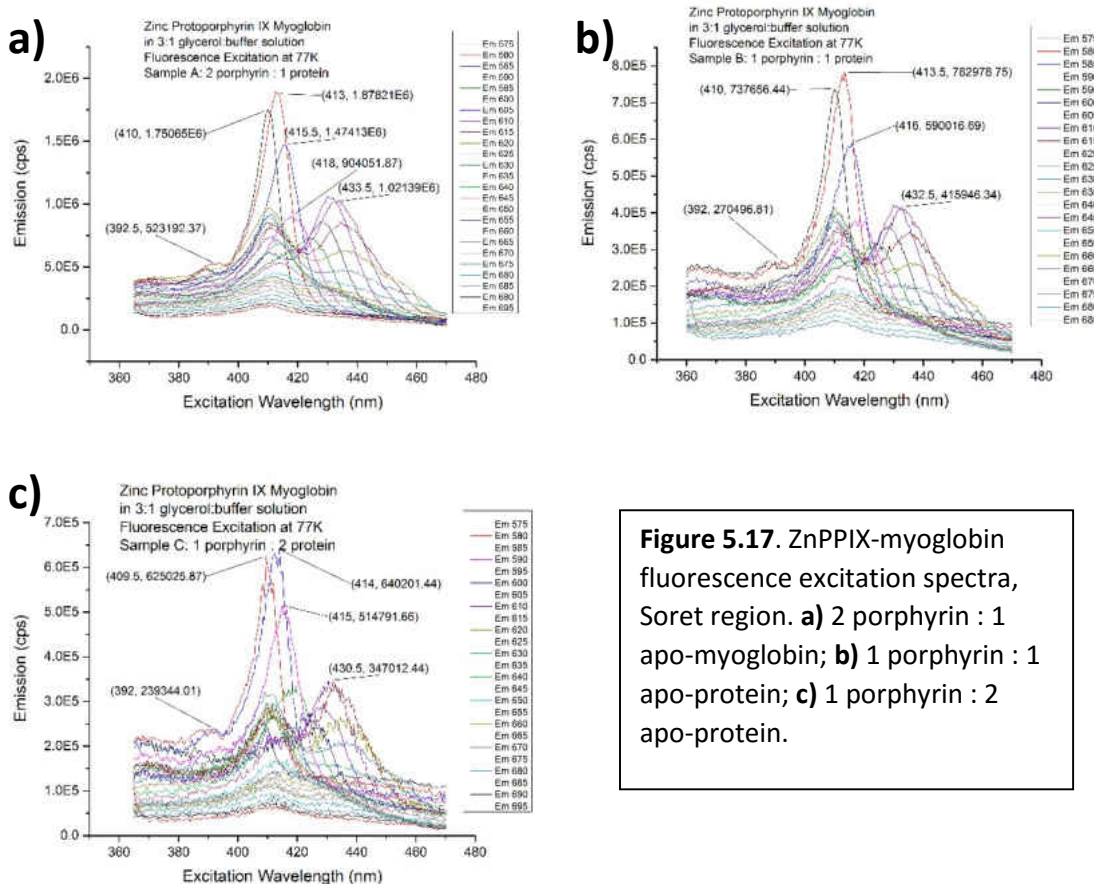
Figure 5.17. ZnPPIX-myoglobin fluorescence excitation spectra, Soret region. **a)** 2 porphyrin : 1 apo-myoglobin; **b)** 1 porphyrin : 1 apo-protein; **c)** 1 porphyrin : 2 apo-protein.

Contrary to the PPIX-myoglobin experiment, there is clear splitting in this sample. Although crystal structures are not available for either, it can be assumed that the protein is covalently attached to the porphyrin through the zinc atom, while that option is not available for the free-base porphyrins. However, like the PPIX-myoglobin sample, there is also no extra peak that appears when there is porphyrin in stoichiometric excess of protein in Sample A. In fact, the samples look qualitatively very similar. However, in glycerol–water solution, the solution

chromophore should peak about 410 nm, which, again, might be obured in these plots. Another

possible explanation is that the porphyrins essentially dimerize inside of the substituted protein,

giving a ratio of 2 porphyrins per myoglobin.

To test this, ZnPPIX in glycerol–water solution was added on top of the existing Sample B

with twice the amount of free porphyrin to porphyrin-substituted protein. The homogeneous

solution was mixed but not allowed to equilibrate for any length of time so the possibility of

substitution was minimized (Fig. 5.18). Here an additional peak does appear at 405 nm and, in

fact, dominates! This indicates that, for ZnPPIX–myoglobin, unsubstituted porphyrin does not

seem to be a major contaminant.
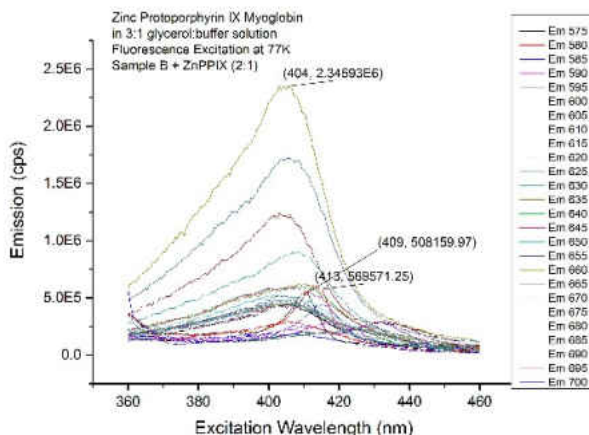
Figure 5.18. ZnPPIX-myoglobin fluorescence excitation spectra, Soret region, with additional, unequilibrated solution porphyrin.

Finally, a denaturation was performed, but this time thermally, since the propionate side

groups in protoporphyrin IX would protonate at low pH. An aliquot of Sample C was heated to

approximately 90 °C for several minutes and immediately transferred to the liquid-nitrogen

sample holder and brought to low temperature by immersion. The denaturation is apparent in

Fig. 5.19, with the peak at 405 nm again dominating.

Zinc Protoporphyrin IX Myoglobin
in 3:1 glycerol:buffer solution
Fluorescence Excitation at 77K
Sample C thermally denatured

(405, 830370.25)
(419.5, 644092.75)
(439, 418712)
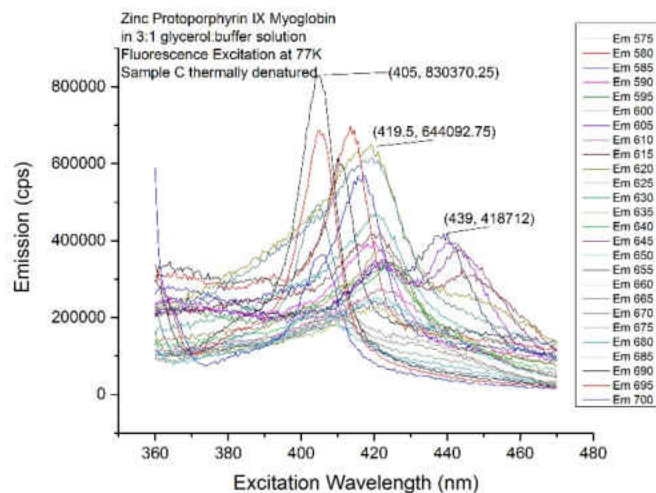
Figure 5.19. ZnPPIX-myoglobin fluorescence excitation spectra, Soret region, Sample C subject to thermal denaturation.

Some comparison between species of ZnPPIX in glycerol–water solutions at low temperature are found in Fig. 5.20. It is clear that protein-substitution is required for Soret-peak splitting in ZnPPIX as well as low temperature. In addition, it shows that the preparations (Samples A, B, and C) that were allowed to equilibrate with porphyrin in apo-protein, have similar shapes distinct from other treatments.
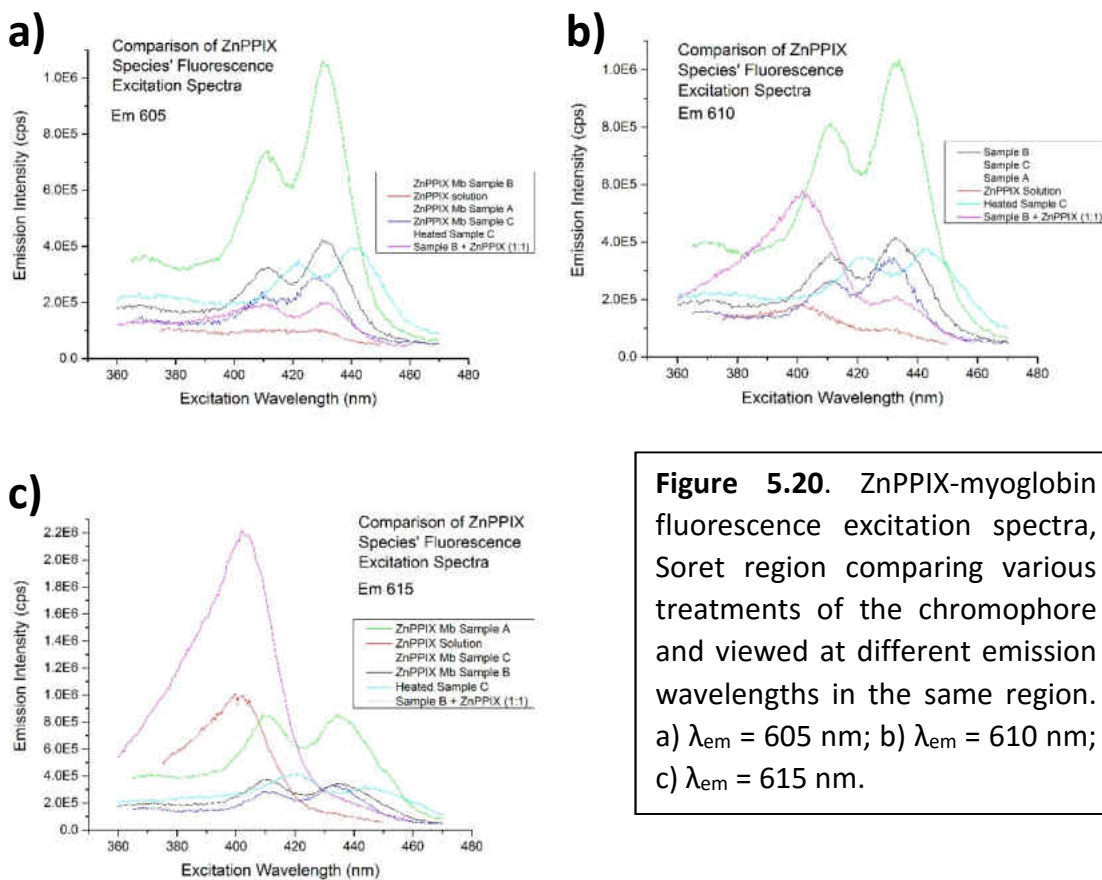
Figure 5.20. ZnPPIX-myoglobin fluorescence excitation spectra, Soret region comparing various treatments of the chromophore and viewed at different emission wavelengths in the same region. a) λ_{em} = 605 nm; b) λ_{em} = 610 nm; c) λ_{em} = 615 nm.

5.3.3.6 Possible explanation for Soret-band splitting in ZnPPIX–myoglobin

Further purification of ZnPPIX–myoglobin Sample C on a size-exclusion column produced a single peak, indicating that aggregation of apo-protein in this sample was not present in large amounts. This, along with the spectral data presented in the last section, seems to rule out contamination as a source of splitting, at least for ZnPPIX–myoglobin. The assessment from Fig. 5.21 below points to the resolution of vibronic sidebands as a source of splitting in this region. Like Fig. 5.13b, Fig. 5.21 is the emission spectra produced at each of five excitation maxima from individual fluorescence excitation spectra. For this species, the emission maxima occur at similar positions, indicating that radiative relaxation shares points of energetic origin, as tabulated in Table 5.13.
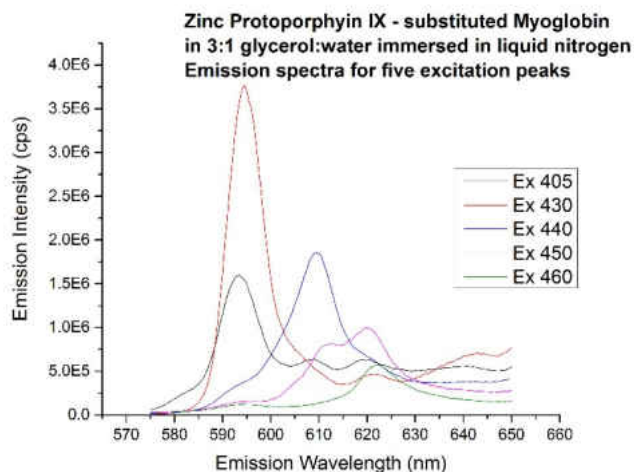
Figure 5.21. ZnPPIX-myoglobin emission pattern comparison in Soret region for five excitation peaks (wavelength in nm).

Table 5.13. Emission maxima of five peak excitation positions for ZnPPIX–myoglobin.

Excitation Peak Position (nm)	Emission Peak 1 (nm)	Emission Peak 2 (nm)	Emission Peak 3 (nm)	Emission Peak 4 (nm)
405	593.5	609.0	620.5	640.5
430	594.5		621.5	643.0
440	*shoulder	609.5	*shoulder	
450	593.5	612.5	620.0	
460	594		622	

5.4   Conclusions

The peak maxima for the absorption and emission spectra of a number of porphyrins in a variety of environments have been tabulated in support of future hole-burning Stark experiments with high resolution but very narrow spectral windows. The porphyrin species tested are protoporphyrin IX (PPIX), zinc protoporphyrin IX (ZnPPIX), protoporphyrin IX dimethyl ester (PPIX DME), zinc protoporphyrin IX dimethyl ester (ZnPPIX DME), and tin protoporphyrin IX dimethyl ester (SnPPIX DME). Both free solution porphyrins as well as myoglobin-substituted porphyrins were tested. The solvents used were water and 3:1 glycerol:water mixture for water-soluble

chromophores; ethanol, chloroform, and their glycerol mixtures for non-polar chromophores; and neutral phosphate buffer and its 3:1 glycerol:buffer mixture for proteins.

Because hole-burning Stark spectroscopy is performed at very low temperatures, spectra of the porphyrin species and environments listed above were also taken while immersed in liquid nitrogen. A method was devised to accomplish these conditions using a commercial instrument. Unusual spectra that revealed splitting in some bands of porphyrins in the UV–VIS region, especially the Soret band, at low-temperature prompted further exploration into the cause. It was determined that porphyrins in solution exhibited band splitting at higher concentrations and was resolved at lower micromolar concentrations. However, band splitting in porphyrin-substituted myoglobin persisted at all concentrations. The cause of splitting seemed to vary based on the porphyrin. PPIX–myoglobin did not show any Soret-band splitting but did exhibit some Q-band splitting. PPIX DME–myoglobin appeared to have Soret splitting at least partially as a results of a mixture of species, i.e., different conformations of the porphyrin substituted into the protein. ZnPPIX–myoglobin, however, seemed to be a sample with a single species and Soret-band splitting appears to be simply the resolution of vibronic bands associated with the electronic Soret transitions.

5.5 References

1. Gouterman, M., Spectra of porphyrins. *Journal of Molecular Spectroscopy* **1961,** *6* (0), 138-163.

2. (a) Gouterman, M., Study of the Effects of Substitution on the Absorption Spectra of Porphin. *The Journal of Chemical Physics* **1959,** *30* (5), 1139-1161; (b) Gouterman, M.; Wagniere, G. H.; Snyder, L. C., Spectra of porphyrins: Part II. Four orbital model. *Journal of Molecular Spectroscopy* **1963,** *11* (1-6), 108-127.

3. Hashimoto, T.; Choe, Y.-K.; Nakano, H.; Hirao, K., Theoretical Study of the Q and B Bands of Free-Base, Magnesium, and Zinc Porphyrins, and Their Derivatives. *The Journal of Physical Chemistry A* **1999,** *103* (12), 1894-1904.

4. (a) Weigl, J. W., The polarization of the fluorescence of tetraphenylporphine. *Journal of Molecular Spectroscopy* **1957,** *1* (1–4), 133-138; (b) Rimington, C.; Mason, S. F.; Kennard, O., Porphin. *Spectrochimica Acta* **1958,** *12* (1), 65-77; (c) Gouterman, M.; Stryer, L., Fluorescence Polarization of Some Porphyrins. *The Journal of Chemical Physics* **1962,** *37* (10), 2260-2266; (d) Sevchenko, A. N.; Gurinovich, G. P.; Solev'ev, K. N., Some Features of the Polarization of the Fluorescence of Porphyrins. *Soviet Physics - Doklady* **1961,** *5*, 808-811.

5. (a) Friedrich, J., Hole Burning Spectroscopy and Physics of Proteins. In *Methods in Enzymology*, Sauer, K., Ed. Academic Press: 1995; Vol. Volume 246, pp 226-259; (b) Friedrich, J.; Haarer, D., Photochemical Hole Burning: A Spectroscopic Study of Relaxation Processes in Polymers and Glasses. *Angew. Chem., Int. Ed. Engl.* **1984,** *23* (2), 113-140.

6. Geissinger, P.; Kohler, B. E.; Kulikov, S. G.; Terpougov, V., Inhomogeneous Broadening of Electronic Spectra in Organic Model Systems. *J. Chem. Phys.* **1998,** *108* (5), 1821-1829.
7. Geissinger, P.; Woehl, J. C.; Kohler, B., Internal electric fields and inhomogeneous broadening: protoporphyrin-IX in myoglobin. *Molecular Crystals Liquid Crystals* **1996,** *283*, 69-74.
8. Geissinger, P.; Kohler, B. E.; Woehl, J. C., Electric Field and Structure in the Myoglobin Heme Pocket. *J. Phys. Chem.* **1995,** *99*, 16527-16529.

9. White, W. I., Aggregation of Porphyrins and Metalloporphyrins. In *The Porphyrins*, Dolphin, D., Ed. Academic Press: New York, 1978; Vol. 5, pp 303-339.

10. Breslow, E.; Koehler, R., Combination of Protoporphyrin IX with Sperm Whale Apomyoglobin. *Journal of Biological Chemistry* **1965,** *240* (5), PC2266-PC2268.

11. Woehl, J. C. Measuring Internal Electrostatic Fields and Potentials at Molecular and Atomic Resolution using Hole-Burning Spectroscopy. University of California, Riverside, Riverside, CA, 1996.

12.   Leonard, J. J.; Yonetani, T.; Callis, J. B., Fluorescence study of hybrid hemoglobins containing free base and zinc protoporphyrin IX. *Biochemistry* **1974,** *13* (7), 1460-1464.

13.   Tamura, M.; Asakura, T.; Yonetani, T., Heme modification studies of myoglobin: I. Purification and some optical and EPR characteristics of synthesized myoglobins containing unnatural hemes. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **1973,** *295* (2), 467-479.

14.   G09 User's Guide: SCRF keyword. In *http://gaussian.com/g_tech/g_ur/k_scrf.htm*, Gaussian, I., Ed. 2009.

15.   Albright, P. S., Experimental Tests of Recent Theories Descriptive of the Salting-out Effect1. *Journal of the American Chemical Society* **1937,** *59* (11), 2098-2104.

16.   Andrews, S. S.; Boxer, S. G., A liquid nitrogen immersion cryostat for optical measurements. *Review of Scientific Instruments* **2000,** *71* (9), 3567-3569.

17.   Gouterman, M., In *Optical Spectra and Electronic Structure of Porphyrins and Related Rings* [Online] Dolphin, D., Ed. Academic Press: New York, 1978; p. 165.

18.   (a) Koch, F.; Kullmann, M.; Selig, U.; Nuernberger, P.; Gotz, D. C. G.; Bringmann, G.; Brixner, T., Coherent two-dimensional electronic spectroscopy in the Soret band of a chiral porphyrin dimer. *New Journal of Physics* **2013,** *15* (2), 025006; (b) Zazubovich, V., Fluorescence Line Narrowing and delta-FLN Spectra in the Presence of Excitation Energy Transfer between Weakly Coupled Chromophores. *The Journal of Physical Chemistry B* **2014**; (c) Matile, S.; Berova, N.; Nakanishi, K.; Fleischhauer, J. r.; Woody, R. W., Structural Studies by Exciton Coupled Circular Dichroism over a Large Distance:â€‰ Porphyrin Derivatives of Steroids, Dimeric Steroids, and Brevetoxin Bâ€¥. *Journal of the American Chemical Society* **1996,** *118* (22), 5198-5206; (d) Prieto, I.; Pedrosa, J. M.; MartÃn-Romero, M. T.; MÃ¶bius, D.; Camacho, L., Characterization and Structure of Molecular Aggregates of a Tetracationic Porphyrin in LB Films with a Lipid Anchor. *The Journal of Physical Chemistry B* **2000,** *104* (43), 9966-9972.

19.   (a) Mandoj, F.; Nardis, S.; Pudi, R.; Lvova, L.; Fronczek, F. R.; Smith, K. M.; Prodi, L.; Genovese, D.; Paolesse, R., Beta-Pyrazino-fused tetrarylporphyrins. *Dyes and Pigments* **2013,** *99* (1), 136-143; (b) Kee, H. L.; Kirmaier, C.; Tang, Q.; Diers, J. R.; Muthiah, C.; Taniguchi, M.; Laha, J. K.; Ptaszek, M.; Lindsey, J. S.; Bocian, D. F.; Holten, D., Effects of Substituents on Synthetic Analogs of Chlorophylls. Part 2: Redox Properties, Optical Spectra and Electronic Structure. *Photochemistry and Photobiology* **2007,** *83* (5), 1125-1143; (c) de Miguel, G.; Hosomizu, K.; Umeyama, T.; Matano, Y.; Imahori, H.; Martín-Romero, M. T.; Camacho, L., Tunable Soret-Band Splitting of an Amphiphilic Porphyrin by Surface Pressure. *ChemPhysChem* **2008,** *9* (11), 1511-1513; (d) Mathew, S.; Yella, A.; Gao, P.; Humphry-Baker, R.; CurchodBasile, F. E.; Ashari-Astani, N.; Tavernelli, I.; Rothlisberger, U.; NazeeruddinMd, K.; Grätzel, M., Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers. *Nat*

Chem **2014,** *6* (3), 242-247; (e) Clezy, P. S.; Morell, D. B., A spectroscopic study of haematin compounds in the Soret region. *Biochimica et Biophysica Acta* **1963,** *71* (0), 165-171.

20.   (a) Sudhakar, K.; Loe, S.; Yonetani, T.; Vanderkooi, J. M., Fluorescent derivatives of human hemoglobin. Differences in interaction of the porphyrin with the protein between the alpha and beta subunits. *Journal of Biological Chemistry* **1994,** *269* (37), 23095-23101; (b) Balog, E.; Kis-Petik, K.; Fidy, J.; Kohler, M.; Friedrich, J., Interpretation of Multiple Q(0,0) Bands in the Absorption Spectrum of Mg-Mesoporphyrin Embedded in Horseradish Peroxidase. *Biophys. J.* **1997,** *73* (1), 397-405; (c) Sudhakar, K.; Laberge, M.; Tsuneshige, A.; Vanderkooi, J. M., Zinc-Substituted Hemoglobins: Alpha- and Beta-Chain Differences Monitored by High-Resolution Emission Spectroscopy. *Biochemistry* **1998,** *37* (20), 7177-7184; (d) Shibata, Y.; Kurita, A.; Kushida, T., Real-Time Observation of Conformational Fluctuations in Zn-Substituted Myoglobin by Time-Resolved Transient Hole-Burning Spectroscopy. *Biophysical Journal* **1998,** *75* (1), 521-527; (e) Shibata, Y.; Kushida, T., Determination of Q_x- and Q_y-Absorption Bands of Zn-Porphyrin Derivatives Contained in Proteins by Hole-Burning Spectroscopy. *Chem. Phys. Lett.* **1998,** *284* (1), 115-120; (f) Manas, E. S.; Wright, W. W.; Sharp, K. A.; Friedrich, J.; Vanderkooi, J. M., The Influence of Protein Environment on the Low Temperature Electronic Spectroscopy of Zn-Substituted Cytochrome c. *The Journal of Physical Chemistry B* **2000,** *104* (29), 6932-6941; (g) Reddy, K. S.; Angiolillo, P. J.; Wright, W. W.; Laberge, M.; Vanderkooi, J. M., Spectral Splitting in the α (Q0,0) Absorption Band of Ferrous Cytochrome c and Other Heme Proteins. *Biochemistry* **1996,** *35* (39), 12820-12830.

21.   (a) Kulikov, S. G.; Veret-Lemarinier, A. V.; Galaup, J. P.; Chaput, F.; Boilot, J. P., Sol-Gel Hosts Doped with Porphyrin Derivatives. Part I. Spectroscopy, Hole-Burning and Spectral Diffusion. *Chem. Phys.* **1997,** *216* (1), 147-161; (b) Levantino, M.; Cupane, A.; Zimanyi, L., Quaternary Structure Dependence of Kinetic Hole Burning and Conformational Substates Interconversion in Hemoglobin *Biochemistry* **2003,** *42* (15), 4499-4505.

22.   (a) Ellerby, L.; Nishida, C.; Nishida, F.; Yamanaka, S.; Dunn, B.; Valentine, J.; Zink, J., Encapsulation of proteins in transparent porous silicate glasses prepared by the sol-gel method. *Science* **1992,** *255* (5048), 1113-1115; (b) Dave, B. C.; Soyez, H.; Miller, J. M.; Dunn, B.; Valentine, J. S.; Zink, J. I., Synthesis of Protein-Doped Sol-Gel SiO2 Thin Films: Evidence for Rotational Mobility of Encapsulated Cytochrome c. *Chemistry of Materials* **1995,** *7* (8), 1431-1434.

23.   Kim, Y.; Kang, E.; Kwon, Y. S.; Cho, W. J.; Cho, C.; Chang, M.; Ree, M.; Chang, T.; Ha, C. S., Electrical properties of silica-polyimide composite dielectric thin films prepared via sol-gel reaction and thermal imidization. *Synthetic Metals* **1997,** *85* (1–3), 1399-1400.

Chapter 6

Hole-Burning Stark Spectroscopy

6.1 Introduction

The experimental measurement of internal electric fields in heme proteins has been conducted using Stark spectroscopy, which is the application of an external electric field while a spectrum is collected. The Stark shift that results in the porphyrin spectrum allows for the measurement of a change of the energy levels of the chromophore. From this, the internal electric field vector can be extracted mathematically. However, as described in Chapter 5, the significant inhomogeneous broadening present in porphyrin spectra obscures the Stark shift. As a means of increasing resolution to measure the Stark shift, spectral hole burning is employed.

The theoretical underpinnings of both Stark spectroscopy as well as spectral hole burning will be presented here in general terms. Further details on the quantum-mechanical Stark analysis, which this work has helped to advance, will be discussed in Chapter 7. Then, the experimental setup for Stark hole-burning experiments will be described. In addition, a related measurement of the dielectric constant of low-temperature glasses will be considered. Finally, experimental spectra of PPIX–myoglobin obtained with the Stark hole-burning apparatus will be presented.

6.2   Theory

6.2.1   Stark spectroscopy

When a spectroscopy sample is placed between charged parallel-plate electrodes, the resulting electric field ($\vec{E}_{ext}$) causes a perturbation to the energy levels of the chromophores in the sample.[2] The magnitude of the applied field is limited by the point of dielectric breakdown for the sample material (i.e., where it begins to conduct electricity). For this experiment's frozen glycerol–water glass solvent, an appropriate range for $\vec{E}_{ext}$ is about $1 - 20$ kV/cm. The Stark energy-level shift increases as the magnitude of the applied field increases. This shift can be described with either a classical or quantum-mechanical approach.

First, the classical Stark equation is:

$$\Delta v\left(\vec{E}_{ext}\right) = -\frac{1}{h}\left[\Delta \vec{\mu} \cdot \vec{E}_{ext} + \frac{1}{2}\vec{E}_{ext} \cdot \Delta \overleftrightarrow{\alpha} \cdot \vec{E}_{ext} + \cdots \right] \tag{6.1}$$

where Δv is the Stark shift, i.e., change in transition frequency of the chromophore (in wavenumbers); h is Planck's constant; c is the speed of light in a vacuum. Let us define, $\Delta \vec{\mu}$, the difference dipole moment and $\Delta \overleftrightarrow{\alpha}$, the difference polarizability, more carefully in Eq. 6.2 and Eq. 6.3; the difference dipole moment is also illustrated in Fig. 6.1.[3] The subscripts "0" and "1" refer to the ground and first electronic excited states, respectively:

$$\Delta \vec{\mu} = \vec{\mu}_{1,perm} - \vec{\mu}_{0,perm} \tag{6.2}$$

$$\Delta \overleftrightarrow{\alpha} = \overleftrightarrow{\alpha}_1 - \overleftrightarrow{\alpha}_0 \tag{6.3}$$
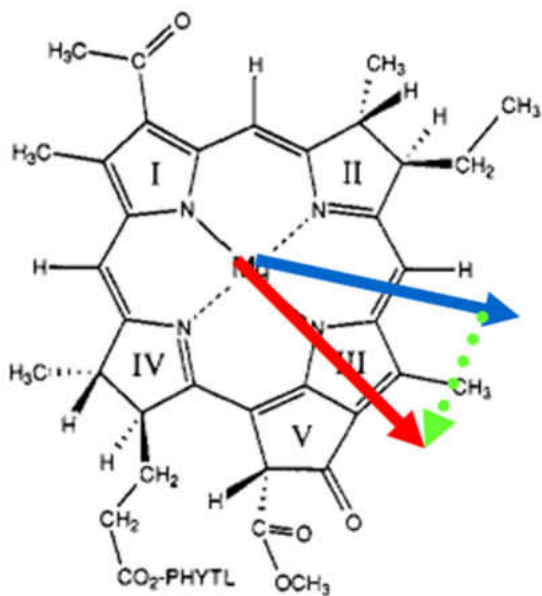
Figure 6.1. Difference dipole moment dramatized for chlorophyll, which would exhibit symmetry-breaking to have a non-zero $\Delta\vec{\mu}$ value. The red arrow represents the ground-state permanent dipole moment vector; the blue arrow represent the first-excited-state permanent dipole moment; the green dashed arrow represents the difference dipole moment, $\Delta\vec{\mu}$.

Clearly, the expression in Eq. 6.1 is a truncated series. Beyond the linear term, containing $\Delta\vec{\mu}$, and quadratic term, containing $\Delta\overleftrightarrow{\alpha}$, higher order terms, containing respective hyperpolarizability coefficients (β, γ, etc.), may also contribute to the classical description of the Stark shift depending on the extent of the non-linear optical nature of the sample. From Eq. 6.1, the field functions by interacting with the electron distribution of the chromophore. The quadratic term in the equation leads to a physical insight into the nature of the interaction: the external electric field interacting with the polarizability of the chromophore creates an *induced* dipole moment $(\vec{\mu}_{ind})$.[4] This is distinct from the *permanent* dipole moment found in Eq. 6.1 because it accounts for the additional charge polarization when the external electric field is applied:

$$\Delta\vec{\mu}_{ind} = \Delta\overleftrightarrow{\alpha} \cdot \vec{E}_{ext} \tag{6.4}$$

Considering now not just a single molecule but a population randomly distributed in a glassy solvent, the linear Stark shift in a molecule with a vanishing permanent-dipole-moment difference can be understood.[1, 5] As shown in Fig. 6.2, the colored arrows arranged in circular

144

distribution represent the permanent dipole moment vectors of a distribution of chromophores. The external electric field vector, as it is applied during a Stark experiment, falls incident upon this distribution and interacts to a different extent, depending on the angle. If we simplify the interaction to three scenarios, we can correspond a spectral change to each: 1) the difference dipole moment vector parallel to the applied field (red arrows) causes a shift in spectrum to lower energy (red curve); 2) the vector perpendicular to the applied field (green arrows) leaves the spectrum relatively unchanged in terms of energy (green curve); 3) the vector anti-parallel (blue arrows) to the applied field causes a shift to higher energy (blue curve). Intuitively, the shifts in spectra correspond to these interactions like one force "opposes" or "encourages" another. Thus, for a random distribution of chromophores, a Stark "shift" will manifest as a spectral broadening. The "zero-field spectrum" (i.e., before the external electric field is applied) and the spectrum after a field has been applied will encompass the same area under the curve because the same number of chromophores are sampled in each spectrum.
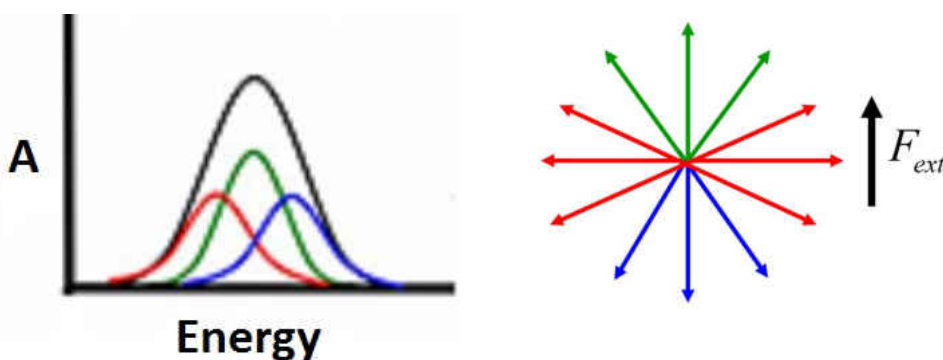
Figure 6.2. The spectrum before the field has been applied is the most intense and shown in black. The three general interactions shown by the vectors have corresponding, less-intense curves that shift in energy accordingly. (Figure adapted from Ref. [1].)

Another factor that arises for a population of chromophores in a glassy solvent is that $\vec{E}_{ext}$ encounters the chromophores' surroundings before interacting with the chromophore. In this work, the surroundings for the chromophores includes not only the frozen glycerol–water solvent but also the protein matrix in which the porphyrin chromophore is substituted. As might be expected, the external electric field is mitigated by the dielectric solvent with contributions from both the glass and the protein and scales with their respective dielectric constant values. However, the geometry of the cavity where the chromophore sits is also a factor, which can, in fact, serve to magnify the field intensity. For myoglobin, the heme cavity is an elongated space with an ill-defined border; however, because an exact treatment for a spherical cavity exists, will serve as the approximation for the heme cavity. These factors are defined by the Lorentz local field factor, f:

$$f = \frac{\varepsilon+2}{3} \qquad (6.5)$$

where ε is the effective dielectric constant of the glycerol–water and protein matrix combined. The local field factor, when considered, is then used as a coefficient for the applied field.[6]

Where does the internal electric field arise in this analysis? From Chapter 3, we defined *total* electric field ($\vec{E}_{tot}$) as the sum of internal electric field (generated by the protein matrix) and external electric field (applied during a Stark experiment) in Eq. 3.3. The classical expression given in Eq. 6.1 when an internal electric field is present contains not solely $\vec{E}_{ext}$ but $\vec{E}_{tot}$. Now that we see the field as it interacts with the solvent, we can define $\vec{E}_{tot}$ more precisely:

$$\vec{E}_{tot} = \vec{E}_{int} + \overleftrightarrow{f} \cdot \vec{E}_{ext} \qquad (6.6)$$

Often, the classical Stark equation will be truncated at only the linear term.[7] This may be due to the difficulty in obtaining experimental values for polarizabilities in excited states, which would be required to generate the $\Delta \overleftrightarrow{\alpha}$ coefficient for the quadratic term. Errors arise when neglecting higher-order terms, especially when working with a centrosymmetric molecule such as porphin. This is because molecules with inversion symmetry theoretically will have a dipole moment equal to zero in both the ground and excited states, therefore $\Delta \vec{\mu} = 0$. The linear term on its own would then not be sufficient to describe the Stark shift that results. For substituted myoglobin, the possibility of an *induced* dipole moment in porphyrins, however, is significant. With the definition in Eq. 6.4, $\vec{\mu}_{ind}$ can be captured to properly describe our system of interest. However, we again rely on the quadratic term for a proper description. Finally, the subpopulation of randomly oriented molecules described in Fig. 6.2 that lies perpendicular with $\Delta \vec{\mu}$ perpendicular to the external electric field does not exhibit a *linear* Stark shift. However, it may undergo a shift that is governed by higher-order terms are thus necessary to properly analyze this subset of sample.[8] It is clear that, especially for our system of interest, we cannot rely on a simple linear analysis.[9]

Might there be a better method for elucidation of internal electric field from Stark shifts in porphyrin-substituted myoglobin?[10] A distinct quantum-mechanical (QM) approach was developed by Kohler, et al. starting in 1992.[11] Here, the energy shift is expressed by a perturbation term to the Hamiltonian:

$$\widehat{H} = \widehat{H}_0 - \vec{E}_{tot} \cdot \hat{\vec{\mu}} = \widehat{H}_0 - (\vec{E}_{int} + f \cdot \vec{E}_{ext}) \cdot \hat{\vec{\mu}} \tag{6.7}$$

where $\widehat{H}$ is the total Hamiltonian (energy) and $\widehat{H}_0$ contains all of the contributions to energy *except* the total electric field, e.g., any conformational changes induced by the protein, short-range electrostatics, etc. The perturbation term, $\vec{E}_{tot} \cdot \hat{\vec{\mu}}$, is the scalar product of the total electric field (defined in Eq. 6.6) and the dipole moment *operator*. The dipole moment matrix element is defined as:

$$\vec{\mu} = \langle \psi_i | \hat{\vec{\mu}} | \psi_j \rangle = \langle \psi_i | q \cdot \hat{\vec{r}} | \psi_j \rangle \tag{6.8}$$

with subscripts *i* and *j* referring to electronic states. For *i=j*, the expression describes the permanent dipole moment; for *i≠j*, the expression describes the *transition* dipole moment, which relates to the probability of transition between states *i* and *j*. This QM approach describes the interaction of the field and chromophore in a unique way. The perturbation term states that the electric field is sensed as a "projection" onto the dipole moment of the chromophore. This has the consequence of defining the resolution of this analysis method: the expression yields the exact internal fields in the limit of the *point dipole approximation*. This means that the measured internal electric field in this work is essentially averaged over the surface of the porphyrin probe (see discussion in Chapter 4).

The QM Stark analysis introduced here will be expanded in detail along with challenges and possible solutions in the next chapter.

6.2.2   Spectral hole burning

The Stark shift in porphyrin-substituted myoglobin is on the order of 10^{-2}–10^{-3} nm for a reasonable external electric field strength. However, the spectral width of the lowest-energy electronic Q band, which is monitored in our experiments, is on the order of 20 nm. Clearly, the

shift would be entirely obscured if no modification to the spectrum were made. Resolution is re-gained through spectral hole burning.[12]

Hole burning selects a subset of chromophores in the sample that resonate at the same energy. The technique essentially photobleaches that subgroup by focusing a narrow-linewidth laser at a particular frequency on the sample, affecting the spatial volume that falls within the beam's cross-section and penetration depth. When the "burned" spectral region is scanned for a second time as readout, the resulting change to the spectrum exhibits a "dip" at the burning frequency. The hole depth and width is dependent on burning time and burning laser intensity, and represents the number of burned chromophores at that frequency. Because the laser linewidth is should ideally be narrower than the homogeneous linewidth of the transition under study (see definition in Chapter 5), the subset that is affected by the burning laser can be accessed at any point in the distribution that the laser crosses in frequency space. This can be visualized with Fig. 5.6b. If any single frequency is chosen for burning, it may lie either at the peak of a homogeneous line or may lie at the overlapping wings of any two, which would give the burning laser access to a subset over both homogeneous lines. The resulting spectral hole reflects both the burning action as well as the light absorbed during the scanning process that is performed in order to read out the hole. After these steps, the spectral hole is, in the ideal case, Lorentzian in shape (although inverted from the homogeneous lineshapes illustrated in Fig. 5.6) with a width that is twice the homogeneous linewidth, reflecting both the width of the burning step and the probing step:

$$\delta_{hole} = 2\,\Gamma_{hom} \tag{6.9}$$

What is the physical process of spectral hole burning? Hole-burning mechanisms are molecular-species-dependent and can be divided into two categories: 1) photochemical burning and 2) photophysical burning.[13] For free-base protoporphyrin IX, the hole-burning mechanism is photochemical, meaning the probe molecule itself undergoes a rearrangement that breaks and make covalent bonds. In this species, it is postulated that an isomerization of the inner hydrogens, i.e., changing positions from one *trans* position to the other, results in a spectral hole because the pyrrole-nitrogen positions at the center of the porphyrin ring are not equivalent due to non-symmetric positions of the vinyl substituents (see Fig. 4.2)[14] and because of the disordered local environment. For metalloporphyrins like zinc protoporphyrin IX, the inner-hydrogen-isomerization option is not available. Thus it is thought that this species undergoes a photophysical process instead. This a less specific mechanism and involves the rearrangement of the chromophore's surroundings or possibly less drastic conformational changes rather than a chemical process.[15] For either mechanism, photochemical or photophysical, the effect is to move the resonance frequency for the burned subset so that the number of molecules absorbing at the burning frequency is depopulated. At very low temperatures (e.g., <4 K), the photoconversion can persist as long as the temperature remains low enough to prevent the reverse reaction and the spectral hole is permanent under such conditions.

Extensive lineshape theories for spectral holes have been developed.[16] These complex expressions nevertheless are interesting in the number of variables that are encompassed, which help to gain a handle on the number of physical factors involved in the process of spectral hole burning. As mentioned above, spectral holes contain contributions from both the burning and the readout ("scanning" or "probing" step), f_B and f_P respectively. The polarization of the laser in

these two steps with respect to the transition dipole moment of the chromophore also plays a

role in the resulting lineshape of the hole. Here, the angles θ (the in-plane polar angle) and φ (the

azimuthal angle, i.e., the projection onto the perpendicular plane) play a role. The overall hole

shape for a photochemical process as a function of frequency ($\Gamma_{hole}(v)$) is given as:

$$\Gamma_{hole}(v) = A_0 - \frac{\sigma}{4\pi} \int_0^{2\pi} d\varphi \cdot \sin\theta \int_{-\infty}^{+\infty} N^\tau \left[f_B(\theta), v' \right] \cdot f_P(\theta, \varphi) \cdot HLF(v_0 - v') dv'$$

(6.10)

where A_0 is the absorption intensity at the burning frequency (v_0) *before* burning; σ is the

absorbance cross-section of the chromophore; N is the number of molecules in the burn volume;

v' is the center frequency of the homogenous line; the normalized homogeneous lineshape

function is designated by $HLF(v_0 - v')$. The burning and scanning steps can be performed

parallel (∥) or perpendicular (⊥) to one another. These vectors interacting with the

chromophore's transition dipole moment are illustrated in Fig. 6.3. The polarization factors of

those steps (with respect to the transition dipole moment) are:

$$f_B(\theta) = cos^2\theta$$ (6.11)

$$f_{P\parallel} = cos^2\theta$$ (6.12)

$$f_{P\perp} = sin^2\theta sin^2\varphi$$ (6.13)

The burning function is further defined:

$$N^\tau[f_B(\theta), v'] = N_0(v') \cdot exp\left[-\frac{\sigma \cdot I}{hv_0} \eta\tau \cdot f_B(\theta) \cdot HLF(v_0 - v') \right]$$ (6.14)

where $N_0(v')$ is the inhomogeneous distribution function before burning; I is the intensity of the burning laser; η is the quantum yield of the burning process (i.e., the proportion of the absorbed photons that results in a chromophore photoconversion to a "burned" product); τ is the burn time.
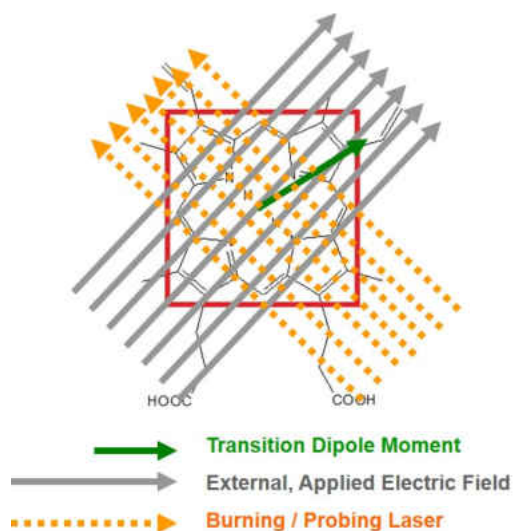
Figure 6.3. Three interacting vectors in a hole-burning experiment: transition dipole moment of a porphyrin chromophore (green with the symmetric porphin core boxed in red); external electric field (gray); the burning and probing laser (applied parallel in this case in orange).

When an external electric field is applied to a spectral hole, a Stark shift occurs. However, there are two distinct effects that can be observed and that depend on the angle of the applied field with respect to the laser polarization: 1) a broadening or 2) a splitting. The two effects can be seen for a single hole in a single sample when the applied field is switched from a parallel to perpendicular interaction with the scanning and burning laser polarization. See Fig. 6.4a for examples of broadened spectral holes and Fig. 6.4b for examples of split holes.
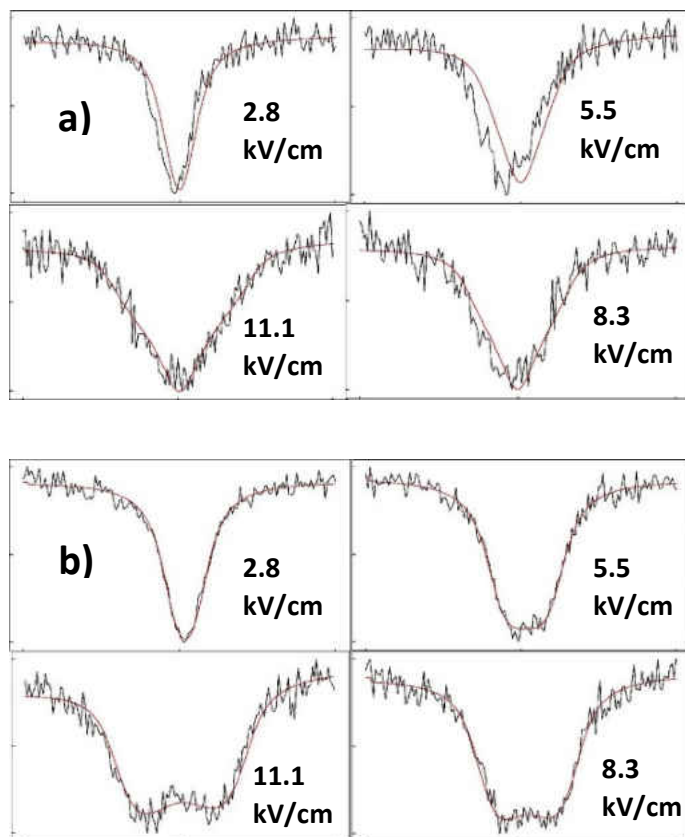
Figure 6.4. Spectral holes burned in the Q_x band of PPIX–myoglobin at two different orientations of laser polarization with respect to external electric field. The Stark shift increases as external electric field increases clockwise in each panel with values noted in kV/cm. **a)** broadened holes with a perpendicular orientation; **b)** split holes with a parallel orientation.

6.3 Experimental

With the aim of measuring internal electric field in heme proteins, an apparatus for hole-burning Stark spectroscopy was assembled by group member Brad Moran. Limited results with the current experimental set-up were obtained and are given here.

6.3.1 Current hole-burning Stark spectroscopy apparatus

The basic components of an experimental set-up for hole-burning Stark spectroscopy are a burning laser and a scanning laser (which in this case are the same) a wavemeter to establish position in frequency space, a polarizer for laser light, a cryostat, a sample holder with electrodes to generate an external electric field, detectors for absorbance or emission, and a way to process

the resulting spectra. In this work, a 5–6 watt, continuous wave, 532-nm Millenia Pro frequency-doubled Nd:YAG laser (Spectra Physics / Newport Co., Santa Clara, CA) was used to pump a Matisse ring dye jet laser (Sirah Lasertechnik, Grevenbroich, Germany). The dye laser had a continuously tunable frequency range, both by the type of dye used and by scanning optics that adjust the size of the laser cavity. The dye with appropriate wavelength range for working with free-base protoporphyrin IX is dicyanomethylene (DCM, 620–670 nm); for blue-shifted metalloporphyrins, rhodamine 6G (R6G) provides a range 570–620 nm.. From the dye laser, part of the beam entered a wavemeter (High Finesse, Tübingen, Germany) to monitor frequency position. After a series of focusing optics, including polarizing elements, the laser beam entered a cryostat via optical fiber and directed to the bottom of the cryostat where the porphyrin-substituted protein sample in glycerol–water solvent was placed. The sample holder was configured to contain the sample between parallel-plate electrodes, which were charged with a direct-current power supply (Bertram [now Pearse–Bertram, Bloomfield, CT]) with a range of up to 25 kV. The fluorescence resulting from laser excitation was captured above the sample where the excitation light entered the cryostat. The fluorescence, which was emitted well red-shifted to the excitation light, was detected by an avalanche photodiode (APD, Perkin-Elmer, Waltham, MA). Thus, the spectra collected is a fluorescence excitation spectrum where the *x*-axis is excitation energy and the *y*-axis is fluorescence intensity. A basic scheme for hole-burning Stark spectroscopy is shown in Fig. 6.5:
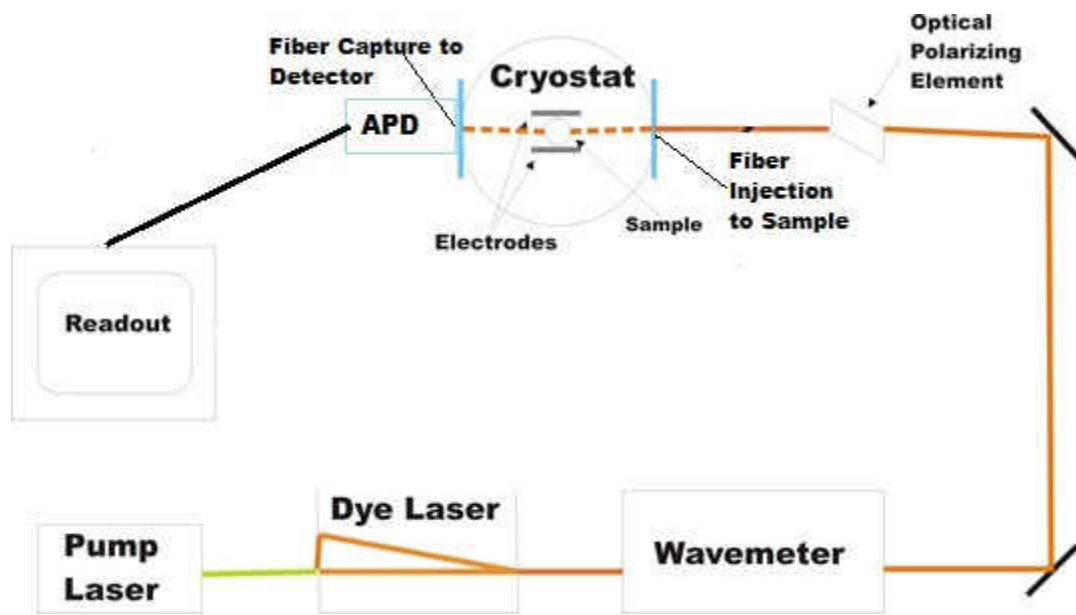
Figure 6.5. Hole-burning Stark spectroscopy experimental set-up. The green laser beam from the pump laser is transformed to a lower-energy orange beam by the ring dye laser. The path of the beam is followed through the components described above.

The careful alignment of the number of optical components in this experiment set-up was essential for the excitation light to reach the sample. A laser power stabilizer ("noise eater"; Cambridge Research and Instrumentation, Woburn, MA) was employed to correct intensity fluctuations from the laser source. One corrective element placed in the beam path was a beam stabilizer (Newport Corp. / New Focus, Santa Clara, CA). This was deemed necessary because the scanning process used by the dye laser results in a slight "beam walk" (see Fig. 6.6), which can displace the beam a few hundred nanometers and cause it to miss a target mirror, lens, or optical-fiber coupling port. The beam stabilizer was actually comprised of several optical elements: a beam sampler to redirect <10% of the laser beam to a position sensitive detector that controls

the position of highly reflective mirrors on motorized mounts to actively correct the beam path in real time.

x = Beam displacement
d = Brewster plate thickness
p = beam pathlength in plate
α = incident angle
β = refracted angle

Equations:

$$\beta = \arcsin(\sin\alpha/n) \qquad p = d/\cos\beta \qquad x = p\,\sin(\alpha-\beta)$$
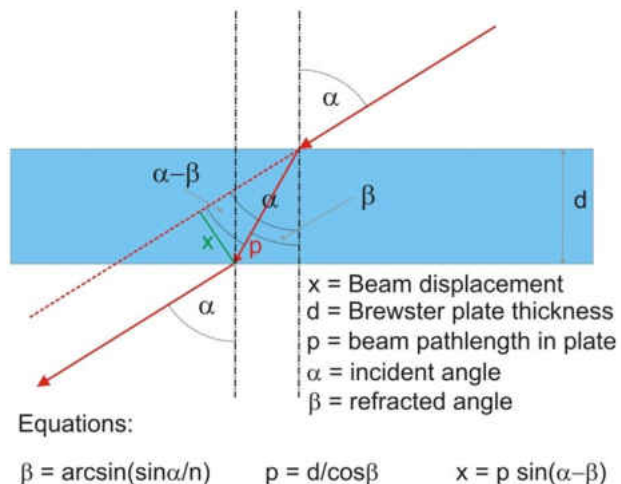
Figure 6.6. Laser beam path through a Brewster plate, which is the optical element in the dye laser responsible for the scanning capability, allows for calculation of the beam displacement it causes.

Once the laser beam traversed the optics on the table, it entered the cryostat, where the sample resided. Based on the discussion of inhomogeneous broadening as well as the need to create persistent spectral holes, the porphyrin-substituted protein was analyzed in a custom-built liquid-helium cryostat (Janis Research Company, Inc., Woburn, MA). The cryostat was comprised of an outer vacuum jacket, a cryogen layer, another vacuum jacket, and an inner sample compartment that could be connected to the cold cryogen gas or could be isolated in vacuum. The cryogen layer was first cooled to liquid nitrogen temperature (77 K), which was subsequently siphoned out, and replaced with liquid helium (4 K). When the liquid helium was placed under vacuum, the temperature was dropped to 1.4 K. The number of steps involved with cooling made it a challenge to achieve the very low temperature required for a successful experiment.

Another challenge with the experimental set-up in this work was an appropriate sample holder for the cryostat described, which resided at the bottom of the > 4-foot-tall vessel. The main criterion for an effective design was that the electrodes needed to sit about a millimeter apart to generate field strengths of 5–20 kV/cm, leaving the sample volume to reside in a narrow channel of insulating material. Explored designs began with a glass microscope slide with electrode-lined channels. First, an attempt was made to deposit layers of nickel–chromium and aluminum by thermoevaporation to a thickness tall enough to form channels. However, the metal flaked away from the glass easily. Then, mounting 1-mm-thick foam tape onto the glass to form a channel and covering it with conductive silver paint was tried, but the adhesive was not appropriate for a low-temperature environment. A similar attempt with gluing cut glass onto the glass mount and covering the surface of the channel with copper tape was made. This, too, had gaps under the adhered channel edges where solvent could escape the channel. Finally, a more complex polypropylene holder with a milled sample channel was obtained. The channel was lined with thin pieces of copper secured without adhesive. Initially, experiments were performed with solvent directly touching the copper electrodes. However, once electric field was applied, it became apparent that the sample needed insulation from the field because heating occurred of the sample compartment when the field was applied. The sample was then placed in a square, glass capillary tube that fit into the narrow channel snugly.

Because the sample resided a few feet away from the detector, the small sample volume need to produce enough fluorescence to register several hundred photon counts on the sensitive APD detector. The signal was impacted not only by the concentration of chromophores but also by the intensity of the incoming laser light (which could fluctuate dramatically over several hours)

as well as the alignment of the sample with respect to the focused beam's focal point. The alignment process was exacerbated by the inability to know the absolute location of the sample inside of the enclosed, opaque cryostat. In addition, piezo-controlled coarse and fine positioning motors lacked the ability to meter the spatial distance covered when an impulse was sent to initiate movement. However, once an appropriate concentration of chromophores was found (3.03×10^{-3} M PPIX at 50% substitution in myoglobin), the alignment issues were eventually overcome.

Finally, the experimental set-up described was capable of producing spectral holes with PPIX–myoglobin. However, the production of an electric field large enough to produce a measureable Stark shift was impeded in a sample chamber under vacuum, which is required to bring the temperature below 4 K. Introducing cold cryogen gas (helium) into the sample chamber allowed for a larger field, but brought the temperature higher than was tolerable to retain resolution.

6.3.2   Past hole-burning Stark spectroscopy apparatus

Because the spectral holes collected with the current apparatus were not able to have an electric field applied (see results and discussion below), the advances in QM Stark analysis described in the next chapter (Chapter 7) were tested using previously collected spectral holes, namely the spectra shown in Fig. 6.4. (The set was labelled Myo A 16131 in past files.) That data collection used an experimental set-up with some differences from the current one.[17] First, the pump and dye laser combination was different: a 7.0 W, 514- nm Coherent Sabre TSM20 argon-ion pump laser with a tunable Coherent 699-29 single frequency ring dye laser operating with Kiton Red 620 in a mixture of methanol and ethylene glycol. Then, the sample holder was a

cylindrical, 1-mm glass capillary, which was immersed in a home-made glass optical cryostat. The exciting laser light fell incident upon the sample directly through the quartz cryostat windows as opposed to the confocal set-up described in the current set-up. The excitation (for reference) and fluorescence were detected with cooled Burle C31034 photomultiplier tube (PMT) (rather than the avalanche photodiodes used in the current set-up) and then amplified 25 times (Stanford Research Systems SR445) before reaching a photon counter (Stanford Research Systems SR400). Finally, to apply an electric field to the sample, the high-voltage cable ran through the helium gas and require insulation in an evacuated stainless-steel tube.

6.3.3   Results and discussion

The following series of spectral holes were created in PPIX–myoglobin in glycerol–water at < 4 K. They were burned in a number of locations in the Q_x band at approximately 620 nm (exact location given in GHz). Figures 6.7a –e show the effect of burn time on spectral holes. It was determined that a 12-second burn time with this particular sample and set of optics was the most effective. In addition, it can be noted that the post-burn spectrum (i.e., the scanning step) loses intensity, which is certainly a consequence of photobleaching and had been mitigated somewhat by scanning at lower intensity by way of filters. A series of repeated scanning steps was performed and shown in Figure 6.8.

Spectral Hole Burned at 483602.13 GHz
Burn Time: 20 seconds
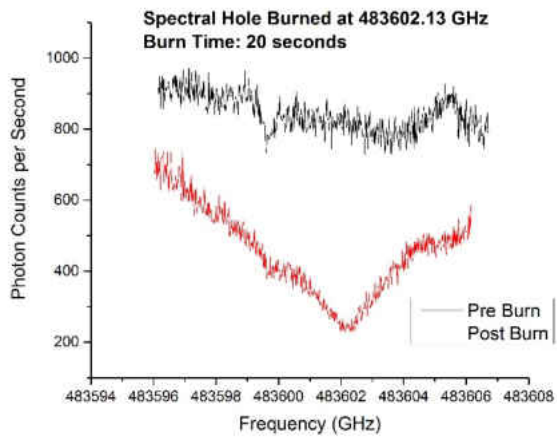
Figure 6.7a. Long burn time (20 s) manifests in a triangular shaped hole rather than a Lorentzian line shape.

Spectral Hole Burned at 483472.43 GHz
Burn Time: 15 seconds
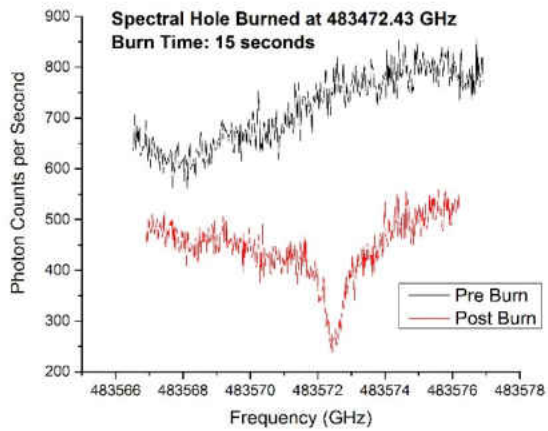
Figure 6.7b. 15-second burn time shows a Lorentzian line shape.

Spectral Hole Burned at 483459.03 GHz
Burn Time: 12 seconds
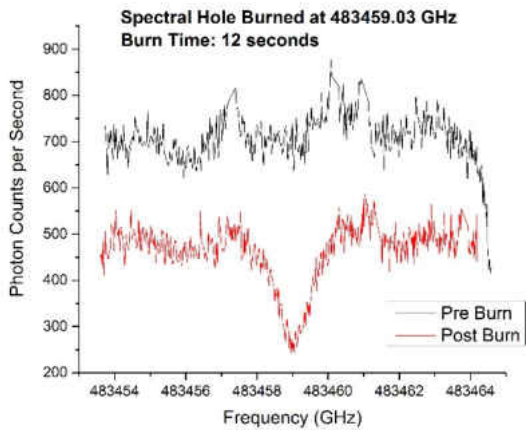
Figure 6.7c. 12-second burn time shows a Lorentzian line shape.
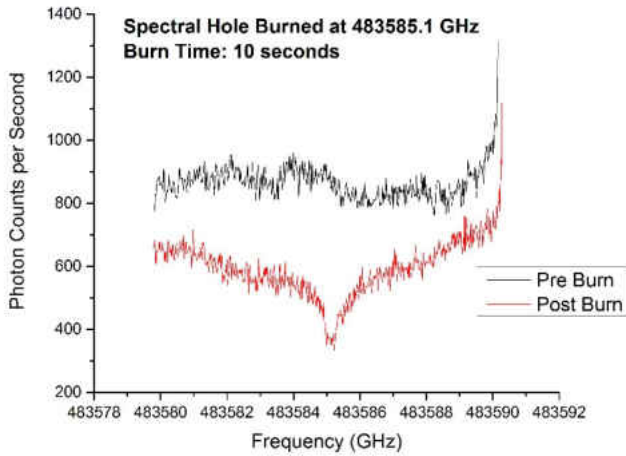
Figure 6.7d. 10-second burn time shows a Lorentzian line shape.
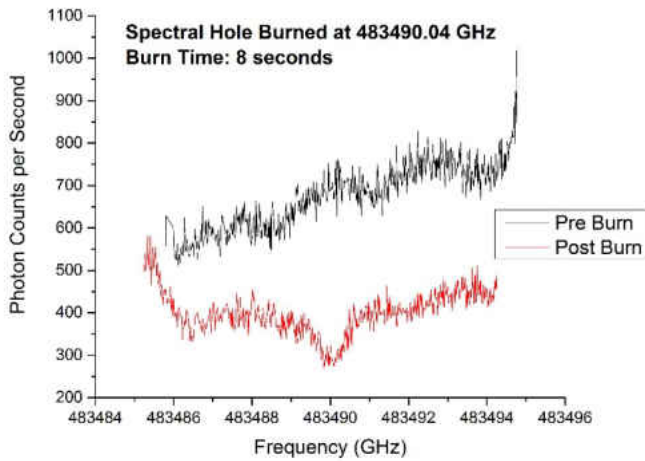
Figure 6.7e. Under-burning (8 sec) also deviates from a Lorentzian line shape, leaving the hole minimum unclear.
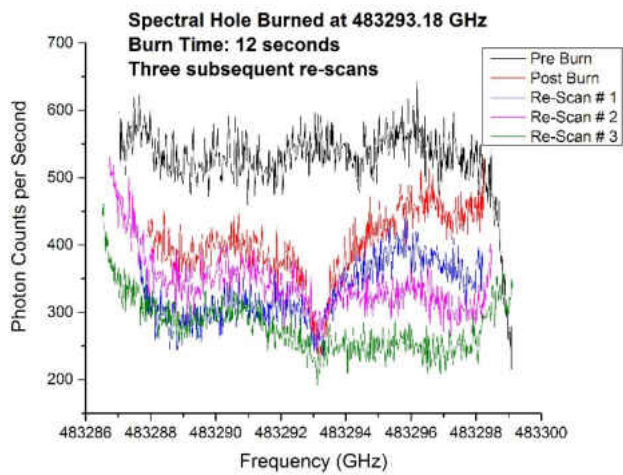
Figure 6.8. A total of four scans after the creation of a spectral hole clearly shows gradual photobleaching of the scanned sample volume.

6.4 Conclusions

Stark spectroscopy can be used to measure the internal electric field inside of heme proteins when they are substituted with porphyrins that exhibit fluorescence. For porphyrins that suffer from significant inhomogeneous broadening, spectral hole burning is used. An extensive experimental set-up to accomplish hole-burning Stark spectroscopy has been assembled and continuous improvements towards the goal of collecting high-resolution data are being made. Spectral holes at liquid helium temperature have been created in PPIX–myoglobin. However, the ability to apply an electric field for Stark experiments in the vacuum environment required for temperatures <4 K is a barrier and the trade-off between these two conditions requires future negotiation.

6.5 References

1.   Boxer, S. G., Stark Realities. *Journal of Physical Chemistry* **2009,** *113* (2972-2983).

2.   Samoilenko, V. D.; Rasumova, N. V.; Personov, R. I., Stark effect in narrow gaps in the absorption bands of complex molecules. *Optika i spektroskopiia* **1982,** *52* (4), 346-348.

3.   (a) Boxer, S. G.; Goldstein, R. A.; Lockhart, D. J.; Middendorf, T. R.; Takiff, L., Excited States, Electron-Transfer Reactions, and Intermediates in Bacterial Photosynthetic Reaction Centers. *Journal of Physical Chemistry* **1989,** *93*, 8280-8294; (b) Hammes, S. L.; Mazzola, L.; Boxer, S. G.; Gaul, D. F.; Schenck, C. C., Stark spectroscopy of the Rhodobacter sphaeroides reaction center heterodimer mutant. *Proceedings of the National Academy of Sciences* **1990,** *87* (15), 5682-5686; (c) Lockhart, D. J.; Hammes, S.; Franzen, S.; Boxer, S. G., Electric Field Effects on Emission Line Shapes When Electron Transfer Competes wlth Emission: An Example from Photosynthetic Reaction Centers. *Journal of Physical Chemistry* **1991,** *95* (2217-2226).

4.   Altmann, R. B.; Renge, I.; Kador, L.; Haarer, D., Dipole moment differences of nonpolar dyes in polymeric matrices: Stark effect and photochemical hole burning. I. *The Journal of Chemical Physics* **1992,** *97* (8), 5316-5322.

5.   Bublitz, G. U.; Boxer, S. G., Stark Spectroscopy: Applications in Chemistry, Biology, and Physics. In *Annual Review of Physical Chemistry*, 1997; Vol. 48, pp 213-242.

6.   (a) Geissinger, P., Kohler, Bryan E., Woehl, Jorg C., Electric Field and Structure in the Myoglobin Heme Pocket. *J. Phys. Chem.* **1995,** *99*, 16527-16529; (b) Andrews, S. S.; Boxer, S. G., Vibrational Stark Effects of Nitriles I. Methods and Experimental Results. *The Journal of Physical Chemistry A* **2000,** *104* (51), 11853-11863; (c) Hashimoto, H.; Fujii, R.; Yanagi, K.; Kusumoto, T.; Gardiner, A. T.; Cogdell, R. J.; Roszak, A. W.; Issacs, N. W.; Pendon, Z.; Niedzwiedski, D.; Frank, H. A., Structures and functions of carotenoids bound to reaction centers from purple photosynthetic bacteria. *Pure Applied Chemistry* **2006,** *78* (8), 1505-1518; (d) Kanchanawong, P.; Dahlbom, M. G.; Treynor, T. P.; Reimers, J. R.; Hush, N. S.; Boxer, S. G., Charge Delocalization in the Special-Pair Radical Cation of Mutant Reaction Centers of Rhodobacter sphaeroides from Stark Spectra and Nonadiabatic Spectral Simulations. *The Journal of Physical Chemistry B* **2006,** *110* (37), 18688-18702.

7.   (a) Sandberg, D. J.; Rudnitskaya, A. N.; Gascon, J. A., QM/MM Prediction of the Stark Shift in the Active Site of a Protein. *Journal of Chemical Theory and Computation* **2012,** *8* (8), 2817-2823; (b) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G., Electric Fields at the Active Site of an Enzyme: Direct Comparison of Experiment with Theory. *Science* **2006,** *313* (5784), 200-204.

8.   Drobizhev, M.; Tillo, S.; Makarov, N. S.; Hughes, T. E.; Rebane, A., Color Hues in Red Fluorescent Proteins Are Due to Internal Quadratic Stark Effect. *The Journal of Physical Chemistry B* **2009,** *113* (39), 12860-12864.

9.   Geissinger, P., Quantitative Measurement of Internal Molecular Electric Fields in Solids. In *Anisotropic Organic Materials: Approaches to Polar Order*, Glaser, R., Kaszynski, Piotr, Ed. American Chemical Society: Washington, DC, 2002.

10.   (a) Kohler, B.; Woehl, J. C., Classical and quantum mechanical models for Stark experiments. *Molecular Crystals Liquid Crystals* **1996,** *291*, 119-134; (b) Geissinger, P., Woehl, Jorg C., Prince, Barry J., A quantum-mechanical model for the determination of internal electric fields at protein active sites from the Stark effect on persistent spectral holes. *Journal of Luminescence* **2004,** *107*, 220-229.

11.   Gradl, G.; Kohler, B. E.; Westerfield, C., Electric field splitting of the octatetraene 1 1Ag-->2 1Ag transition in n-hexane. *Journal of Chemical Physics* **1992,** *97* (9), 6064.

12.   (a) Purchase, R.; Sellars, M. J.; Krausz, E.; Manson, N. B., Electric-field-induced broadening of spectral holes in zinc phthalocyanine. *Chemical Physics Letters* **2000,** *327* (3-4), 189-196; (b) Jankowiak, R.; Reppert, M.; Zazubovich, V.; Pieper, J.; Reinot, T., Site Selective and Single Complex Laser-Based Spectroscopies: A Window on Excited State Electronic Structure, Excitation Energy Transfer, and Electron-Phonon Coupling of Selected Photosynthetic Complexes. *Chem. Rev. (Washington, DC, U. S.)* **2011,** *111* (8), 4546-4598; (c) Volker, S., Hole-Burning Spectroscopy. *Annu. Rev. Phys. Chem.* **1989,** *40*, 499-530; (d) Jankowiak, R.; Hayes, J. M.; Small, G. J., Spectral Hole-Burning Spectroscopy in Amorphous Molecular Solids and Proteins. *Chem. Rev. (Washington, DC, U. S.)* **1993,** *93* (4), 1471-1502; (e) Kohler, B. E.; Personov, R. I.; Woehl, J. C., Electric Field Effects in Molecular Systems Studied via Persistent Hole Burning. In *Laser Techniques in Chemistry*, Myers, A. B.; Rizzo, T. R., Eds. John Wiley & Sons, Inc.: 1995; pp 283-323; (f) Friedrich, J., Hole Burning Spectroscopy and Physics of Proteins. In *Methods in Enzymology*, Sauer, K., Ed. Academic Press: 1995; Vol. Volume 246, pp 226-259.

13.   Friedrich, J.; Haarer, D., Photochemical Hole Burning: A Spectroscopic Study of Relaxation Processes in Polymers and Glasses. *Angew. Chem., Int. Ed. Engl.* **1984,** *23* (2), 113-140.

14.   Volker, S.; van der Waals, J. H., Laser-Induced Photochemical Isomerization of Free Base Porphyrin in an *n*-Octane Crystal at 4.2 K. *Mol. Phys.* **1976,** *32* (6), 1703-1718.

15.   Carter, T. P.; Small, G. J., Nonphotochemical Hole Burning of Self-Aggregated Dimers of Chlorophyll a in Polystyrene. *J. Phys. Chem.* **1986,** *90* (10), 1997-1998.

16.   (a) Haarer, D., Photochemical Hole-Burning in Electronic Transitions. In *Persistent Spectral Hole-Burning: Science and Applications*, Moerner, W. E., Ed. Springer-Verlag: New York, 1987; Vol. 44; (b) Rebane, K. K.; Rebane, L. A., Basic Principles and Methods of Persistent

Spectral Hole-Burning. In *Persistent Spectral Hole-Burning: Science and Applications*, Moerner, W. E., Ed. Springer-Verlag: New York, 1987; Vol. 44.

17.   Woehl, J. C. Measuring Internal Electrostatic Fields and Potentials at Molecular and Atomic Resolution using Hole-Burning Spectroscopy. University of California, Riverside, Riverside, CA, 1996.

Chapter 7

Quantum-Mechanical Stark Analysis

7.1 Introduction

The quantum-mechanical (QM) Stark analysis method briefly introduced in Chapter 6 is expanded in detail here. First, the theory and method of analysis is continued, followed by an explanation of a number of challenges in implementation. Then, an in-house line-fitting program to implement the QM Stark analysis, *Stark06*, is described. Attempts to overcome the previously incomplete implementation are made and eventually achieved. Theoretical underpinnings of the new implementation process are offered. Finally, a complete protocol for the QM Stark analysis is presented.

7.2 QM Stark Analysis Theory

An expansion of Eq. 6.8 reveals that the QM approach to Stark analysis is also a series like the classical approach in Eq. 6.1:

$$\widehat{H} = \widehat{H}^{(0)} - \vec{E} \cdot \hat{\vec{\mu}} - \frac{1}{6}\sum_i \sum_j \widehat{Q}_{ij}\frac{\partial E_j}{\partial x_i} - \ldots \tag{7.1}$$

where $\widehat{Q}_{ij}$ is the electric quadrupole moment operator. Further terms in the expression are based on the subsequent multipole expansion (octapole, hexadecapole, etc.). Here, we restrict the expression to the dipole moment term, which is not an arbitrary neglect of terms but rather an approximation that the gradient of the electric field ($\frac{\partial E_j}{\partial x_i}$) is zero. This means that we are measuring the field as homogeneous over the plane of the porphin probe and will yield an exact value within this point-dipole approximation.

Expressing Eq. 7.1 specifically for porphin in matrix form (Eq. 7.2), the diagonal contains

electronic state energies (because the permanent dipole moment - $\vec{\mu}_{ij}$ where $i{=}j$ – is equal to

zero in the centrosymmetric molecule) and the off-diagonal elements contains the product of the

electric field and transition dipole moment for the transition between the two electronic states,

i and *j*. Solving this equation means that a diagonalization procedure calculates new electronic

state energies taking the perturbation terms in the off-diagonal elements into account. In other

words, the goal is to find the new electronic state energies once the electric field is applied.

Solving the equation also requires calculated input for all transitions in the matrix: 1) transition

dipole moment and 2) state energy of each electronic state considered in the matrix. This

amounts to calculating *N* electronic states for the diagonal of an *N* x *N* matrix and *[(N² / 2) – N]*

unique transition dipole moment values for the off-diagonal elements of the matrix (considering

that $\mu_{ij} = \mu_{ji}$ in this case). The process of calculating these excited-state values with *ab initio*

and density functional theory methods in the *Gaussian09* (G09) computational package will be

detailed in Chapter 8.

$$\hat{H} = \begin{pmatrix} 0 & -E_x\mu_{Q_x} & -E_y\mu_{Q_y} & -E_x\mu_{B_x} & -E_y\mu_{B_y} \\ -E_x\mu_{Q_x} & \varepsilon_{Q_x}^{(0)} & -\vec{E}\cdot\vec{\mu}_{12} & -\vec{E}\cdot\vec{\mu}_{13} & -\vec{E}\cdot\vec{\mu}_{14} \\ -E_y\mu_{Q_y} & -\vec{E}\cdot\vec{\mu}_{21} & \varepsilon_{Q_y}^{(0)} & -\vec{E}\cdot\vec{\mu}_{23} & -\vec{E}\cdot\vec{\mu}_{24} \\ -E_x\mu_{B_x} & -\vec{E}\cdot\vec{\mu}_{31} & -\vec{E}\cdot\vec{\mu}_{32} & \varepsilon_{B_x}^{(0)} & -\vec{E}\cdot\vec{\mu}_{34} \\ -E_y\mu_{B_y} & -\vec{E}\cdot\vec{\mu}_{41} & -\vec{E}\cdot\vec{\mu}_{42} & -\vec{E}\cdot\vec{\mu}_{43} & \varepsilon_{B_y}^{(0)} \end{pmatrix}$$

From State i → *To State j*

(7.2)

This 5 x 5 matrix accounts for only the electronic transitions in the visible region of porphin (Q_x, Q_y, B_x, B_y,) but reveals that the quantum-mechanical analysis of Stark shifts relies not only on the experimental transition of interest (here the Q_x transition between the state 0 (the ground state) and state 1 (the first excited state)) but also transitions surrounding it. In fact, theoretically, *all* other electronic states beyond the states of interest (comprising the experimental transition of interest) can influence the analysis. The application of an electric field to a sample in a Stark experiment has the effect of extensive mixing between electronic states, unlike the Zeeman effect where the application of a magnetic field is relatively straightforward in its effect. The effect of electronic-state mixing in this case amounts to a kind of repulsion between electronic states when an electric field is applied, increasing or decreasing the energy of individual states (see Fig. 7.1). The extent of repulsion increases with the strength of the applied electric field but no linearly as the varying influence of other electronic states must be accounted for.
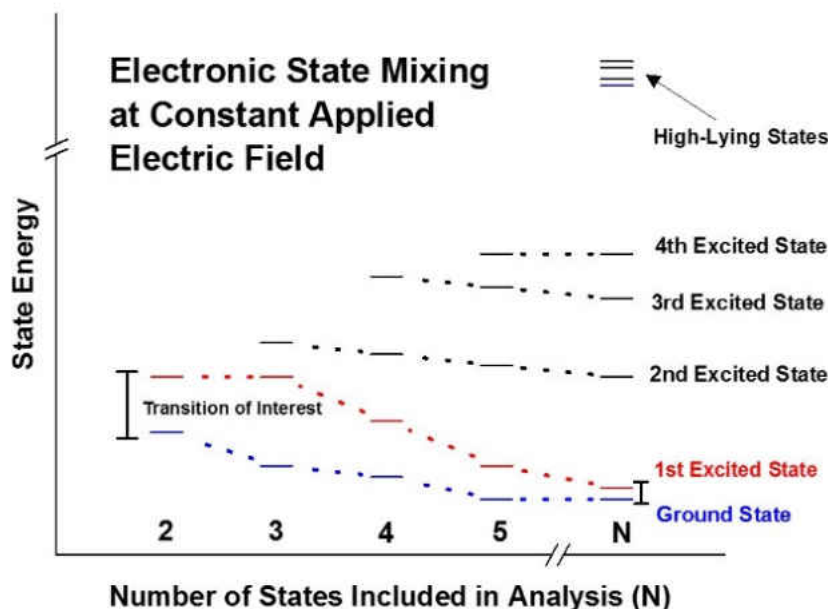
Figure 7.1. The effect of an applied electric field on the energetic positions on electronic state energies of a chromophore in a Stark experiment. State energies required for QM analysis depend on the number of other states considered.

The influence of one state on another is not constant but varies in coupling strength according to Eq. 7.3:

$$\frac{\langle \psi_i | \mu_r | \psi_j \rangle \langle \psi_j | \mu_s | \psi_i \rangle}{\varepsilon_i^{(0)} - \varepsilon_j^{(0)}} \tag{7.3}$$

where $\langle \psi_i | \mu_r | \psi_j \rangle$ is the transition dipole moment between two states, i and j, and $\varepsilon^{(0)}$ is the electronic state energy of the coupled states. The subscripts r and s refer to the components of the transition dipole moment vector, i.e. they reflect the spatial polarization of the transition dipole moment For centrosymmetric porphin oriented according to Gouterman (see Chapter 5 and Fig. 7.2), all transition dipole moments are polarized along a single axis, x, y, or z, so the

numerator in Eq. 7.3 is simply the square of the transition dipole moment for the transition considered.
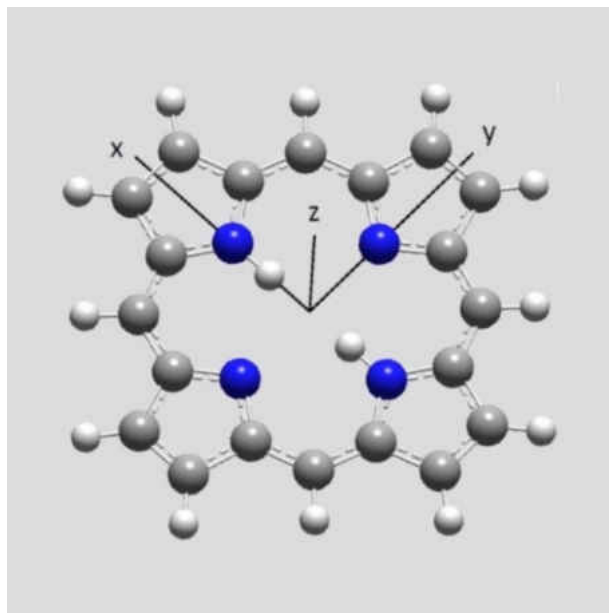
Figure 7.2. Free-base porphin with axes aligned in a standard orientation according to Gouterman. This orientation is adhered to throughout the following computational results.

Unfortunately, molecules theoretically have an *infinite* number of electronic states, which is a seemingly insurmountable barrier to implementing this approach to the analysis of Stark spectra. However, examining Eq. 7.3, we can see how the most influential electronic states on our transition of interest might be elucidated. First, the denominator simply states that the farther an electronic state is from the two states, 0 and 1, in our transition of interest (Q_x), the smaller the influence it has, which is intuitive. This suggests that there may be a point where a state is just too far away to have significance in the analysis. Then, the numerator contains a value (transition dipole moment) that gives the probability of that particular transition occurring. For centrosymmetric molecules like porphin, the physical basis of spectroscopic selection rules is symmetry: two electronic states of identical symmetry will have a transition dipole moment of zero while two electronic states of differing symmetry will have a non-zero transition dipole

moment. (This is also why the permanent dipole moment of any electronic state of porphin is zero.) This might suggest that an extremely probable transition might overcome a large energy difference, especially because the numerator is squared while the denominator is not. However, there appears to be a theoretical limit to the transition dipole moment value that is related to transition energy that will become apparent in the discussion of oscillator strength below.

7.3 Generating a finite matrix

As mentioned above, the theoretically infinite number of electronic states required to solve the Hamiltonian in the QM Stark analysis was the major barrier to effective implementation of the method to determine internal electric field. A couple of studies were undertaken by Geissinger, et al. previously to try to determine which electronic states were necessary for the analysis of porphyrin-substituted protein.[2] Limited largely by insufficient computational resources, the completeness of the set of excited-state values was never certain. However, these studies made it clear that the final experimental value for internal electric field analyzed in this manner relies heavily on the number of states included in the analysis, i.e., the matrix dimensions. (Comparisons between methods to calculate excited-state properties, i.e., state energies and transition dipole moment, will be made in Chapter 8, where these earlier results will be discussed.)

The following two sections describe two attempts at gauging completeness of a matrix to thoroughly analyze existing Stark spectra but with the computational resources to examine a matrix of much larger dimensions. Here, access to 252 electronic states (see Fig. 7.3a and 7.3b) for analysis are used, as opposed to ≤17 states previously available. Figures 7.3a and 7.3b are the

same data presented in two formats, the number of states populating higher energies much more densely than lower energies near the transition of interest (Q_x, ground- to first-excited state).
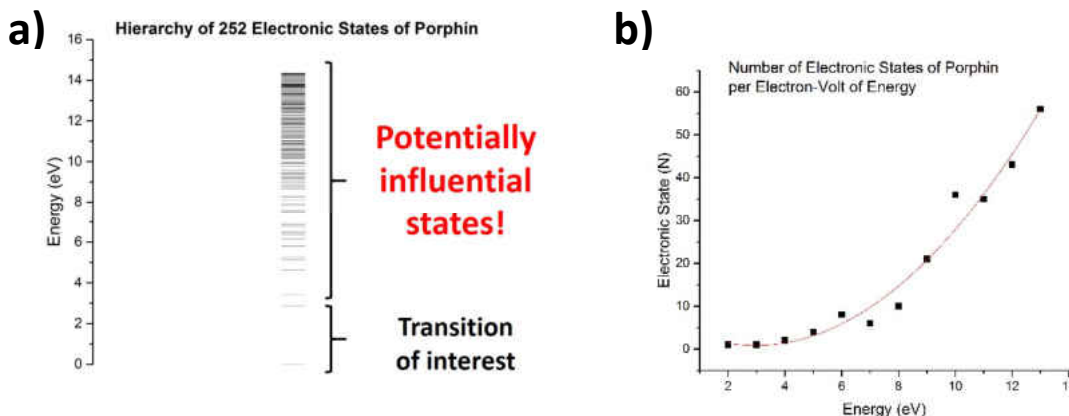
Figure 7.3. 252 electronic states of porphin calculated with CIS / sdd (method / basis set). **a)** A "hierarchy" viewpoint that emphasizes the low-energy transition of interest and the large number of states at much higher energy of potential influence. **b)** The exponential growth of the density of electronic states from low to high energy.

7.3.1 Completeness via sum-over-states polarizability

A first attempt to understand whether the number of excited states included in analysis were sufficient was gauged using the sum-over-states (SOS) polarizability (α_{SOS}) (see derivation in Appendix E). Eq. 7.4 is familiar in comparison to Eq. 7.3:

$$\alpha_{SOS,rs} = -2 \sum_{i \neq j} \frac{\langle \psi_i | \mu_r | \psi_j \rangle \langle \psi_j | \mu_s | \psi_i \rangle}{\varepsilon_i^{(0)} - \varepsilon_j^{(0)}} \tag{7.4}$$

where variables are defined as in Eq. 7.3. Here, instead of gauging the influence on one state on another, the sum of these contributions is an approximation of the physical property of polarizability, which was described in Chapter 6 with respect to the classical Stark analysis. While polarizability does not enter into the QM Stark analysis directly like the classical, this interesting

172

perspective allows us to see the "piece-by-piece" contributions of each electronic state that exists for a molecular probe.

The calculation of electronic properties of the ground state of porphin using the HF / sdd method and basis set in G09 yielded an "approximate polarizability," which is calculated using the perturbation theory "sum-over-states" method, in addition to an "exact polarizability," which is defined as $\alpha_{exact} = \frac{\partial^2 \varepsilon}{\partial \vec{E}^2}$, the second derivative of the molecular energy (ε) with respect to an applied electric field ($\vec{E}$). The polarizability, as a tensor, yields nine components, of which only three are non-zero for porphin and represent the components parallel to the three axes. For free-base porphin, this method yielded α_{xx} = 411.991 a.u.; α_{yy} = 383.297 a.u.; α_{zz} = 95.998 a.u., where a.u. are atomic units (see Appendix F for conversions between atomic units and typical units for a number of electronic properties). Not surprisingly, the *x*- and *y*-polarized components, which are in-plane, are about the same magnitude, while the *z*-polarized component, which is perpendicular to the plane, is much smaller. Figures 7.4a, 7.4b, and 7.4c plot the SOS polarizability (calculated with Eq. 7.4) as more electronic states are considered. Then, the percentage of the total SOS polarizability is compared at N=10 states are included and at N=250 states included. For the *x*- and *y*-components, the SOS polarizability is reasonably well-described at 86.9% and 90.0%, respectively. However, the slope of the progress to 100% total SOS polarizability becomes quite shallow by N=250 and likely requires many more states to fully describe the property. On the other hand, the *z*-component is not at all completely described at N=250 (only 14.1% of the total SOS polarizability), even though its absolute value is so much smaller than the in-plane components. This indicates that *z*-polarized electronic transitions tend to be much higher energy than in-plane transitions, which is also not surprising physically. This is

one of the first indications that the z-component of the internal electric field is not adequately

measured with a planar porphin model.
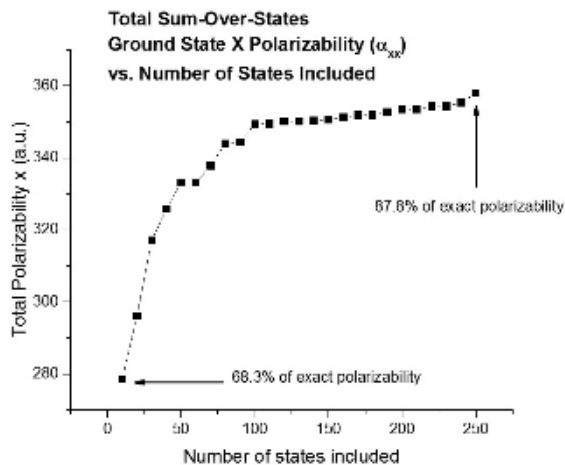
Figure 7.4a. The *x*-component of the summed SOS polarizability as more electronic state are included in the sum.
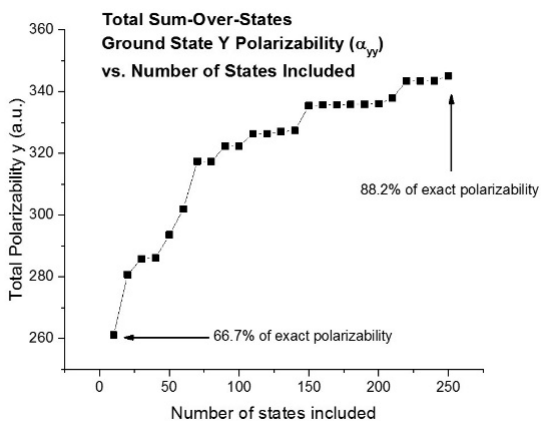
Figure 7.4b. The *y*-component of the summed SOS polarizability as more electronic state are included in the sum.
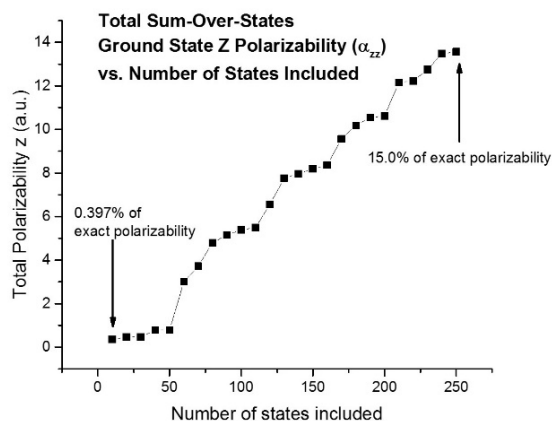
Figure 7.4c. The *z*-component of the summed SOS polarizability as more electronic state are included in the sum.

This examination of progress towards total SOS polarizability gives an indication that the number of electronic states calculated is a reasonably complete description for the *x*- and *y*-components but a 250 x 250 matrix would make for a burdensome calculation of internal electric field.

7.3.2 Completeness via the change in Q_x transition energy

Another viewpoint for gauging completeness of the matrix used for QM Stark analysis is to focus on the transition of interest, the Q_x transition in porphin, which is the value used for internal field calculation once the diagonalization step is complete. According to the denominator in Eq. 7.3, the influence of higher lying states scales with energetic "distance" from the states of interest. It seems, then, that there may be a point where electronic states are so far away from states 0 (ground state) and 1 (first-excited state) that it can be approximated as having the same influence on each. The *absolute* energies of these two states may continue to evolve as more electronic states are added to the matrix, but the difference in energy (i.e., the *transition* energy) of the two states will remain approximately the same after a certain number of electronic states are included.

Now, because we are not interested in the coupling between any two states, *i*, and *j*, we will now use the *n* to designate a state of interest (0 or 1) and *p* to designate a higher lying electronic state. First, the influence of each electronic state higher than the states of interest (*p* = states 2 – 252) were evaluated with respect to state *n* = 0 and then again to state *n* = 1 according to Eq. 7.3. Then, the difference in value for each individual state was calculated per Eq. 7.5 and plotted in order of energy. Here, α refers not to the summed polarizability value but to individual contributions from Eq. 7.3. The result are predictive plots for electronic states of influence on the

Q_x transition of porphin. The plots in Figs. 7.5a, 7.5b, and 7.5c show the behavior of individual x-, y-, and z-components for porphin; Fig. 7.5d shows all three components combined.

$$(\alpha_{xx,1p} - \alpha_{xx,0p}) + (\alpha_{yy,1p} - \alpha_{yy,0p}) + (\alpha_{zz,1p} - \alpha_{zz,0p}) =: \alpha_{1p} - \alpha_{0p} \qquad (7.5)$$
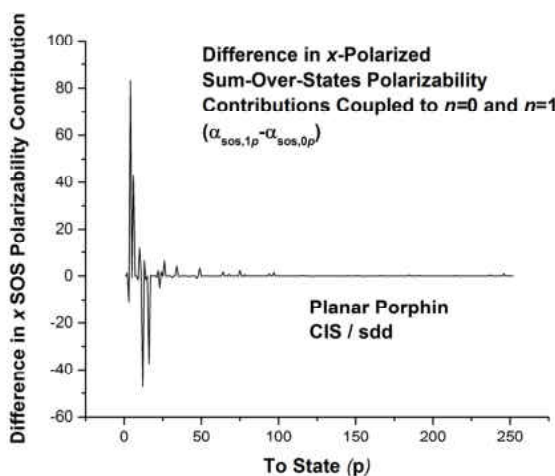
Figure 7.5a. Predictive plot for x-polarized states of influence on the Q_x transition of planar porphin.
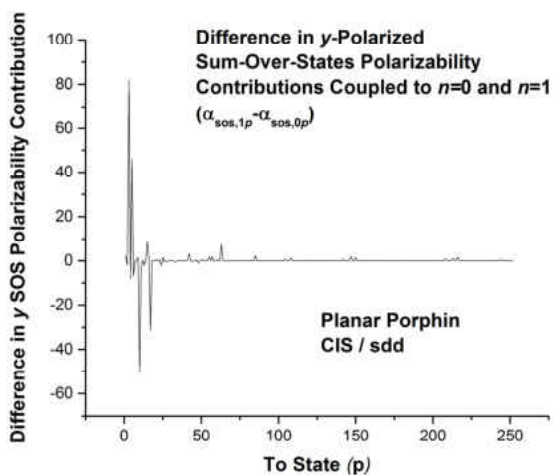
Figure 7.5b. Predictive plot for y-polarized states of influence on the Q_x transition of planar porphin.
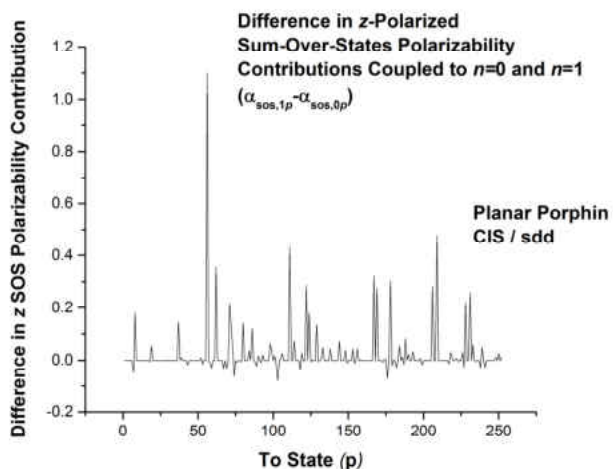
Figure 7.5c. Predictive plot for *z*-polarized states of influence on the Q_x transition of planar porphin.
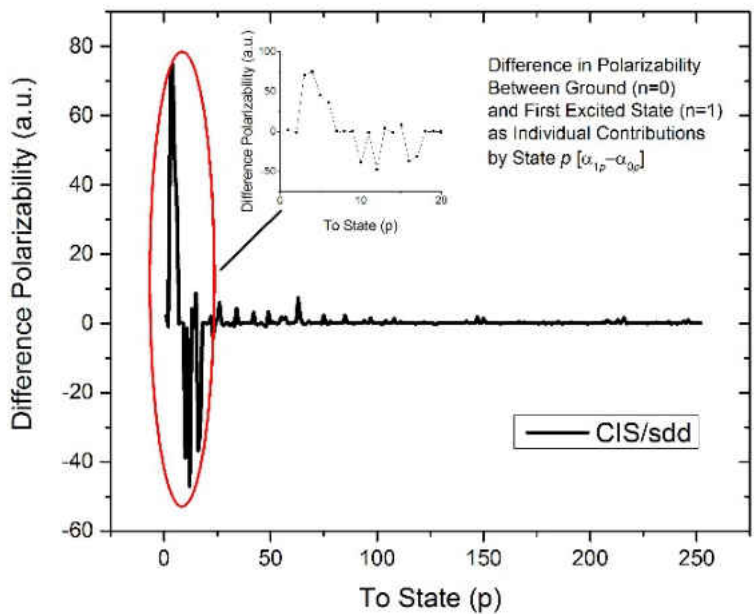
Figure 7.5d. Predictive plot for all states of influence on the Q_x transition of planar porphin. The inset shows more detail for states *p* = 1–20.

The plots for the *x*- and *y*-polarized states show convergence for influence certainly for *p* > 100 and possibly by *p* > 50. For *z*-polarized states, no convergence is seen by *p* = 252. This is reflective of the results for progress to SOS polarizability given in Figs. 7.4a–c. Again, this suggests that the *z*-

component of the internal electric field may not be adequately sensed by a planar probe molecule. Combining the three components, the results are similar to the more influential x- and y-component plots: states with very large influence from p = 1–20 followed by less influential states and essentially no additional influence past p = 100.

These results suggested that there may indeed be finite number of relatively influential electronic states on our transition of interest, which would allow for the implementation of the QM Stark analysis method with reasonable certainty of completeness. A 100 x 100 matrix (10 000 elements) is certainly still a large matrix, but a sizable reduction from the initially considered 250 x 250 matrix (62 500 elements).

7.3.3   Determining essential electronic states

Now that a reasonable selection of states based on the energy denominator of Eq. 7.3 could be made by finding a convergence point for the change in Q$_x$ transition energy, further narrowing the list of the most influential electronic states was undertaken. Based on perturbation theory, an equation to rank electronic states that couple directly to the states of interest, n = 0, 1, was derived. Then, further consideration of *indirectly* coupled electronic states to the states of interest was made. Indirect coupling refers to electronic states that may not affect the ground and first-excited state with a large transition dipole moment but certainly couple strongly to a state that does, having influence through an intermediate state.

Perturbation theory begins with the unperturbed energy of a molecular state n, $\varepsilon_n^{(0)}$, and adds a series of perturbation terms to yield the overall perturbed energy of that molecular electronic state, ε_n :

$$\varepsilon_n = \varepsilon_n^{(0)} + \varepsilon_n^{(1)} + \varepsilon_n^{(2)} + \varepsilon_n^{(3)} + \varepsilon_n^{(4)} + \cdots \tag{7.6}$$

For our application, the unperturbed energy is the Q_x transition energy of the porphyrin–substituted myoglobin with no electric field. The first-order energy correction, $\varepsilon_n^{(1)}$, for centrosymmetric porphin vanishes. This is because the correction is the simply $-\vec{E} \cdot \vec{\mu}$, in which the permanent dipole moment for either $n=0$ or $n=1$ is zero. Then, the second-order energy correction, $\varepsilon_n^{(2)}$, corresponds to the influence of electronic states directly coupled states on the Q_x transition energy. The third-order energy correction, $\varepsilon_n^{(3)}$, corresponds to the influence of the states indirectly coupled to the Q_x transition energy by way of the second-order correction states. The higher-order energy corrections also correspond to indirectly coupled states, however, not by way of the second-order correction states but one level removed by way of the third- and higher-order correction states. Then, influence on the transition of interest uses the same idea as presented in Eq. 7.5, i.e., to look not only at the influence on each state n individually, but at the difference between the two:

$$\varepsilon_1^{(2)} + \varepsilon_1^{(3)} + \varepsilon_1^{(4)} + \cdots - \varepsilon_0^{(2)} - \varepsilon_0^{(3)} - \varepsilon_0^{(4)} - \cdots \tag{7.7}$$

Formulas for the second- and third-order corrections (i.e., single- and double-state coupling, respectively) are given and discussed below. (See Appendix G for derivation of each of these correction terms.)

7.3.3.1 Single-state coupling (2nd-order energy correction)

A very familiar formulation arises for the second-order correction term (Eq. 7.8), looking quite like the SOS polarizability in Eq. 7.4.:

$$\varepsilon_n^{(2)} = \sum_{p \neq n} \frac{\langle n|\vec{E} \cdot \hat{\vec{\mu}}|p\rangle\langle p|\vec{E} \cdot \hat{\vec{\mu}}|n\rangle}{\varepsilon_n^{(0)} - \varepsilon_p^{(0)}} = \sum_{p \neq n} E_{\mu np}^2 \frac{\mu_{np}^2}{\varepsilon_n^{(0)} - \varepsilon_p^{(0)}} \qquad (7.8)$$

Here, the equation is specific for other electronic states, *p*, and their influence on states of interest, *n*. In this work, *p* will always be a higher-lying state (and may be referred to as such) because the states of interest are *n*=0,1. In addition, the familiar transition dipole moment also includes the perturbation, $E_{\mu np}$, the component of the electric field projected onto the dipole moment, μ_{np}. Specific for porphin, or other centrosymmetric probe molecules, the transition dipole moment lies along only one molecular axis and the numerator is simply the square of the scalar value. For the purposes of ranking influence of states *p* on states *n*, the direction of the transition is not relevant. Then, the influence on the *transition* (not just individual states) is given by:

$$\left\{\varepsilon_1^{(2)} - \varepsilon_0^{(2)}\right\} = 2E_{\mu 01}^2 \frac{\mu_{01}^2}{\varepsilon_1^{(0)} - \varepsilon_0^{(0)}} + \sum_{p \neq 0,1} \left(E_{\mu 0p}^2 \frac{\mu_{0p}^2}{\varepsilon_p^{(0)} - \varepsilon_0^{(0)}} - E_{\mu 1p}^2 \frac{\mu_{1p}^2}{\varepsilon_p^{(0)} - \varepsilon_1^{(0)}} \right) \qquad (7.9)$$

Because we are only comparing the influence of higher-lying states, the absolute value of this expression is not important and the perturbing electric field terms, $E_{\mu np}^2$, are set to unity. In addition, because the first term is a constant, it can be neglected when used for ranking purposes.

Considering only the directly coupled states on the transition of interest, a unique form of the Hamiltonian matrix arises (Eq. 7.10). This means that the only elements that are occupied are contained on the diagonal and in the first and second rows and columns. All other off-diagonal elements are set to zero. This was a significant improvement over the previous matrix dimensions and contributed to the analysis routine's efficiency. In addition, a new variable is introduce, A_i, to indicate matrix position, not in terms of energy, but in terms of influence. The

matrix in Eq. 7.10 has the states of interest, $n=0$ and 1, in the first and second positions. The third position, then, is occupied by the higher-lying state, p, that has the highest value for Eq. 7.9.

$$\hat{H} = \begin{pmatrix} \varepsilon_0^{(0)} & -\vec{E}\cdot\vec{\mu}_{01} & -\vec{E}\cdot\vec{\mu}_{0A_2} & -\vec{E}\cdot\vec{\mu}_{0A_3} & \cdots & -\vec{E}\cdot\vec{\mu}_{0A_N} \\ -\vec{E}\cdot\vec{\mu}_{10} & \varepsilon_1^{(0)} & -\vec{E}\cdot\vec{\mu}_{1A_2} & -\vec{E}\cdot\vec{\mu}_{1A_3} & & -\vec{E}\cdot\vec{\mu}_{1A_N} \\ -\vec{E}\cdot\vec{\mu}_{A_20} & -\vec{E}\cdot\vec{\mu}_{A_21} & \varepsilon_{A_2}^{(0)} & 0 & \cdots & 0 \\ -\vec{E}\cdot\vec{\mu}_{A_30} & -\vec{E}\cdot\vec{\mu}_{A_31} & 0 & \varepsilon_{A_3}^{(0)} & & 0 \\ \vdots & & & & \ddots & \\ -\vec{E}\cdot\vec{\mu}_{A_N0} & -\vec{E}\cdot\vec{\mu}_{A_N1} & 0 & 0 & & \varepsilon_{A_N}^{(0)} \end{pmatrix} \qquad (7.10)$$

7.3.3.2 Double-state coupling (3rd-order energy correction)

The third-order energy correction involves not just the influence of higher-lying state, p, on states of interest, n, but also the influence of another higher-lying state, q, on state p, hypothesizing that strong interaction between the two remote states could have an indirect influence on the transition of interest:

$$\varepsilon_n^{(3)} = -\sum_{p\neq n}\sum_{q\neq n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qn}} \frac{\mu_{np}\mu_{pq}\mu_{qn}}{\left(\varepsilon_n^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)}-\varepsilon_q^{(0)}\right)} + E_{\mu_n}\mu_n \sum_{p\neq n} E_{\mu_{np}}^2 \frac{\mu_{np}^2}{\left(\varepsilon_n^{(0)}-\varepsilon_p^{(0)}\right)^2}$$

$$(7.11)$$

This evaluation of indirectly coupled states was for the purpose of populating the off-diagonal elements in Eq. 7.10. The third-order correction difference between states $n=0$ and 1 then is expressed:

$$\left\{\varepsilon_1^{(3)} - \varepsilon_0^{(3)}\right\} = 2E_{\mu_{01}} \frac{\mu_{01}}{\varepsilon_1^{(0)}-\varepsilon_0^{(0)}} \sum_{p\neq 0,1}\left[E_{\mu_{0p}} E_{\mu_{1p}} \left(\frac{\mu_{0p}\mu_{1p}}{\varepsilon_p^{(0)}-\varepsilon_0^{(0)}} + \frac{\mu_{0p}\mu_{1p}}{\varepsilon_p^{(0)}-\varepsilon_1^{(0)}}\right)\right] + $$

$$\sum_{p\neq 0,1}\sum_{q\neq 0,1}\left[E_{\mu_{pq}}\mu_{pq}\left(E_{\mu_{0p}}E_{\mu_{0q}}\frac{\mu_{0p}\mu_{0q}}{\left(\varepsilon_p^{(0)}-\varepsilon_0^{(0)}\right)\left(\varepsilon_q^{(0)}-\varepsilon_0^{(0)}\right)} - E_{\mu_{1p}}E_{\mu_{1q}}\frac{\mu_{1p}\mu_{1q}}{\left(\varepsilon_p^{(0)}-\varepsilon_1^{(0)}\right)\left(\varepsilon_q^{(0)}-\varepsilon_1^{(0)}\right)}\right)\right]$$

$$(7.12)$$

The electric field in this expression cannot simply be set to unity as with the second-order correction, but it is reasonable to assume that the values are about equal. For the evaluation of this expression, the field was set to 1 MV/cm = 1.94 x 10^{-4} a.u.

The first term in Eq. 7.12 can be called the "weighting term" and the second term the "ranking term." The weighting term considers only the direct-coupling influence of state p. The ranking term includes the indirectly coupled state, q, but mitigated by the directly coupled state, p. However, thoroughly considering a set of higher lying electronic states (N=2–100 in this case) means that each state has a "turn" acting as p and another acting as q. This does not have an effect on the ranking term but does have an effect on the weighting term, depending on which state is under consideration as state p. For example, if the (p,q) pair is the third excited state and the twenty-third excited state, the value of $\left\{\varepsilon_1^{(3)} - \varepsilon_0^{(3)}\right\}_{p=3,q=23} \neq \left\{\varepsilon_1^{(3)} - \varepsilon_0^{(3)}\right\}_{p=23,q=3}$. This is problematic for this application because the interest is in evaluating *pairs* of higher-lying electronic states. This means that any two mirrored values in the Hamiltonian matrix will be equal (i.e., $-\vec{E} \cdot \vec{\mu}_{3,23} = -\vec{E} \cdot \vec{\mu}_{23,3}$), so the corresponding terms from Eq. 7.12 should also be equal. The weighting term can be adjusted to account for this. The coefficient, 2, is distributed as each state in the pair acts as p' (the directly coupled state under consideration) in turn:

$$\frac{\mu_{01}}{\varepsilon_1^{(0)} - \varepsilon_0^{(0)}} \sum_{\substack{p \neq 0,1 \\ p \neq q}} \left[\left(\frac{\mu_{0p'=p}\mu_{1p'=p}}{\varepsilon_{p'=p}^{(0)} - \varepsilon_0^{(0)}} + \frac{\mu_{0p'=p}\mu_{1p'=p}}{\varepsilon_{p'=p}^{(0)} - \varepsilon_1^{(0)}} + \frac{\mu_{0p'=q}\mu_{1p'=q}}{\varepsilon_{p'=q}^{(0)} - \varepsilon_0^{(0)}} + \frac{\mu_{0p'=q}\mu_{1p'=q}}{\varepsilon_{p'=q}^{(0)} - \varepsilon_1^{(0)}} \right) \right]$$

(7.13)

Once this accommodation had been made, the almost 10,000 possible pairs were evaluated and plotted in Fig. 7.6. The comparison shows that less than 10% of those pairs have any possibility of a significant influence, again greatly limiting the size of the Hamiltonian matrix that must be evaluated.
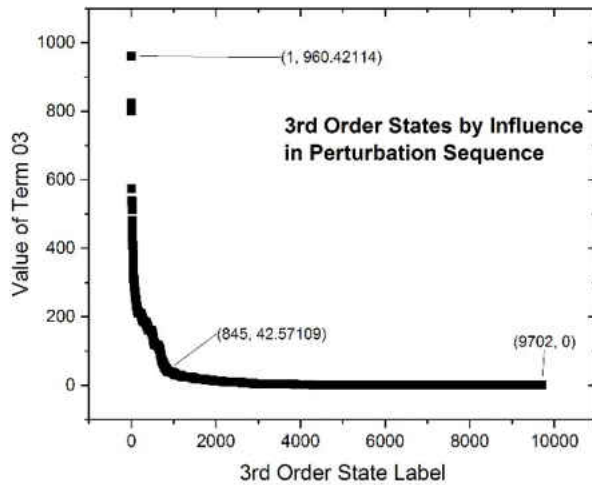
Figure 7.6. The values of the third-order (double-state coupling) terms show that the influence drops off steeply.

7.4 Internal Electric Field Calculation with *Stark06*

The next step then was to do the actual internal electric field calculation using the QM Stark analysis method, increasing the number of included states for each calculation. First, the states were added according to energy to see whether the convergence region predicted by Figs. 7.5a-d was accurate for the calculated internal electric field, i.e., that the field value did not change appreciably with higher energy states than $\sim N=100$. Then, another set of internal electric field calculations was performed, i.e., a "perturbation sequence." Here, calculations using only directly coupled states were initially performed, where states were added one by one to the internal field calculation in the order of influence determined by Eq. 7.9 to yield a converged value. Following

that determination, indirectly coupled states were added to the number of directly coupled states required for a converged value in the order of influence determined by Eq. 7.12

The internal electric field calculation from experimental hole-burning Stark spectra was performed with the in-house *Stark06* program (written by a former postdoctoral research associate, Dr. Barry Prince, which had implemented QM analysis. The number of electronic states included in the analysis was flexible and designated by the user for each calculation. Thus an effective comparison between an analysis using just a few p states and an analysis using 100 p states was possible.

7.4.1 *Stark06* program

This line-fitting program takes three types of input to calculate internal electric field: 1) spectral data; 2) excited-state values (state energy and transition dipole moment) for the Hamiltonian matrix; 3) line-fitting parameters. Table 7.1 lists values required for each type. The spectral data is uploaded in the form of a .DAT file for the actual data points for the Stark spectra and also a .SPL file for the environmental parameters like applied field strength. The excited-state data is uploaded in the form of a .HST file. An example of each of these unique file types is given in Appendix B.

Table 7.1. Three types of required input for *Stark06* program.

Spectral Data	Hamiltonian States	Fitting Parameters
FWHM of the Lorentzian-shaped spectral hole before E_{ext} is applied (cm^{-1})	Electronic state energies (cm^{-1})	Lorentz local field tensor (*f*)
Angle of the laser polarization (hole-burning & scanning with respect to E_{ext}	Transition dipole moment for each pair of states (D)	Number of steps in orientational averaging over all Euler angles
Applied Field E_{ext} (V/cm)	Polarization of the transition (*x,y,z*)	Number of excited states (*N*) to include in the Hamiltonian matrix
Spectral data: fluorescence intensity vs. frequency applied (cm^{-1})		Initial internal electric field (E_{int}) guesses for each component
		Vertex: initial excursion distance for optimization

This is the general process for *Stark06*. The analysis starts with diagonalization of the Hamiltonian using an initial guess for the internal electric field, which yields *H'*, the perturbed energy of the transition of interest. Then, a Stark spectrum is generated for the entered parameters. Finally, this generated spectrum is compared to the experimental spectral holes loaded as spectral data files. The process is repeated until pre-determined fitting criteria are met. This process is detailed in the following.

The first diagonalization step utilizes the Hamiltonian states information: electronic state energies, transition dipole moment, and the polarization of each transition. The matrix is then built but requires a value for $\vec{E}$ for the off-diagonal elements. This value is the *total* electric field, defined earlier by Eq. 6.6:

$$\vec{E}_{tot} = \vec{E}_{int} + \overleftrightarrow{f} \cdot \vec{E}_{ext} \qquad\qquad (6.6)$$

The *external* electric field is the value applied during the experiment as the voltage applied over the gap between the electrodes. The local field factor is given by the Lorentz field factor (Eq. 6.5). Here the value for $\overleftrightarrow{f}$ is chosen to be 2.1 (see Refs. [2a, 3]). Then, an initial guess for *internal* electric field must be provided to generate the matrix. This is loaded as individual spatial components, *x, y,* and *z.* Each component is then considered separately with the transition dipole moment values of a particular polarization contributing to the corresponding component. For example, if no *z-*polarized transition dipole moment values are included in the matrix, then no *z-*component of the internal electric field can be generated. The Stark shift is then the difference between the unperturbed Q_x transition energy entered as part of the .HST file and the calculated perturbed value generated by the diagonalization procedure.

The next step that the program undertakes is to generate a "stick spectrum" based on the perturbed energy. However, if only a single value was used for the placement of the perturbed energy, there would not be a distributed spectrum. The distribution is a result of the various orientations of chromophores with respect to the applied field (see Section 6.2.1). This spatial distribution of probe molecules can be taken into account through an orientational averaging procedure because experiments were specifically conducted in an amorphous, glass-forming solvent, meaning the statistical distribution of orientations is truly random. Thus, the weight of each orientation is equal and calculating the intensity of each orientation at its respective energetic position (*v'*) is based only on the angle of its transition dipole moment with respect to the burning and scanning laser (see Eqs. 6.11–6.13). It was previously determined that stepping through the angles θ (with a range of 180°) and φ (with a range of 360°) in 6° increments was sufficient resolution for the analysis. Therefore, 90 considered orientations (180/6 +360/6=90

steps) generate 90 "sticks" at individual perturbed energies and with individual intensity values. Once the stick spectrum has been generated, a line is fit over the top.

Once the theoretical stick spectrum is produced for a particular internal electric field value, it is compared to the corresponding experimental spectra (many of which are included in the fitting process simultaneously). Then the goodness of fit for all spectra included in the analysis is quantified with a least squares fitting process. The smaller the least-squares residual value, the better the internal electric field value. The process of adjusting internal electric field value to re-diagonalize the matrix and produce a new stick spectrum for comparison to the experimental spectrum is done iteratively. New internal field values are chosen according to a "vertex" value specified by the user. In these calculations, a high value of 1000 was chosen, meaning that smaller steps in internal electric field guesses were made. The program terminates when a minimum residuals value is found.

For the following "energy sequence" and "perturbation sequence," a set of eight Stark spectra (shown in Fig. 6.4) previously acquired were re-analyzed, with the four split holes (laser polarization parallel to the external field) weighted double, i.e., analyzed twice per iteration, because they exhibited a better signal-to-noise ratio than those recorded with the laser polarization perpendicular to the external field. Using this set of experimental spectra allowed direct comparison to the results presented in Geissinger (2004)[2a] and Geissinger (2005).[2b] As electronic states were added for each new analysis, the effect was new fit "landscape" with a unique set of local maxima and minima. For each analysis, the goal was to find the new landscape's global minimum. A map of the converged value's "fit landscape" is given in Fig. 7.7a–c with the *x*-axis corresponding to the *x*-components of the internal electric field ($\vec{E}_x$) and the *y*-

axis corresponding to the *y*-components of the internal electric field ($\vec{E}_y$). The colored contours correspond to fit values (least-squares residuals) with a low value being a minimum. The plots contain the same data but each panel further "zoomed in" to a pairing of $\vec{E}_x$ and $\vec{E}_y$.
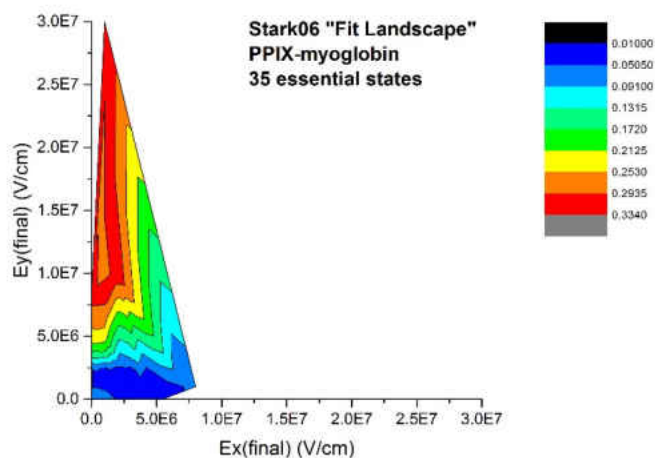
Figure 7.7a. The "fit landscape" for the analysis of Stark spectra created in PPIX-myoglobin with the inclusion of 35 essential states in the Stark06 analysis. The borders of the colored region represent values where the fit value was essentially infinity.
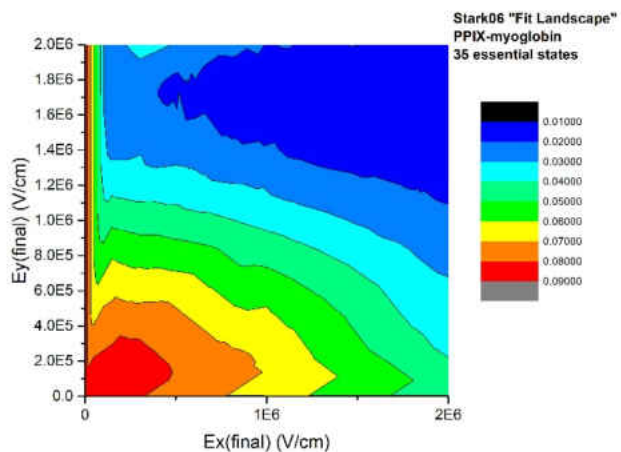
Figure 7.7b. The minimum area from the above plot seems to restrict the best fit values to less than 2.5 MV/cm for E_y and less than 7.5 MV/cm for E_x. Shown here is a square area for each component less than 2.0 MV/cm.
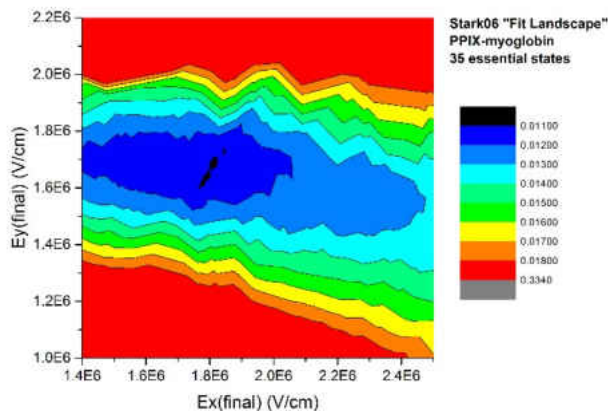
Figure 7.7c. Further resolution of the minimum area pinpoints the global minimum to 1.7–1.8 MV for E_x and 1.6–1.7 MV/cm for E_y.

7.4.2 Energy sequence

First, internal electric field values were calculated with the Stark06 program to a minimum least-squares residual value (approximately 0.13 for this spectral data and Hamiltonian states). The higher-lying states, p, were added in order of energy (i.e., $p = 2$ was the state involved in the Q_y transition, $p = 3$ was the state involved in the B_x transition, etc.) to the states of interest, $n = 0,1$. The matrix dimensions started with a 3 x 3 matrix and states were added one by one until about $p = 17$; then, states were added five by five until a 140 x 140 matrix as defined by the row matrix in Eq. 7.10 was achieved. The results for each spatial component of the electric field, $\vec{E}_x$, $\vec{E}_y$, $\vec{E}_z$, are given in Figs. 7.8a, 7.8b, and 7.8c respectively.
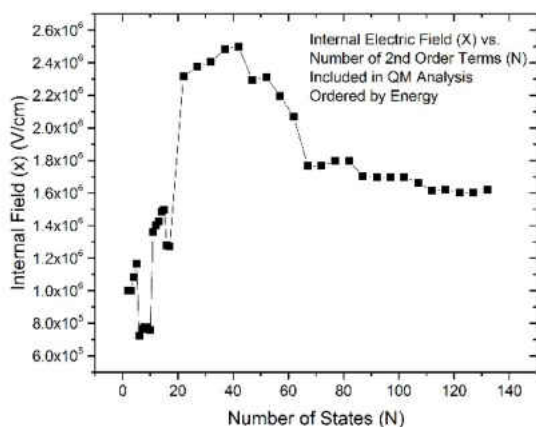
Figure 7.8a. Evolution of $\vec{E}_x$ as higher-lying electronic states are added to the analysis in order of energy.
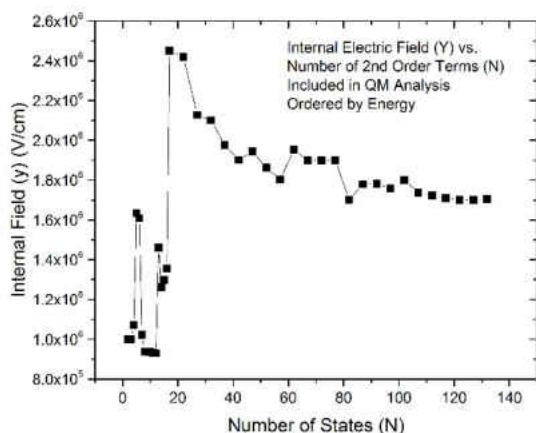
Figure 7.8b. Evolution of $\vec{E}_y$ as higher-lying electronic states are added to the analysis in order of energy.
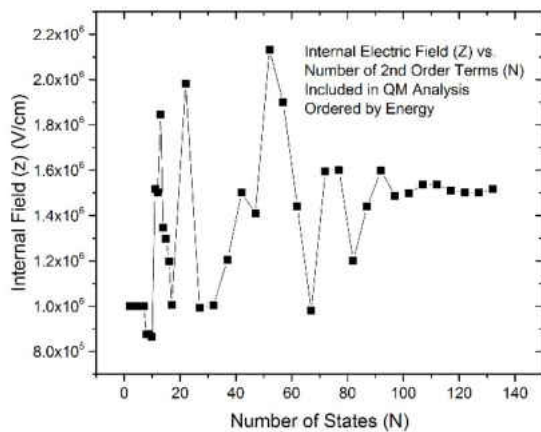
Figure 7.8c. Evolution of $\vec{E}_z$ as higher-lying electronic states are added to the analysis in order of energy.

For the *x*- and *y*-components, Figs. 7.8a and 7.8b both show quite a bit of oscillation for about N = 100 electronic states included until the value stabilizes considerably at $\vec{E}_x$ = 1.701 MV/cm and $\vec{E}_y$ = 1.746 MV/cm. For $\vec{E}_x$, the values prior to convergence vary over a range from 0.7 – 2.3 MV/cm; for $\vec{E}_y$, the values vary over a range from 0.9 – 2.5 MV/cm. In contrast to the in-plane components, however, the out-of-plane *z*-component, Fig. 7.8c does not show a gradual convergence for $\vec{E}_z$ in a range of 140 states but a rather sudden one about N = 100 like the other two components. This trend seems unlike the other two and may indicate that the energy denominator in the SOS polarizability equation could be the key component in convergence. Three conclusions are apparent from this exercise: 1) this QM Stark analysis method is sensitive to the number of Hamiltonian states included in the process; 2) convergence in eventually achieved; 3) the method of predicting convergence described in Section 3.2 and illustrated in Fig. 7.5 appears to the useful.

As a way to more generally assess the impact of indirectly coupled states, more electronic states were added to the set of N = 100 determined in the exercise above. States were added in this manner: electronic transitions from each state p = 2 – 100 to states 2 -252 were considered with Eq. 7.4. The resulting values were sorted for each state p in order of influence. Then, the highest influence value for each state p was added to the 100 x 100 converged-value row matrix from above, and the internal electric field value calculated (labeled as "2nd order" in Fig. 7.9 below). That calculation was followed by another in which the *two* highest influence values were added ("3rd order" in Fig. 7.9). This was continued until four indirectly coupled states were included. This exercise is equivalent to adding a pair of mirrored off-diagonal elements in every

row for each calculation. Figure 7.9 shows that the value determined with only directly coupled states does not change appreciably with the addition of indirectly coupled states. This suggests that for this particular set of spectral and Hamiltonian-states data, the electronic states directly coupled to the transition of interest are sufficient for QM Stark analysis.
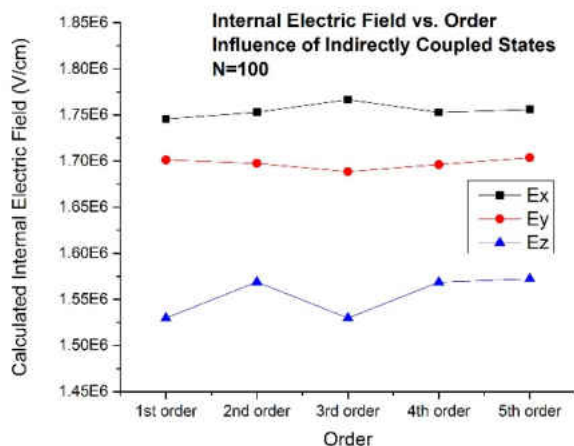
Figure 7.9. Relatively stable values for $\vec{E}_{int}$ when states indirectly coupled to the transition of interest added 100 at a time based on coupling strength to higher-lying states, p.

7.4.3 Perturbation sequence

Next, a more intentional series of calculations was performed. Instead of adding higher-lying electronic states, p, in order of energy, the higher-lying states were ranked in order of influence according to Eq. 7.9 for directly coupled states. Here, we refer to adding electronic states in terms of matrix position, A_i. The internal field calculations were performed as in the energy sequence but with the new ordering according to the perturbation theory equations. Results are given in Figs. 7.10a, 7.10b, and 7.10c.
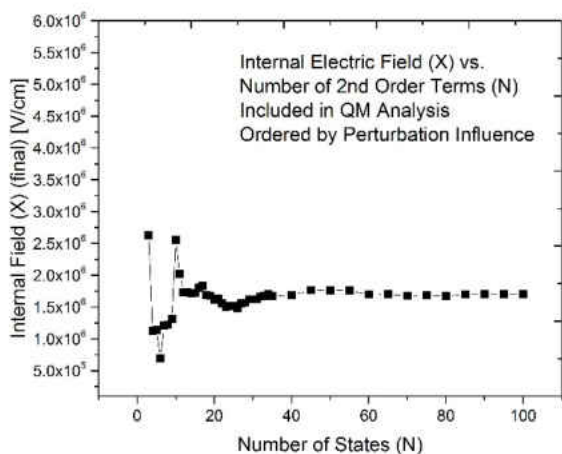
Figure 7.10a. Evolution of $\vec{E}_x$ as higher-lying electronic states are added to the analysis in order of perturbation influence.
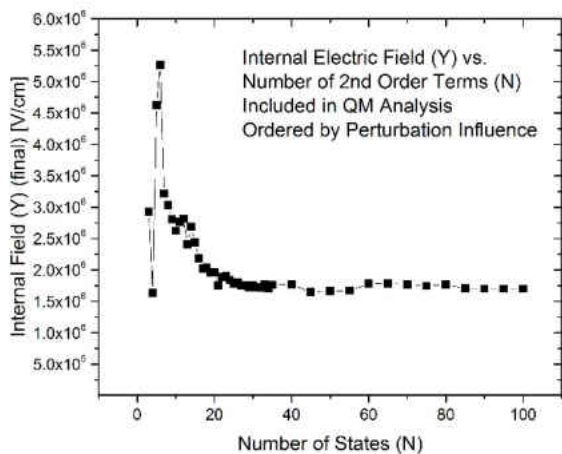
Figure 7.10b. Evolution of $\vec{E}_y$ as higher-lying electronic states are added to the analysis in order of perturbation influence.
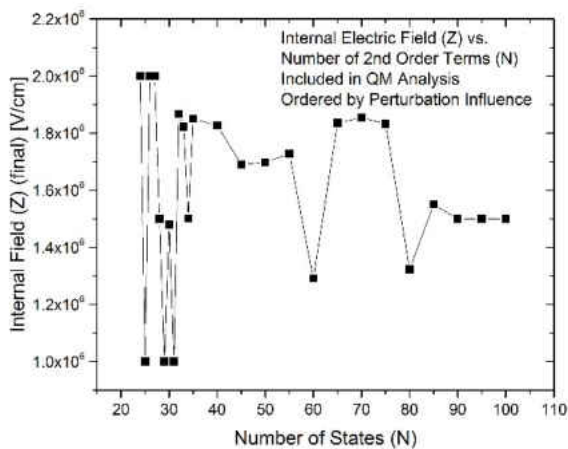
Figure 7.10a. Evolution of $\vec{E}_z$ as higher-lying electronic states are added to the analysis in order of perturbation influence.

The results for the perturbation sequence has similarities to the energy sequence in that the values of the internal electric field for smaller matrices oscillate and then converge for the in-plane components. However, the convergence point appears much more quickly, at approximately A_i = 35 states rather than 100. The internal electric field value at this point is $E_{int,x}$ = 1.700 MV/cm and $E_{int,y}$ = 1.700 MV/cm. The calculations were performed for this set of calculations one by one until 35 states had been added to the matrix, and then five by five until 100 states had been added. The out-of-plane z-component clearly lacks at convergence point at 35 states like the in-plane components did. Further conclusions from this set of calculations are: 1) faster convergence of the internal electric field value can be achieved for the QM Stark analysis by adding the states in order of perturbation influence according to Eq. 7.9; 2) the out-of-plane component is not adequately modeled by a planar porphin model.

The results illustrated in Fig. 7.9 suggested that indirectly coupled states, q, likely are not influential for this particular data set. However, using Eq. 7.12, the effect was tested more quantitatively. The pairs of states plotted in Fig. 7.6 were added one at a time to the 35 x 35 row matrix containing directly coupled states and the internal electric field calculated as above. Interestingly, the initial matrix dimensions had to be expanded slightly to accommodate two indirectly coupled states that acted through directly coupled states that did not initially appear in the 35 x 35 row matrix. This indicates that the transition of interest might indeed be indirectly affected by a higher-lying electronic state, q, to a significant degree via a weakly coupled state, p, that does not have a relatively strong influence. In the end, after making approximately 14 q-state additions to the matrix, the internal electric field value was hardly affected and it was confirmed that indirectly coupled states do not have influence for this data set.

For the QM Stark analysis of PPIX-myoglobin with the current set of Hamiltonian state values, 35 essential directly coupled states were determined to be sufficient for convergence. These values are provided in Appendix H. The two most influential states, not surprisingly, are the highly allowed and close in energy Soret transitions, B_x and B_y.

7.4.4 Perturbed visible transition energies with QM Stark analysis

In addition to returning an internal electric field value, the Stark06 program also returns the perturbed transition values for all of the states included in the matrix as a result of the diagonalization process. Plotting the transition energies of the four visible transitions as returned with the energy sequence of calculations (Figs. 7.11a–d), the pattern of oscillation and eventual convergence is familiar from the internal field values themselves as the changes are only fractions of the total transition energy. This is true for the *x*- and *y*-components of the field, which correlates with the polarizations of the four visible transitions: Q_x, Q_y, B_x, B_y.
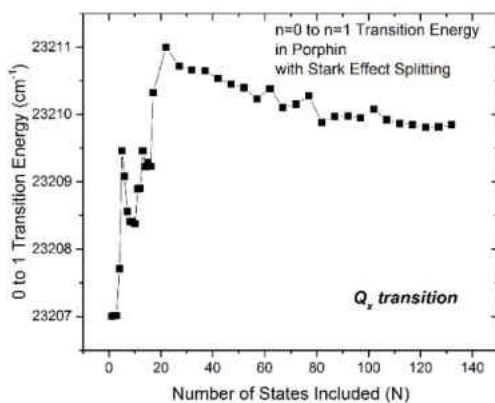
Figure 7.11a. Evolution of the Q_x transition energy as more higher-lying electronic states, *p*, are added to the analysis in order of energy.
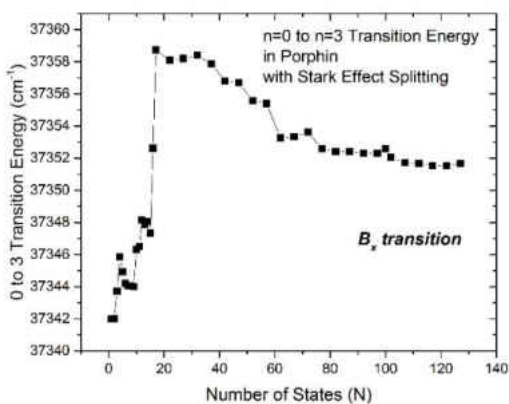
Figure 7.11b. Evolution of the B_x transition energy as more higher-lying electronic states, *p*, are added to the analysis in order of energy.
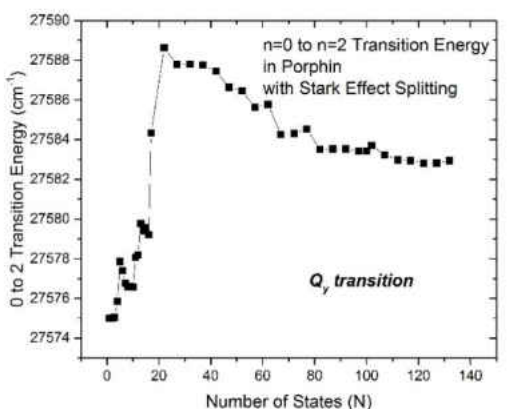
Figure 7.11c. Evolution of the Q_y transition energy as more higher-lying electronic states, *p*, are added to the analysis in order of energy.
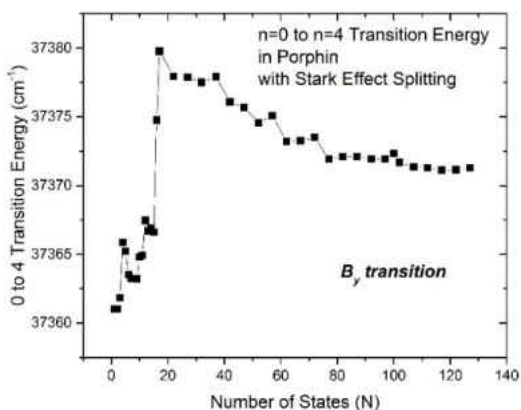
Figure 7.11d. Evolution of the B_y transition energy as more higher-lying electronic states, *p*, are added to the analysis in order of energy.

7.4.5 Non-planar porphin model

Based on the conclusion from Section 4.3 that a planar porphin model is not adequate to calculate the out-of-plane, $\vec{E}_z$ value, a non-planar porphin model was explored. That the heme plane is distorted in proteins is well-established by x-ray crystal structures, and patterns of distortion have been linked functionally to various heme proteins with various enzymatic tasks by the Shelnutt group.[1, 4] Six out-of-plane conformations of the heme macrocycle (i.e., the porphin core), were characterized and illustrated in Fig. 7.12.
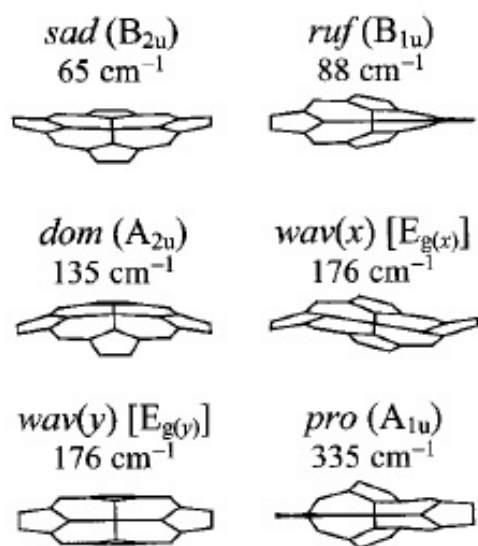
Figure 7.12. Six out-of-plane conformations of the heme macrocycle: saddling, ruffling, doming, waving (x- and y-polarized), and propellering. (Figure from Ref. [1])

It was hypothesized that retaining the *x–y* symmetry while distributing the spatial coordinates along the *z*-axis would increase the intensity of *z*-polarized transition dipole moments without losing the intensity of the in-plane transition dipole moments. For this purpose, a domed porphin structure was chosen because of the symmetry as well as the occurrence of this conformation specifically in myoglobin.

Once a method to optimize a domed structure was determined (details in Chapter 8), the displacement along the z-axis of the pyrrole nitrogens with respect to eight hydrogen atoms on the edge of the ring was adjusted. A structure with a z-displacement of 0.400000 Å was used to execute the same perturbation sequence procedure described in Section 4.3. The results showed that the sum of z-polarized transition dipole moment values for 250 excited states coupled to states of interest, $n = 0,1$, did indeed increase by 25% while retaining comparable magnitude for in-plane polarizations. The latter is expected, as the spectroscopic signature of the porphyrin is clearly identifiable when incorporated into the protein and subject to the distortion induced by the protein environment. Subsequent plots of Eq. 7.5 for this structure gave predictions for the convergence of internal electric field values. The in-plane components converged similarly to the plots in Figs. 7.5a and 7.5b for planar porphin. The z-component did not converge as hoped. The exercise was repeated for an increased displacement of porphin along the z-axis of 0.600000 Å. Results were similar. Displacement past this point was deemed both unphysical (with 0.4–0.5 Å being the average for myoglobin structures) and difficult to optimize.
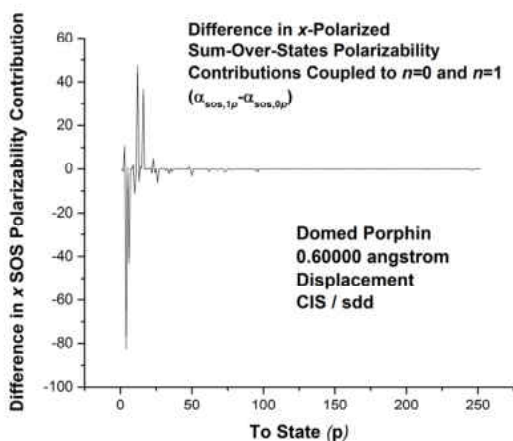
Figure 7.13a. Predictive plot for x-polarized states of influence on the Q_x transition of domed porphin (0.6 angstrom displacement).
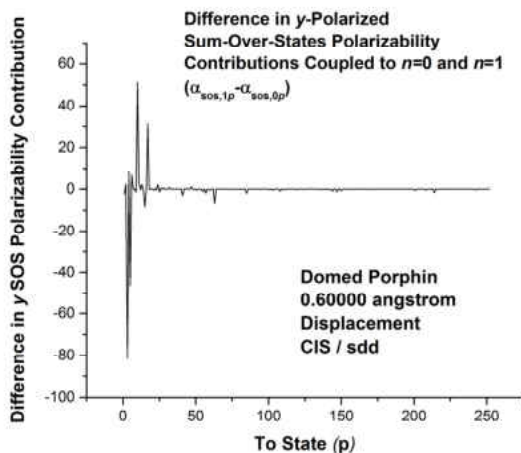
Figure 7.13b. Predictive plot for *y*-polarized states of influence on the Q_x transition of domed porphin (0.6 angstrom displacement).
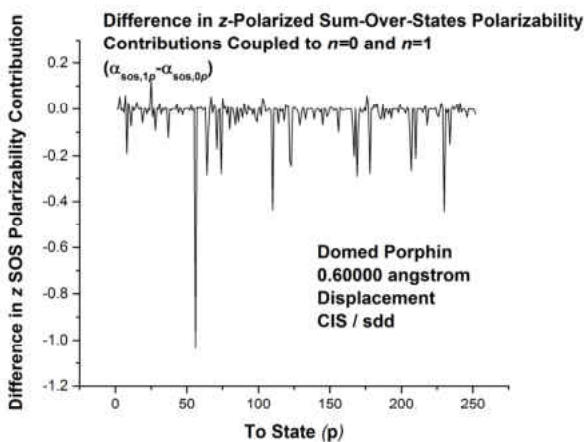
Figure 7.13c. Predictive plot for *z*-polarized states of influence on the Q_x transition of domed porphin (0.6 angstrom displacement).

However, the *z*-polarized transitions did seem to affect the transition of interest differently as the structure took on a greater domed character. As a way to begin to quantify the effect, the *z*-polarized transitions dipole moments coupled to states 0 and 1 were summed as the absolute value ($\sum |\mu_{z,np}|$) (where *n*=0,1 and *p*=2-252) for three conformations of porphin: 14.0758 a.u. for the planar model, 17.6648 a.u. for the 0.400000 Å domed model, and 18.7900 a.u. for the 0.600000 Å domed model. This was the desired trend in examining a domed conformation: an increase in magnitude of *z*-polarized transition dipole moment values.

In addition, the sum of difference z-polarized sum-over-states polarizability values $(\sum|\alpha_{SOS,z,1p} - \alpha_{SOS,z,1p}|$, i.e., the sum of values in Figs. 7.5c and 7.13c) for each of the three conformations was calculated. This number attempts to capture how disparately the set of z-polarized transitions affects the two states in the transition of interest, i.e., whether, in general, they tend to influence the ground state over the first excited state or vice versa. Interestingly, this number was calculated as 285.342 a.u. for the planar model, then dropped dramatically to 21.108 a.u. for the 0.400000 Å domed model, and then finally to 8.3407 a.u. for the 0.600000 Å domed model. While this does not definitively show an "improvement" of this non-planar model over the planar one for purposes of calculating the out-of-plane internal electric field, it does encourage further exploration of the topic.

7.5 Theoretical evidence for convergence

Finally, is the observed convergence of internal electric field values in the QM Stark analysis with a relatively low-energy set of electronic states something to be expected? If the theory behind the QM Stark equation invokes an infinite number of electronic states, is it reasonable to see convergence at a finite number of values and as few as the results gave? As rare as it might be, could there be an electronic state that lies high above the observed with a transition dipole moment coupled to the states of interest that is so large that it entirely overcomes the energy denominator to dominate the analysis? Two theoretical pieces of evidence provide reasonable assurance that this special case will not arise and that convergence with this analysis method is entirely to be expected. See also Appendix I for derivations related to these two theoretical topics.

7.5.1   Oscillator strength & maximum transition dipole moment

The first piece of evidence lies in the definition of oscillator strength (f_{osc}). This is a value well known to experimental spectroscopists as quantitative measure of a transition's intensity. It is typically calculated by integrating under the peak of experimental spectra and expressed as the ratio of this value to the value obtained for the harmonic oscillator, i.e., the "ideal" transition.[5] However, oscillator strength can also be defined as a bridge between classical notions of transition intensity and the quantum-mechanical notion, i.e., the transition dipole moment. This expression essentially is the ratio of transition dipole moment of the species of interest to the transition dipole moment of the ideal harmonic oscillator:

$$f_{Osc,pq} = \frac{8\pi^2 m_e v}{3 e^2 h} \left| \vec{\mu}_{pq} \right|^2 \tag{7.14}$$

where m_e = electron mass, e = elementary charge, h = Planck's constant, v = frequency of the transition of interest, and μ_{01}= transition dipole moment between states p and q. Because this value is a ratio, $f_{osc} = 1$ implies a fully allowed transition. This is not a perfect comparison as a molecular transition like porphin cannot be reduced to the harmonic oscillator model. Indeed many experimental values for strongly allowed transitions, like the Soret transitions in porphin, result in oscillator strength values greater than one. However, it is a reference with a physical basis that can be calculated exactly.

Examining Eq. 7.13, we can see that frequency of the transition is inversely related to transition dipole moment. If we are concerned about the possibility of a fully allowed transition far away from the transition of interest, this equation states that the higher energy the transition is, the lower the transition dipole moment will be. This is very helpful when implementing Eq. 7.8

because influence decreases for transitions "far away" from the transition of interest via the denominator, but now Eq. 7.13 also guarantees that the transition dipole moment coupling the states will also decrease. In addition, it is the *square* of the transition dipole moment in Eq. 7.8 that appears in the numerator, which even more rapidly diminishes the influence of very high-lying states.

A quantitative example allows us to gauge the possible "missing" influence from states past the $N = 100$ convergence region determined above. Choosing state $n = 100$ (energy $\varepsilon = 0.4133$ Ha) as an electronic state well within the convergence region, an "ideal transition" from ground state has a transition dipole moment of 1.9 a.u. with Eq. 7.13. This compares to the most influential Soret (*B*-band) transition of 3.8 a.u. Then, estimating contribution of a state with those excited-state properties with Eq. 7.4, the maximum influence on the ground state is $\alpha_{sos} \approx 8.8$. This would be about 2.6% of the total sum-over-states polarizability for the lowest 100 states combined. It can be safely assumed that such an "ideal" higher-lying transition is exceedingly rare and, if existed, would be marginally influential.

7.5.2   *f-sum rule*

The Thomas-Reiche-Kuhn f - sum rule[6] says that the sum of oscillator strengths from one state n to all other states is equal to 3Z, where Z = number of electrons in the molecule:

$$\sum f_{osc,np} = 3Z \tag{7.15}$$

In answer to the concern that the convergence approach to this QM Stark analysis must neglect an *infinite* set of electronic state, Eq. 7.14 states that there is a finite value for oscillator strengths, therefore any other quantity derived from transition dipole moment, e.g., polarizability, must

also be finite. These electronic properties are measurable and no doubt have finite values. It can then be concluded that an infinite set of electronic states is not necessary for this analysis.

7.6 Conclusions

The quantum-mechanical Stark analysis method conceived of previously showed promise for more accurate determination internal electric fields. A line-fitting program, *Stark06,* had be written to implement the method. However, implementation was challenging because of the theoretical requirement that an infinite number of excited-states be accounted for in the analysis. With a large number of excited-state energies and transition dipole moment values in hand (see Chapter 8 for methods used to obtain the values), the sum-over-states polarizability of the ground state was first used to try to gauge how complete that set of values was. Then, a method was tested to search for a point of convergence in the internal electric field calculated in *Stark06,* with the number of excited states included. When the difference between sum-over-states polarizability values between the ground and first-excited state was plotted against the number of states included, that convergence point was found. At that point, it was concluded that a finite set of excited states could be used for QM Stark analysis.

Then, the protocol was refined to pull out only the most influential electronic states within the set of values required for convergence. Based on perturbation theory, equations for single-state (direct) coupling to the transition of interest as well as double- and triple-state (both indirect) coupling were derived. For PPIX–myoglobin, it was determined that only directly coupled excited states were necessary for a thorough analysis and that only 35 states were deemed essential. In addition, it was also shown that only the in-plane components of the internal electric field for porphin converged while the *z*-component of the field could not be

adequately sensed by a planar porphin model. As a result, a domed porphin model was also explored to increase the sensitivity of the probe to that component but no firm conclusions were reached.

Finally, two theoretical pieces of evidence that a finite set of excited states could be used with this method were presented. First, the quantum-mechanical definition of oscillator strength made it clear that as transition frequency increased, the maximum transition dipole moment value had to decrease. This put to rest fears of extremely high-lying states with very large transition dipole moment values that could not be neglected and allowed for an error calculation of the contribution that might be lost if high-lying states were not included in analysis. Then, an examination of the Thomas–Reiche–Kuhn f-sum rule provided a separate reassurance that a finite set of input was entirely to be expected.

7.7 References

1.   Jentzen, W.; Ma, J.-G.; Shelnutt, J. A., Conservation of the Conformation of the Porphyrin Macrocycle in Hemoproteins. *Biophysical Journal* **1998,** *74* (2), 753-763.

2.   (a) Geissinger, P., Woehl, Jorg C., Prince, Barry J., A quantum-mechanical model for the determination of internal electric fields at protein active sites from the Stark effect on persistent spectral holes. *Journal of Luminescence* **2004,** *107*, 220-229; (b) Geissinger, P.; DuPrey, M. R.; Schwabacher, I. J.; Woehl, J. C.; Prince, B. J., Evaluation and Properties of a Model for the Determination of Internal Electric Fields in Proteins from the Stark Effect of Spectral Holes. *Optics & Spectroscopy* **2005,** *98* (5), 669-674.

3.   Geissinger, P., Kohler, Bryan E., Woehl, Jorg C., Electric Field and Structure in the Myoglobin Heme Pocket. *J. Phys. Chem.* **1995,** *99*, 16527-16529.

4.   (a) Shelnutt, J. A.; Medforth, C. J.; Berber, M. D.; Barkigia, K. M.; Smith, K. M., Relationships between structural parameters and Raman frequencies for some planar and nonplanar nickel(II) porphyrins. *Journal of the American Chemical Society* **1991,** *113* (11), 4077-4087; (b) Jentzen, W.; Song, X.-Z.; Shelnutt, J. A., Structural Characterization of Synthetic and Protein-Bound Porphyrins in Terms of the Lowest-Frequency Normal Coordinates of the Macrocycle. *The Journal of Physical Chemistry B* **1997,** *101* (9), 1684-1699; (c) A. Shelnutt, J.; Song, X.-Z.; Ma, J.-G.; Jia, S.-L.; Jentzen, W.; J. Medforth, C., Nonplanar porphyrins and their significance in proteins. *Chemical Society Reviews* **1998,** *27* (1), 31-42.

5.   (a) Rohatgi-Mukherjee, K. K., Theoretical Absorption Intensity. In *Fundamentals of Photochemistry*, 2nd ed.; New Age International: New Delhi, 1978; pp 63-65; (b) Valeur, B., *Molecular Fluorescence: Principles & Applications*. Wiley: Weinheim, Germany, 2002.

6.   (a) Bethe, H. A.; Jackiw, R. W., Sum Rules. In *Intermediate Quantum Mechanics*, 2nd ed.; W. A. Benjamin, Inc.: New York, 1968; pp 209-215; (b) Thomas, W., Über die Zahl der Dispersionselektronen, die einem stationären Zustande zugeordnet sind. (Vorläufige Mitteilung). *Naturwissenschaften* **1925,** *13* (28), 627-627; (c) Kuhn, W., Über die Gesamtstärke der von einem Zustande ausgehenden Absorptionslinien. *Zeitschrift für Physik* **1925,** *33* (1), 408-412.

Chapter 8

Ab initio Calculations of Ground- and
Excited-State Properties in Porphyrins

8.1 Introduction

Computational resources to calculate excited-state energies and properties of porphyrins were available with the *Gaussian09, Rev. A.02 (G09)* computational package[1] and abundantly enough with a 144-core, 12-node computational cluster (the Department of Chemistry & Biochemistry's "Cleve") to produce 252 electronic excited states of the relatively large 38-atom porphin system. Details of the methods to produce these calculations, configuration interaction – singles (CIS) and time-dependent density functional theory (TDDFT) are provided here and a comparison of the capabilities of the methods is also made. Then, a look at the SOS polarizability of ground and excited states to gauge accuracy of excited-state results in porphin is made along with the higher-order SOS hyperpolarizability. The impact of accuracy on QM internal electric field calculations with the *Stark06* program is examined. Finally, results for a number of structural perturbations to the porphin ring are presented.

The computational resources were also engaged to calculate ground-state electronic properties of porphyrin with interest in modeling the internal electric field produced by myoglobin. These calculations come full circle from Chapter 3, where approaches to modeling the internal electric field were first discussed. The influence of medium-range electrostatics on electron distribution is illustrated with qualitative conclusions about the sensitivity of the porphin probe to its environment. Finally, inquiries about the effect high-energy transitions, such as the

ones discussed in Chapter 7, on the integrity of the porphin structure are entertained with calculations of ionization and dissociation energy.

8.2 Procedure for performing *ab initio* calculations

The process for calculating any molecular property, whether energy or electronic property, whether ground or excited state, starts with a geometry optimization. Then, the position of that structure on the molecule's potential energy surface, e.g., potential well, saddle point, etc., is checked with a frequency calculation. Finally, the wavefunction of the ground or excited state can be calculated based on this optimized structure, which makes available a plethora of energies and electronic properties. In addition, *G09* has the capability to alter the environment of the system of interest from virtual vacuum (no environmental constraints) to a number of solvents to applied electrostatics in the form of a field or point charges. Various keywords and examples of route sections applicable to this work is given in Appendix A.

8.2.1   Geometry optimization

All *ab initio* calculations require a starting structure of nuclear coordinates for the molecular system of interest. However, the structure must undergo refinement to arrive at the conformation that produces a wavefunction of minimum energy in order to carry out a single-point energy calculation. This minimum may be either local or global potential wells, corresponding to different equilibrium conformations of the molecule. The other option for a minimum is a saddle-point on the potential energy surface of the molecule, which corresponds not to an equilibrium conformation but to a transition state. These two different types stationary points on the potential energy surface can be differentiated with a calculation of vibrational

frequencies: all real frequencies indicates that the structure is at a potential minimum while one imaginary frequency indicates a saddle point. Geometry optimizations are iterative and proceed by making incremental changes to the given nuclear coordinates and checking the energy of the molecule at that point to find a minimum. This is rather like slowly crawling around a hilly landscape and getting trapped in a valley. Thus, the optimization is most likely to find whatever minimum, global or not, is closest to the structure provided by the user at the outset. An idea of experimental structures can greatly assist in the process of arriving at a reasonable molecular geometry. For porphyrins, it is known that a planar configuration for species with relatively small substituents is lowest energy in solution. However, results below show that for even a relatively well-known structure like porphin, this optimization step is quite influential on results.

8.2.2 Ground-state calculations

Once a molecular geometry has been optimized, an energy calculation on the ground state of that molecule can be performed along with a number of electronic properties such as dipole moment, partial atomic charge, polarizability, etc. Two methods were used in this work: 1) Hartree-Fock (HF) and 2) density functional theory (DFT). The approaches begin with unique ways of modeling electron distribution and, as a result, can produce distinct results. A simple description of these ground-state methods is given here.

8.2.2.1 Hartree-Fock theory (HF)[3]

This approach begins with the Schrödinger equation, which relies on wavefunctions, ψ, to describe electron distribution in a molecular system. The square of the wavefunction, $|\psi|^2$, gives the electrons' probability density within the molecule. The time-independent version is

used as a basis for HF and can be expressed in terms of the Hamiltonian operator, $\widehat{H}$, to yield the system's energy:

$$\widehat{H}\psi(\vec{r}) = \varepsilon\psi(\vec{r}) \tag{8.1}$$

where ε is the energy of the system, which is now only dependent on the position of the system's electrons and nuclei. The Hamiltonian operator consists of the kinetic ($\widehat{T}$) and potential ($\widehat{V}$) energy of the electrons and nuclei.

$$\widehat{H} = \widehat{T}_{elec}(\vec{r}) + \widehat{T}_{nucl}(\vec{R}) + \widehat{V}_{elec-nucl}(\vec{R},\vec{r}) + \widehat{V}_{elec}(\vec{r}) + \widehat{V}_{nucl}(\vec{R}) \tag{8.2}$$

This expression accounts for the kinetic energy of particles ($\widehat{T}$) with massive size differences as well as both attractive potential ($\widehat{V}_{elec-nucl}(\vec{R},\vec{r})$ between electrons and nuclei) and repulsive potential (between particles of the same type, $\widehat{V}_{elec}(\vec{r})$ and $\widehat{V}_{nucl}(\vec{R})$). The nuclear part of the problem is more straightforward than the electronic part, so Hartree-Fock theory is solely concerned with the solution to the electronic part of Eq. 8.2. Fortunately this is accomplished with the Born–Oppenheimer approximation (i.e., the nuclear motion can be neglected relative to the very fast-moving electrons). Thus the energy contributions for the two types of particles are separable.

The electronic wavefunction is described by molecular-orbital (MO) theory as a combination of a set of molecular orbitals, ϕ_i, that are normalized and orthogonal. In the orbital representation, the Hartree product combines molecular orbitals as:

$$\psi(\vec{r}) = \phi_1(\vec{r}_1)\phi_2(\vec{r}_2) \dots \phi_n(\vec{r}_n) \tag{8.3}$$

A *basis set*, then, expresses MOs as linear combinations of single-electron, nuclei-centered functions:

$$\phi_i = \sum_\mu c_{\mu i} \chi_\mu = \sum_\mu c_{\mu i} \left(\sum_p d_{\mu p} g_p \right) \tag{8.4}$$

where c is the molecular orbital expansion coefficient and χ is some basis function, which is usually modeled as a Gaussian function, g, and d is a weight for the basis function. Different forms of Gaussian functions, for example, are utilized for the uniquely shaped *s,p,* and *d* orbitals. The size of the combination of g functions can influence the accuracy of calculations that uses it and is specified as a pre-packaged basis set at the start of any *ab initio* calculation.

Unfortunately, Eq. 8.3 does not meet all of the criteria required to locate an electron within a molecular system. First, as fermions, electrons must be described by an antisymmetric function, i.e., if any two electrons are interchanged, their wavefunction must change sign. In addition, electron spin must be accounted for in double-occupancy molecular orbitals. To meet these criteria, a wavefunction in HF is expressed as a determinant, not simply a product of the MOs but rows and columns with every combination of position and spin. This leads to the physical description of a location probability, not a set location.

Now with a way to solve for molecular orbitals, φ_i, by way of Eq. 8.4 by specifying a basis set and then a way to incorporate the molecular orbitals into a wavefunction that follows all of the rules for electrons by way of a determinant, a successful Hartree–Fock method then needs to solve for the molecular orbital expansion coefficient, c. The method relies on the variational principle to iteratively solve for c, which means it seeks the coefficients that converge at the

lowest energy for the given wavefunction. A formulation for values of *c* is provided by the Roothaan–Hall equations, of which a matrix form is:

$$Fc = Sc\varepsilon \tag{8.5}$$

With interest in the physical factors considered, we will not define the variables with equations. However, *F*, the Fock matrix, expresses the influence of all of the electrons in the system on a particular orbital. It relies on a density matrix for occupied orbitals. Then, *S* is the overlap matrix describing overlap between orbitals. The orbital energy matrix, ε, contains single-electron orbital energies comprising the basis function, χ, for a molecular orbital, ϕ. The iterative procedure to search for *c* is called the Self-Consistent Field (SCF), which produces a set of occupied and unoccupied orbitals (utilized by the CIS method described below).

The Hartree–Fock method is then limited in accuracy by the chosen basis set, the effectiveness of the algorithm that iteratively searches for molecular orbital coefficients, *c*, and also by the incomplete treatment of electron correlation. The problem of electron correlation, i.e., instead experiencing an average field, electrons "see" and avoid other individual electrons, can be addressed by a configuration interaction (CI) procedure, which will be explained in the CIS section as well.

8.2.2.2 Density functional theory (DFT)[3a, 4]

Density functional theory, on the other hand, does not deal with wavefunctions. Rather, DFT energies involve functions of the electron density, $\rho(\vec{r})$, called *functionals*. Focusing again on the electronic part of the problem, the electronic energy contributions in typical DFT

functionals according to Kohn & Sham have some familiar pieces in kinetic (T) and potential (V)

energy:

$$\varepsilon_{elec} = T + V + J + \varepsilon_{XC} \tag{8.6}$$

The other terms are J, the Coulomb self-interaction, and ε_{XC}, the exchange-correlation energy,

which are not stated explicitly in HF but play a role when CI is included (see CIS section below).

The term J is described as "Coulomb self-interaction" rather than electron–electron repulsion

according to Hohenberg–Kohn theorem (the original work proposing the use of functionals)

because an external potential is used in its place.

Different DFT functionals can have strengths and weaknesses in describing various parts

of the sum in Eq. 8.6, especially the exchange–correlation part. However, the method is certainly

an improvement over HF-quality calculations as this pivotal contribution to the energy is not well

addressed in an HF approach. In a DFT calculation, a functional is specified for the calculation,

but a basis set is also required in order to calculate the kinetic energy terms, which cannot be

described in terms of electron density. DFT calculations in *G09* also use a specified Gaussian basis

set for a procedure called "density fitting" for increased calculation efficiency.[5]

8.2.3   Excited-state calculations

The two excited-state methods explored in this work take two distinct approaches to

modeling electronic transitions. CIS is an older method[6] and TDDFT is a more recent

development.[7] CIS focuses on the "physics of the state," i.e., the excited state itself, while TDDFT

focuses on the "physics of the transition," i.e., the excitation from ground to excited state. As

implemented in *G09*, however, both methods are capable of excited-state structure optimization and therefore can calculate electronic properties of excited states.

A distinction in the type of excited-state transitions must be made for this work. The transitions observed in the spectroscopic experiments are *vertical* transitions with reference to Franck-Condon principle, meaning that transitions originating from the lowest vibrational state of the electronic ground state will often reach the electronic excited state at a higher vibrational level in the instantaneous timeframe of such a transition. The relaxation process then is slow enough to allow a non-radiative transition to the lowest vibrational level of the excited state before a final radiative transition (i.e., luminescent emission) back to the electronic ground state. This cycle can be viewed in terms of molecular geometry equilibria: a transition from an equilibrium ground state to a non-equilibrium excited state followed by relaxation to an equilibrium position of the excited state and termination upon transition back to an equilibrium ground state.

This has bearing on how excited-state calculations are carried out to yield the vertical transitions required by QM Stark analysis. A typical calculation in which an optimized molecular structure is used to request excited-state energies and transition dipole moments are the transitions originating solely from the electronic ground state, $n = 0$, of the molecule. The returned values are for vertical transitions. However, obtaining vertical transition values for excited-state-to-excited-state transitions are not automatically returned. Based on the conclusions in Chapter 7, transitions originating from the first excited state ($n = 1$) are necessary to include for QM Stark analysis of the PPIX–myoglobin system, and transitions from other excited states might be necessary for the analysis of other molecular systems. The energy of a

vertical transition from one excited state, p, to another, q, is a straightforward calculation from ground-state values:

$$\varepsilon_{0q} - \varepsilon_{0p} = \varepsilon_{pq} \tag{8.7}$$

i.e., the difference in energy between the two transitions from the ground state. However, the transition dipole moment between the two excited states is not able to be calculated based on their transitions from the ground state. The method typically employed by *G09* to calculate properties for excited states involves the optimization of that state, followed by another excited state calculation to determine energies and transition dipole moment values. However, for our application, that would amount to transition dipole moment values originating from an excited state at *equilibrium* geometry, which are not "vertical." Fortunately, an internal option ("IOps") for the calculation of the desired vertical transition dipole moment values from non-equilibrium excited states was available in *G09 Rev. A.02*, but only for the CIS method (see Appendix A). Although it will be seen that TDDFT had some advantage in accuracy over CIS, this availability ultimately restricted the values that could be used for subsequent use in QM Stark analysis to those generated by CIS.

8.2.3.1 Configuration interaction – singles (CIS)[3a, 4a, 6]

As mentioned above, including electron correlation with a configuration interaction (CI) matrix in an HF calculation improves the method's results. Understanding this CI process, which is intended to better describe the *ground state*, can set the stage for understanding the related *excited-state* method, CIS. The CI process involves promoting one or more electrons in an HF determinant from occupied to unoccupied orbitals, which are generated as part of the SCF

process in an HF calculation. As substitutions of unoccupied orbitals enter the original HF determinant, a new determinant is formed and a collection of determinants is created as more substitutions are made. If all possible substitutions are made, "full CI" is reached as the initial HF determinant is combined with substituted ones:

$$\psi = c_0 \psi_{HF} + \sum_i^{occupied} \sum_a^{virtual} c_i^a \psi_i^a + \sum_{i<j}^{occupied} \sum_{a<b}^{virtual} c_{ij}^{ab} \psi_{ij}^{ab} + \cdots \qquad (8.8)$$

This description sounds a lot like mimicking electronic transitions, which the CI process is essentially doing. However, for the CIS process, the ground-state HF determinant is used as a reference determinant in order to generate excited-state wavefunctions:

$$\psi_{CIS} = \sum_i^{occupied} \sum_a^{virtual} c_i^a \psi_i^a \qquad (8.9)$$

The method uses only singly-excited wavefunctions in a linear combination, as stated in Eq. 8.9, followed by diagonalization.

The excited-state energies generated by the method are eigenvalues of:

$$\langle \psi_{ia} | H | \psi_{jb} \rangle = [\varepsilon_{HF} + \varepsilon_a - \varepsilon_i] - (ja\|ib) \qquad (8.10)$$

where ε_{HF} is the HF ground-state energy, ε_a the energy of unoccupied orbital a, ε_i the energy of occupied orbital i, and the final term is the anti-symmetrized two-electron integral. In addition, an excited-state potential energy surface is constructed with the calculation because a wavefunction is generated, making an excited-state optimization possible. Historically, the optimization was the strength of the method, with quantitative values like state energies less reliable and are considered "roughly HF quality."[4a]

8.2.3.2 Time-dependent density functional theory (TDDFT)

Rather than calculating the excited state itself like CIS, TDDFT takes the approach of applying a perturbation to the ground state to model a transition incurred by the application of electromagnetic radiation. When small perturbations applied, the time-dependence can be modeled with some approximations. Runge & Gross developed TDDFT by treating the external potential portion of the DFT equation in a time-dependent fashion. The way that most computational programs go about modeling perturbations and to calculate excited state energies is to search for the poles of a linear response function, $[f(\vec{r}, t)]$. This is the point where the value of the function increases very quickly. The model must be that of a true small perturbation, thus the excited state density is similar to that of the ground state. The linear density–density response function quantifies how the ground state responds to a perturbation of the external potential as part of the exchange-correlation kernel:

$$f_{xc}[n_{GS}]\left(\vec{r}, \vec{r}\,', t - t\,'\right) = \frac{\delta v_{xc}(\vec{r}, t)}{\delta n(\vec{r}\,', t\,')} \tag{8.11}$$

where n_{GS} is the ground-state density and v_{xc} is the exchange-correlation potential. This describes how the density at $\vec{r}$ and t changes if it is perturbed at $\vec{r}\,'$ and t'.

This simplistic description of the method nevertheless demonstrates that TDDFT is a distinct approach to calculating excited states. The results can be expected to also be distinct from the CIS results, which will be demonstrated in this work. Accuracy of TDDFT calculations depends heavily on the functional used like DFT, again largely due to the exchange-correlation formulation of the functional.

8.3 Results of excited-state calculations

For the following calculations, the process of geometry optimization followed by a check of stationary point type (minimum or saddle point) with a frequency calculation was upheld before proceeding to any excited-state calculation. For all TDDFT calculations, the porphin structure was optimized with DFT using the same functional and basis set combination used for the excited-state calculation. For all CIS calculations, the porphin structure was optimized with HF using the same basis set used for the excited-state calculation. Differences in the structure of porphin arose between the two optimization methods, which will be discussed after the excited-state calibration process.

8.3.1 Calibration process

The ongoing challenge for *ab initio* calculations is the selection of an appropriate basis set and, for DFT calculations, a functional. For well-characterized systems like porphin, for which experimental transition values are readily available, a "calibration" process is best for choosing a basis set and functional, i.e., calculating known values with a number of such parameters and comparing the accuracy of results to the published experimental values. For TDDFT, functionals can be categorized into families of related formulations that can be expected to return similar results, so only choosing a functional from each of families can assist with the initial calibration process. Then the functionals from within the family that best matches known values can be further assessed for accuracy. For free-base porphin, the four visible transitions, Q_x, Q_y, B_x, B_y, serve as the values for comparison in a functional / basis set calibration.

8.3.1.1 Previously published results for porphin

Accepted experimental values for the transition energy and oscillator strengths of the four visible transitions of gas-phase porphin are given in Table 8.1.[8] These are the values used for comparison to calculated results in the calibration process.

Table 8.1. Accepted experimental values for the transition energies and oscillator strengths of the four visible transitions of gas-phase porphin.

Transition	Transition Energy (various units)				Oscillator Strength	Transition Dipole Moment
Units	eV	nm	cm^{-1}	Hartree (a.u.)	No units	a.u.
1 (Q_x)	1.98	626	15970	0.0728	0.02	0.642
2 (Q_y)	2.42	512	19519	0.0890	0.07	1.09
3 (B_x)	3.33	372	26858	0.1225	1.15	3.75
4 (B_y)	3.33	372	26858	0.1225	1.15	3.75

Because a number of computational studies have used porphin as a highly symmetric test molecule, there are many results available for comparison.[2] Fig. 8.1 plots the Q_x transition energy calculated by various computational methods and suggests TDDFT as a high accuracy method for the molecule.
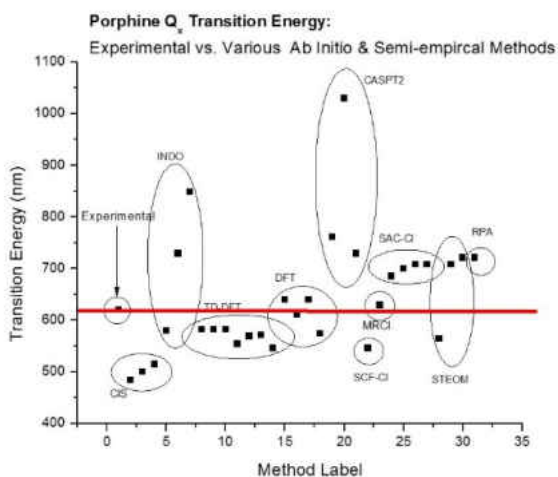
Figure 8.1. Thirty values of the Q_x transition energy of porphin as calculated by a number of methods and reports collected in Ref. [2] as compared to the experimental value that sits on the red line. The methods of interest, CIS and TDDFT, sit slightly below the experimental value (higher energy) with TDDFT more accurate than CIS.

8.3.1.1.1   Results of calibrations

8.3.1.1.1.1 TDDFT calibration

First, the TDDFT method was calibrated. The seven functional families (groups of functionals that have a common formulation) that were examined initially with the functional representative used in the calibration are given here:

1) LSDA (local spin-density approximation) family: LSDA functional

2) GGA (generalized gradient approximation) family: BVP86 functional

3) M-GGA (meta-generalized gradient approximation) family: TPSS functional

4) H-GGA (hybrid-generalized gradient approximation) family: B3LYP functional

5) HM-GGA (hybrid–meta-generalized gradient approximation) family: M06 and M06-2X functionals

6) long-range correction family: CAM-B3LYP functional

The first three families are considered "pure functionals," i.e., HF exchange is not included explicitly. The two hybrid functional families begin to incorporate HF exchange. The HM-GGA family encompassed two functionals that had very different amounts of HF exchange, M06 with a small percentage and M06-2X with a much larger amount, so both were included. The final long-range correction functional extends the "tail" of the functional. Transition energy results are plotted against experimental transition energies in Fig. 8.2. The H-GGA functional family was chosen as the most accurate in this calibration.
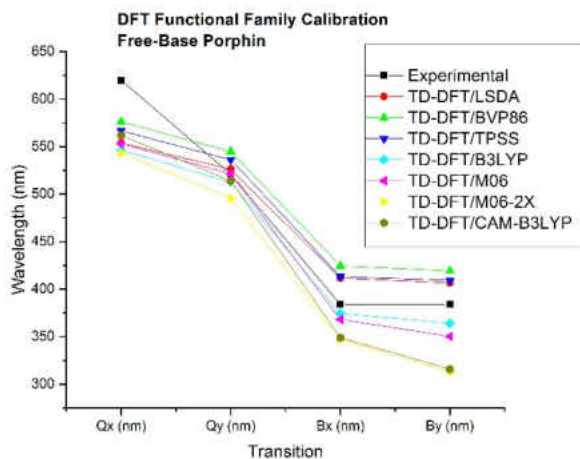
Figure 8.2. TDDFT functional family calibration plots transition energies returned by various functionals versus experimental values. The H-GGA family, represented by B3LYP, returned the most accurate values.

Next, five available functionals in the H-GGA functional family were calibrated in the same manner. The H-GGA functionals examined were:

1) B3LYP

2) O3LYP

3) B3P86

4) B3VP86

5) PBE1PBE

Results are plotted in Fig. 8.3 similarly to the functional family calibration. The O3LYP functional was chosen as the most accurate in this calibration as compared to experimental transition energy values, especially for the Q_y, B_x, and B_y transitions. In general, the Q_x transition energy is the least accurately reproduced energy, with most TDDFT functionals predicting too high an energy for the transition.

222
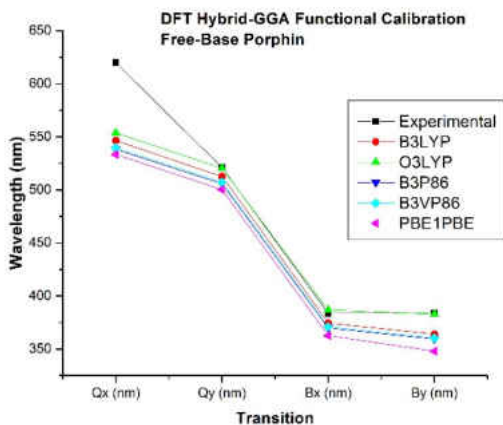
Figure 8.3. TDDFT H-GGA functional calibration plots transition energies returned by various functionals versus experimental values. The O3LYP functional appears to be most accurate.

The first two TDDFT calibration steps above were performed using the 6-311++G** (also sometimes labelled 6-311++G(d,p)) basis set as a reasonably large basis set. However, as the size of basis set does not guarantee accuracy, a final basis-set calibration using the O3LYP functional was performed. The twenty basis sets examined were:

1) STO-3G

2) 3-21G

3) 6-31G

4) 6-31G*

5) 6-31+G*

6) 6-31+G**

7) 6-31++G**

8) 6-311G

9) 6-311G*

10) 6-311+G*

11) 6-311+G**

12) 6-311++G**

13) cc-pVDZ

14) cc-pVTZ

15) cc-pVQZ

16) LanL2DZ

17) SDD

18) DGDZVP

19) DGDZVP2

20) DGTZVP

Results are plotted in Fig. 8.4 as above. In this case, the choice of basis set was more difficult as the functional performed reasonably well with a number of basis sets. A combination of accuracy and computational efficiency was found with the cc-pVDZ basis set. The related cc-pVTZ and cc-pVQZ technically returned slightly more accurate values, however the computational time increased exponentially with the larger basis sets and their use would have made calculation of a large number of excited state (as was deemed necessary in Chapter 7) rather impractical.
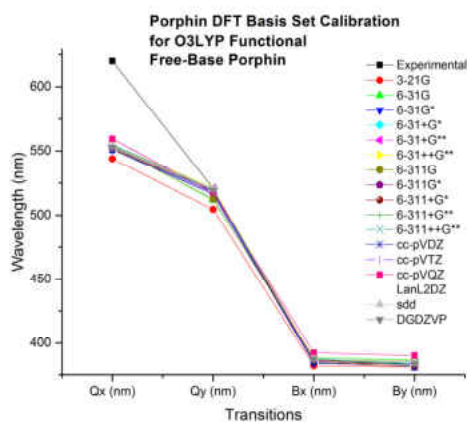
Figure 8.4. The TDDFT basis-set calibration plots transition energies returned by various basis sets used with the O3LYP functional versus experimental values. The cc-pVDZ basis set was determined as the most appropriate basis set to use in combination with the O3LYP functional.

8.3.1.1.1.1.1 **CIS calibration**

Then, a similar process with the CIS method was undertaken to calibrate the method and choose a basis set. [Note that the state-energy axis has changed to state-energies in electron-volts (eV) from wavelength in nm in the TDDFT calibration section.] The results for the transition energies are plotted versus experimental values in Fig. 8.5. The same twenty basis sets were used for CIS as with TDDFT. All of the basis sets significantly overestimated the energy of all four transitions, much more than TDDFT. Most of the basis sets performed similarly within the method, other than two that overestimated the energies even more. Not surprisingly the two most inaccurate basis sets were the smallest, STO-3G, and a related basis set, lanl2mb. A closer look at the middle "pack" reveals the moderately sized sdd and related lanl2dz basis set as the most accurate with sdd returning slightly more accurate results in terms of transition energy.
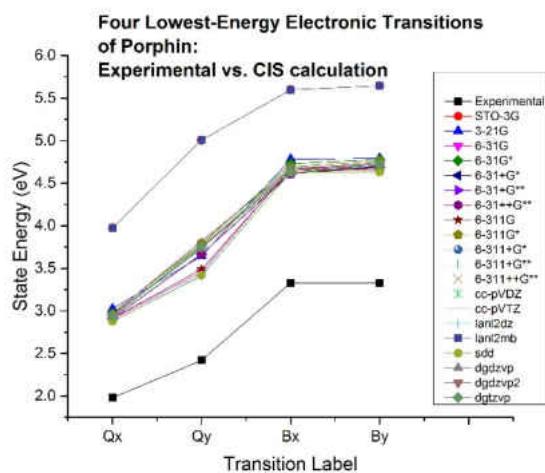
Figure 8.5. The CIS basis-set calibration plots transition energies returned by various basis sets versus experimental values. The sdd and lanl2dz basis sets were determined as the most accurate basis sets in terms of transition energy results.

In addition to a transition energy calibration, the oscillator strengths returned by CIS were also examined here. The results of the four most accurate basis sets in terms of oscillator strength

are plotted in Fig. 8.6. Interestingly, the transition energy accuracy is not a predictor of oscillator strength accuracy as both STO-3G and lanl2mb generate the most accurate oscillator strength values. It was determined that sdd returned the best combination of transition energy and oscillator strength with the CIS method.
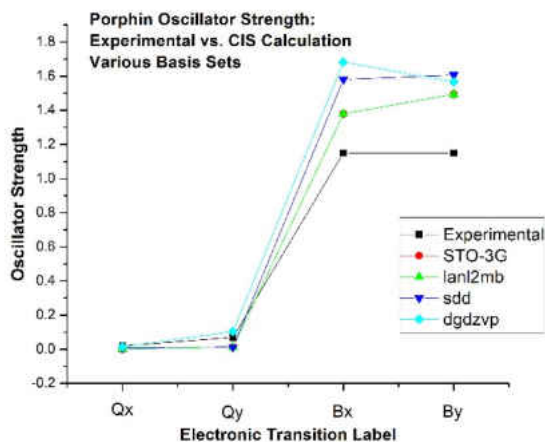
Figure 8.6. The CIS basis-set calibration plots oscillator strengths returned by four basis sets versus experimental values. The STO-3G and lanl2mb basis sets were determined as the most accurate basis set in terms of oscillator strength, which was one of the *least* accurate in terms of transition energy.

8.3.1.1.2 Calibration comparison

The results from the TDDFT and CIS calibrations are compared in Fig. 8.7 (transition energy) and Fig. 8.8 (oscillator strength). It is determined that TDDFT returns more accurate transition energies for free-base porphin but CIS returns more accurate oscillator strengths for the molecule. However, CIS showed positive qualities in both categories where TDDFT did not. Although the absolute values of transition energies were incorrect with TDDFT, the trend given by CIS / sdd values mirrored the experimental values, i.e., had a more consistent error. (See, however, discussion of errors in transition polarization in optimization section below.) More strikingly, TDDFT generated a more serious error with oscillator strength values when it returned a value of *zero* for one of the two strongly allowed Soret (B) transitions.

226
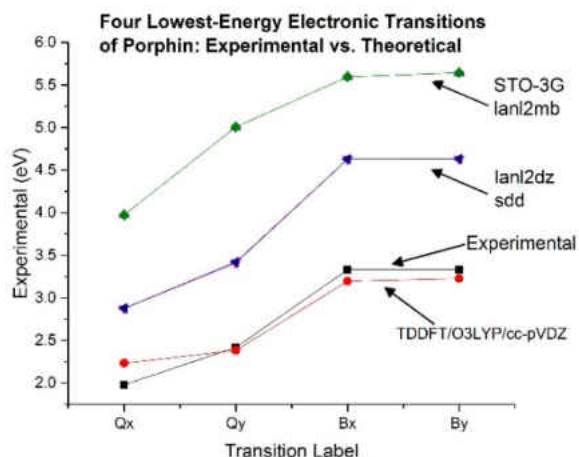
Figure 8.7. Comparison of the visible transition energies of porphin as determined experimentally and calculated with two excited-state methods, CIS and TDDFT. TDDFT returns results closest to experimental data.
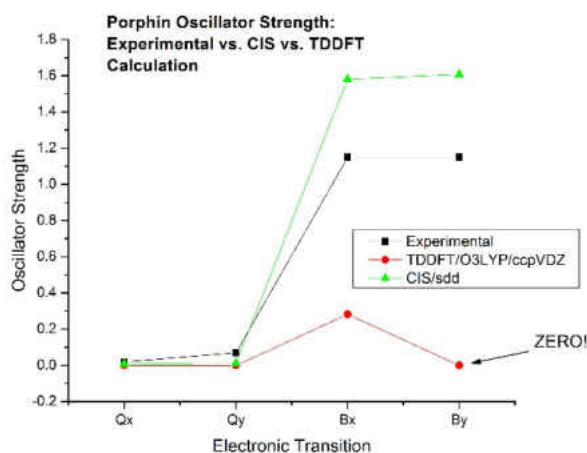
Figure 8.8. Comparison of the visible transition oscillator strengths of free-base porphin as determined experimentally and calculated with two excited-state methods, CIS and TDDFT. CIS returns results closest to experimental.

In addition, Table 8.2 gives tabulated values for experimental, TDDFT, CIS, and ZINDO transition energies and oscillator strengths. ZINDO is a semi-empirical method that had been used in earlier reports related to this work[9] and is included for comparison. The impact of these various methods' excited-state results on internal electric field calculations is discussed below.

227

Table 8.2. Values, both experimental and calculated, related to the four visible-region transitions of free-base porphin.

Transition from ground state	Transition Energy (eV)				Oscillator Strength			
	Experi-mental	TDDFT/ O3LYP/ cc-pVDZ	CIS/ sdd	ZINDO	Experi-mental	TDDFT/ O3LYP/ cc-pVDZ	CIS/ sdd	ZINDO
1	1.98	2.23	2.88	1.71	0.02	3.00×10^{-4}	0.01	0.02
2	2.42	2.38	3.42	2.05	0.07	3.00×10^{-4}	0.01	0.03
3	3.33	3.20	4.63	3.39	1.15	0.28	1.58	1.64
4	3.33	3.23	4.63	3.55	1.15	0.00	1.61	2.49

8.3.1.1.3   Optimized porphin structures

8.3.1.1.4   Planar structures

The optimization process performed prior to the excited-state calculations described above have bearing on the results, with each method producing small but significant differences. (The structure files for each are given as Cartesian coordinates in Appendix J.) Each for each method, HF (for use with CIS) and DFT (for use with TDDFT), a structure of identical symmetry was provided for the optimization process: a planar D_{2h} structure with inner hydrogens in the *trans* position and oriented as illustrated by Fig. 7.2. Both methods retained the planarity of the macrocycle. However, while DFT retained the inversion symmetry of the initial structure at the end of optimization, HF generated a structure with two pyrrole nitrogens slightly displaced by 0.014485 Å along the negative x-axis. This appears as pyrrole rings I and IV rotated slightly clockwise and counterclockwise, respectively, although the deviation is so slight that it is only visual upon careful inspection. The non-symmetric structure can be traced to the optimization process. When a totally symmetric file was loaded to optimize with the HF method, the stationary

point at which the job termination was found to be not a minimum but a saddle point, i.e., the subsequent frequency job produced one imaginary frequency. A tactic to move the structure away from the saddle point is to animate the imaginary frequency and displace the coordinates along that frequency's normal mode axis. Not surprisingly the displacement used to move the structure away from the saddle point remained as a slight asymmetry when the structure found a nearby minimum.

Interestingly, this slightly asymmetric structure produced excited-state results that were unexpected in two aspects. First, the HF-optimized structure predicted transition polarizations with the CIS method that were opposite to the experimental and TDDFT-method results. The CIS jobs calculated the first and third transitions (Q_x and B_x, respectively) as *y*-polarized and the second and fourth transitions (Q_y and B_y, respectively) as *x*-polarized. Second, even though the asymmetry was significant enough to apparently switch transition polarizations, it was not significant enough to mix polarizations for individual transitions, i.e., each transition remained exclusively polarized along a single axis (*x, y,* or *z*).

These results prompted a closer look to determine whether this was an anomaly of method /basis set combination. The switched polarizations were in fact consistent across the CIS-method results with the asymmetry, regardless of basis set. In addition, when the symmetric structure that technically landed on a saddle point of the potential energy surface was fed into a CIS calculation, the polarizations returned to the experimental positions. However, this tactic cannot be taken as correct for a ground-state structure described by HF as it does not reside in a potential-energy minimum. Another tool for determining the accuracy of an optimized structure

is through a wavefunction stability test (keyword: stable for a *G09* energy job). An HF wavefunction calculated for a totally symmetric porphin structure returned as unstable.

An additional set of calculations looked at whether the porphin structure with slightly deviated pyrrole nitrogens was the only asymmetric structure to find a minimum. Other normal-mode deviations were explored to produce other asymmetric porphin structures, including elongating the structure along the inner hydrogens and rotating both pyrrole rings clockwise. In each case, the structure returned to the symmetric saddle point rather than finding a new minimum. It was determined that HF did not consider the totally symmetric structure to be the lowest-energy ground-state structure. One plausible explanation might be that the HF method could be assessing the repulsive interaction of inner hydrogens as more unfavorable than the DFT method, enough so that the generally favorable symmetry advantages do not overcome this unfavorable contribution.

8.3.1.1.5 Non-planar structures

The domed porphin model explored in Chapter 7 to try to increase the intensity of *z*-polarized transitions (i.e. perpendicular to the porphin plane) required some constraints to successfully optimize such a structure, especially since structure optimizations allowed to optimize freely found planar configurations, regardless of method. The tactic used to accomplish this was to freeze certain coordinates to guarantee a *z*-displacement, while retaining in-plane *x*–*y* symmetry. This was best accomplished with a symbolic Cartesian input format, where coordinates are labelled as either variable (allow to optimize) or constant (frozen along that axis during optimization). This option had the advantage of being able to fix a displacement in the *z*-direction while optimization could still proceed in the *x*- and *y*-directions,

even for "frozen" atoms. The frozen atoms were chosen as eight peripheral hydrogens bonded to pyrrole rings at $z = 0.000000$ and the four pyrrole nitrogens at $z = 0.400000$ or 0.600000. An example of a non-planar porphin input files in symbolic Cartesian coordinates is also given in Appendix A.

8.3.1.2 Variance of SOS polarizability with basis set

The approach to predicting convergence for internal electric field calculations with difference in SOS polarizability described in Chapter 7 was shown to be effective. However, only results from a single basis set was presented, CIS / sdd. The approach was examined with the twenty basis sets listed earlier to see if a convergence point was generally found for the x- and y-components and none for the z-component. This was found to be the case across all basis sets as calculated with CIS. Figs. 8.9a–d show the x- and y-components for two basis sets, STO-3G and cc-pVDZ, which are distinct in formulation from sdd. Convergence is found well within 50 states.
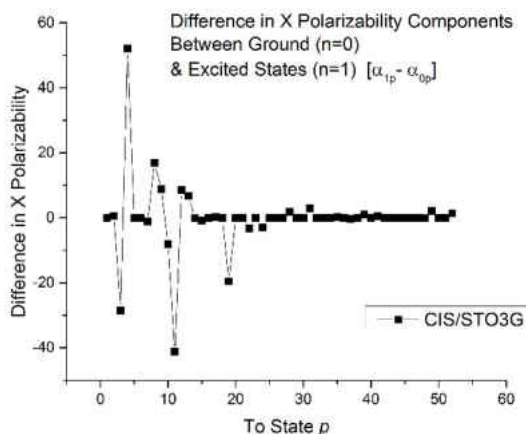
Figure 8.9a. Difference in SOS polarizability for two basis sets calculated with CIS show convergence; x-component, STO-3G.
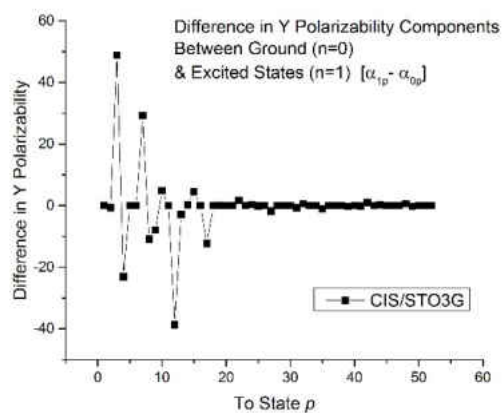
Figure 8.9b. Difference in SOS polarizability for two basis sets calculated with CIS show convergence; *y*-component, STO-3G.
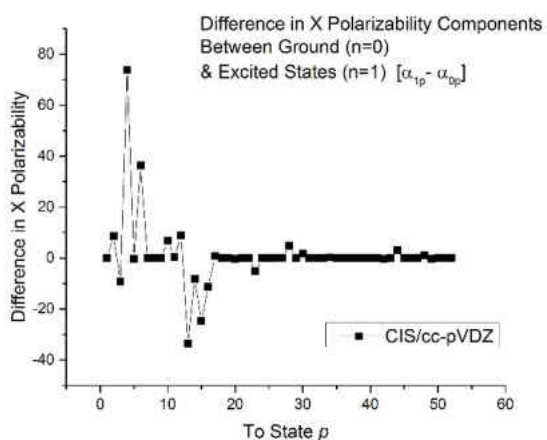
Figure 8.9c. Difference in SOS polarizability for two basis sets calculated with CIS show convergence; *x*-component, cc-pVDZ.
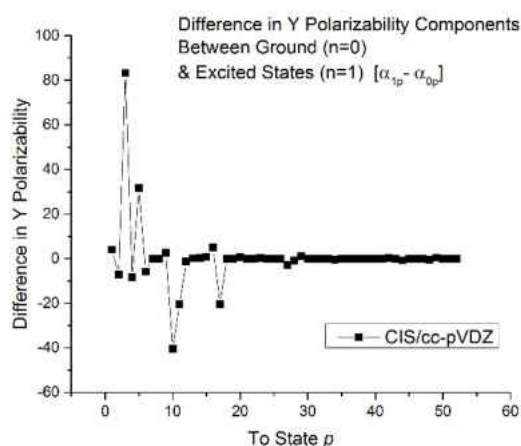
Figure 8.9d. Difference in SOS polarizability for two basis sets calculated with CIS show convergence; *y*-component, cc-pVDZ.

8.3.1.3 SOS hyperpolarizability

As an approach to gauge the accuracy of excited-state calculations, the sum-over-states (SOS) hyperpolarizability, *β*, was calculated:

$$\beta_{m,xyz} = \sum_{p\neq m}\sum_{q\neq m}\frac{\langle m|\hat{\mu}_x|p\rangle\langle p|y\hat{\mu}_y|q\rangle\langle q|\hat{\mu}_z|m\rangle}{(\varepsilon_p-\varepsilon_m)(\varepsilon_q-\varepsilon_m)} \qquad (8.11)$$

where *m* is the electronic state of interest, *p* and *q* are directly and indirectly coupled electronic states, and *x, y,* and *z* spatial components. Hyperpolarizability is a tensor with twenty-seven components: xxx, xxy, xxz, xyx, xyy, xyz, xzx, xzy, xzz, yxx, yxy, yxz, yyx, yyy, yyz, yzx, yzy, yzz, zxx, zxy, zxz, zyx, zyy, zyz, zzx, zzy, zzz. It appears in the classical Stark equation (Eq. 6.1) in the third-order term as *Δβ*. Like the first-order term in the equation containing the difference dipole moment (*Δμ*), this first hyperpolarizability term should also vanish because of porphin's inversion symmetry.

Using the data calculated with CIS / sdd, hyperpolarizability values for 252 excited states of porphin were calculated. Interestingly, twenty out of the twenty-seven components were indeed zero and each of the seven non-zero components occurred with the same *x*-containing polarizations for each and every state of interest. This seems as if it may be related to the slight deviation from symmetry along the *x*-axis. (Contribution to the hyperpolarizability of conjugated molecules in this "sum-over-states" fashion is also explored in [10].)

8.3.1.4 Comparison of methods with *Stark06*

The results from the excited-state calculations were implemented in *Stark06* to calculate internal electric field and to compare to previous results generated with excited-state input from

the semi-empirical method, ZINDO. The previous results had just 17 electronic states to use in

internal electric field calculations and so the 17 lowest-energy, *x*- or *y*-polarized states were used

for each method to generate the results shown in Table 8.3. In addition to the results from ZINDO,

CIS / sdd, and TDDFT / O3LYP / cc-pVDZ, a calculation using a mixture of the more accurate TDDFT

transition energies and the more accurate CIS oscillator-strength values was performed. All of

the results were distinct in terms of direction (see angle column) but fall within about an order

of magnitude, with the *x*-component larger than the *y*-component for all but the mixed CIS /

TDDFT calculation, where the magnitude was about the same for each component.

Table 8.3. Comparison of internal electric field values generated with difference method / basis set combinations.

Method / Basis Set	E_x Final (V / cm)	E_y Final (V / cm)	Angle (degrees)
ZINDO	6.14×10^5	8.68×10^5	55
CIS/sdd	3.44×10^5	8.37×10^5	68
TDDFT / O3LYP/cc-pVDZ	7.34×10^5	1.07×10^6	56
TDDFT / CIS Mix (ε_n / f_{osc})	5.27×10^5	4.09×10^5	38

8.3.2 Electronic properties of excited states

As described in the theoretical basis of CIS, the method was developed to be able to optimize excited-state structures. To examine some electronic properties of the first excited states (i.e., state of interest, $n = 1$), the excited-state structure was optimized with CIS / sdd.

8.3.2.1 Excited-state optimization

Excited-state optimization yielded a structure with total energy compared to the ground-state in Table 8.4. The keywords related to this optimization and the coordinates of the optimized structure are presented in Appendix J.

Table 8.4. Total energy of the ground state and first excited state of free-base porphin as calculated with HF / sdd.

Electronic State	SCF Energy (a.u.)	Zero-Point Energy Correction (a.u.)	Total Energy (a.u.)
$n = 0$ (ground)	-982.9589936	0.31516948212	-982.6438241
$n = 1$ (1st excited state)	-982.9553994	0.31367172504	-982.6417277
$\Delta\varepsilon$ (difference)			0.0020965

As expected, the first-excited state has a higher (less negative) energy than the ground state.

8.3.2.2 Polarizability of excited states

Once a structure was optimized, a frequency calculation yields the exact polarizability of the ground state and first excited state of porphin using HF / sdd. The centrosymmetric molecule, as predicted by theory, generated only the three diagonal components, α_{xx}, α_{yy}, α_{zz}, of the nine polarizability tensor components. Values for exact polarizability of two states is given in Table 8.5. Theoretically, the polarizability of an excited state should be greater than that of the ground state, and that expectation is met with this *ab initio* calculation for all three components, which was not the case for comparable semi-empirical ZINDO calculations.[11]

Table 8.5. Exact polarizability of the ground and first-excited state of porphin as calculated with HF / sdd, given as the three non-zero components.

Electronic State	α_{xx} Tensor Component (a.u.)	α_{yy} Tensor Component (a.u.)	α_{zz} Tensor Component (a.u.)
$n = 0$ (ground)	407.770	391.212	90.526
$n = 1$ (1st excited state)	445.916	433.789	91.288

The optimized first excited state of porphin was also used to calculate transitions from that state using CIS / sdd. The SOS polarizability was calculated for this state and plotted below in Figs. 8.10a–c as more electronic states are included in the calculation similarly to the series of plots in Chapter 7 for the ground state in Figs. 7.4a–c. The first-excited-state SOS polarizability for the in-plane components reaches to not even half the exact polarizability at N = 252 states,

where the ground state reaches to almost 90% for the same number of states included. In addition, the first excited-state plots of the in-plane components also show a much sharper increase before 25 states are included followed by a very slow increase over the next 225 states. The progression of the z-component in Fig. 8.10c is similar to that for the ground state: very slow, more linear progression. Figure 8.11 plots the SOS polarizability of the combined spatial components for the ground and first excited state up to $N = 252$. Even though the total exact polarizability shows that the excited state is more polarizable than the ground state as expected, the proportion contributed by much higher-lying states is far greater for the excited state.

a)

Total Sum-Over-States
First Excited State X Polarizability (α_{xx})
vs. Number of States Included

48.4% of exact polarizability

5.90% of exact polarizability

Sum-Over-States Polarizability (a.u.)

Number of States Included (N)

b)

Total Sum-Over-States
First Excited State Y Polarizability (α_{yy})
vs. Number of States Included

51.0% of exact polarizability

30.8% of exact polarizability

Sum-Over-States Polarizability (a.u.)

Number of States Included (N)

c)

Total Sum-Over-States
First Excited State Z Polarizability (α_{zz})
vs. Number of States Included

1.61% of exact polarizability

0.107% of exact polarizability

Sum-Over-States Polarizability (a.u.)

Number of States Included (N)

Figure 8.10. SOS polarizability of the first excited state of porphin (CIS / sdd) for comparison to Figs. 7.4 for the ground state. **a)** *x*-polarized; **b)** *y*-polarized; **c)** *z*-polarized.

238
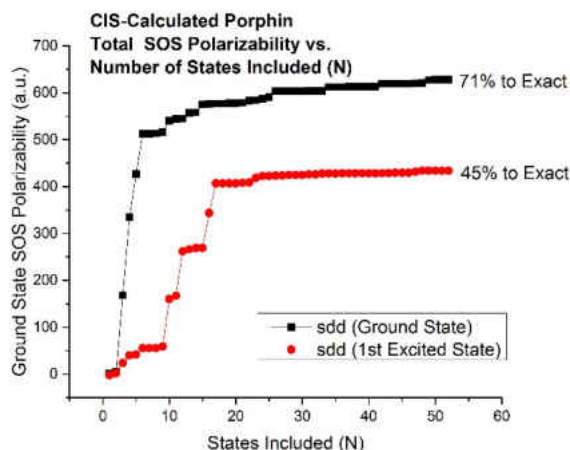
Figure 8.11. SOS polarizability of the ground and first excited state of porphin (CIS / sdd) shows different progress towards the exact values as more states are included in the analysis.

8.3.2.3 Implementation in classical Stark analysis

A motivation for calculating electronic properties of the excited state of porphin was to open a path to perform an analogous classical Stark analysis (Eq. 6.1) with the data for comparison to the QM Stark analysis. While an algorithm for such a comparison calculation remains in progress, the pivotal difference polarizability value is generated with the data in Table 8.5, the definition of difference polarizability ($\Delta\alpha$) from Eq. 6.3, and the relevant Eq. 8.12a–c below. The difference polarizability value produced with CIS / sdd is calculated to be $\Delta \alpha_{01} = 3.990 \; a.u.$

$$\overleftrightarrow{\alpha}_{principal\;axes} = \begin{pmatrix} \alpha_{xx} & 0 & 0 \\ 0 & \alpha_{yy} & 0 \\ 0 & 0 & \alpha_{zz} \end{pmatrix} \tag{8.12a}$$

$$Scalar \; \alpha = \frac{1}{3} Tr(\overleftrightarrow{\alpha}) \tag{8.12b}$$

$$Tr(\alpha) = \sum \alpha_{ii} \tag{8.12c}$$

8.3.3 Excited-state transitions in perturbations to porphin

Beyond the model probe, porphin, the real, experimental systems of interest have unique transition energies. Different small "perturbations" to the highly symmetric theoretical probe will, of course, alter transition energies: metallation with zinc, adding substituents to yield protoporphyrin, and finally subtle conformation changes, especially to ring planarity, in real systems such as the real heme conformation in the 1MBO file of oxygenated human myoglobin. Computational method comparisons were made for several of these perturbations for future reference. [Note that in the figures below the energy axis is again presented as wavelengths in nm.]

8.3.3.1 Metallation

First, free-base porphin was compared to zinc porphin in terms of transition energy to gauge the effect of metallation on the macrocycle and to examine how theoretical methods dealt with the addition of a transition metal, which sometimes poses a problem computationally. Porphin energies are reported as calculated in the calibration section above with gas-phase experimental values reported from the same source. Zinc porphin, however, has comparatively few spectroscopic values reported. Apparently, the molecule is synthesized from free-base porphin in solution, so no gas-phase values could be obtained for comparison. The available sources gave spectra for zinc porphin only in in organic solvents[12], and a low dielectric-constant source was used (n-octane, $\varepsilon \approx 2$) for the comparison below. The results in Figs. 8.12a and 8.12b show that TDDFT / O3LYP / cc-pVDZ makes a very reasonable prediction for both the free-base and metallated versions, even predicting the degenerate nature of both the Q and B transitions

in the metallated porphin. This means that TDDFT as a computational method can be robust with

transition metals (at least when no unpaired electrons are present).
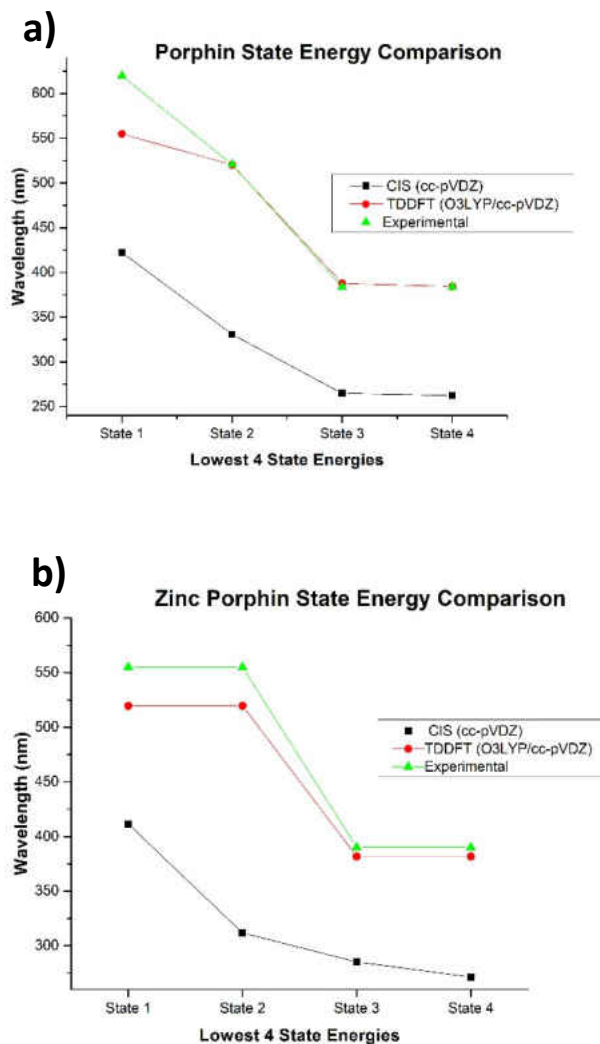
Figure 8.12. The four visible transitions of porphin calculated with two computational methods compared to experimental values. **a)** Free-base porphin; **b)** zinc porphin.

8.3.3.2 Substituents

Next, substituents were added to porphin to create protoporphyrin IX as described in

Chapter 3, i.e., two negatively charged propionate groups in addition to methyl and vinyl groups.

The optimization process was initiated with a fully planar model; however, these optimizations

were quickly unsuccessful, meaning that a minimum on the potential energy surface could not

be located within the default parameters in terms of number of iterations or in terms of step size. The solution arrived with a structural "clean-up" feature in GaussView 5 that uses a quick force-field calculation to arrive at a plausible structure. This step altered the PPIX structure immediately to force this *in vacuo* structure to curl each charged propionate group in opposite directions either above or below the porphin plane to look quite similar to their respective conformation when heme is in the protein environment (Fig. 3.8a). Clearly, the negative charge on the propionates needed distance in the optimized structure. The optimized structure of PPIX is provided in Appendix J.

An unexpected result occurred during calculations of transition energies for any form of PPIX: the O3LYP functional that was so successful for porphin predicted transition energies for these substituted porphyrins well into the infrared (IR) region. This was clearly incorrect, so another calibration was undertaken to see if a functional existed to predict reasonable values for PPIX. An HM-GGA functional with a larger amount of HF exchange, M06-2X made reasonable predictions for PPIX-based structures. It is unclear whether the overall -2 charge or the substituents themselves caused the breakdown in accuracy for the O3LYP functional. The values in Figs. 8.13a-b are compared to PPIX and ZnPPIX in neutral phosphate buffer, so these do not provide exact comparison to the *in vacuo* calculations, but give a reasonable idea of how each method performs. The TDDFT / M06-2X functional gave results similar results to the CIS results. The metallation with zinc gives similar performance. Note the large break in the energy axis.

a)

Protoporphyrin IX (PPIX)
State Energy Comparison

b)

Zinc Protoporphyrin IX (ZnPPIX)
State Energy Comparison
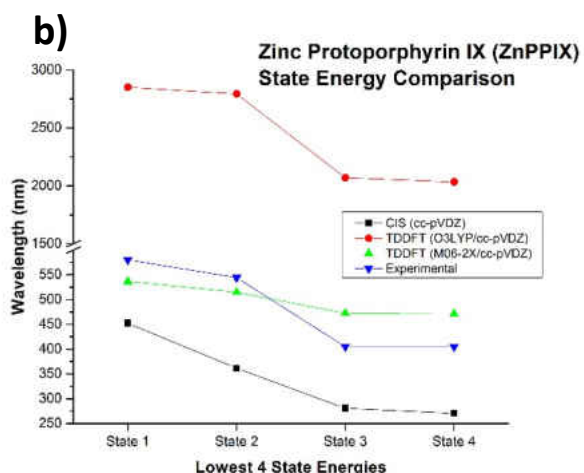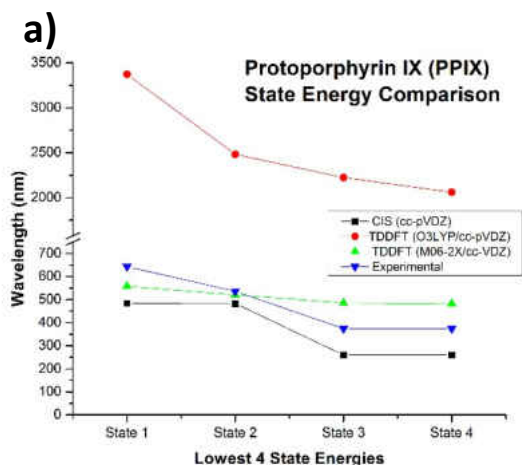
Figure 8.13. The four visible transitions of protoporphyrin IX calculated with two computational methods compared to experimental values. **a)** Free-base PPIX; **b)** ZnPPIX.

8.3.3.3 Planarity of porphyrin ring

Finally, the effect of small deviations to the PPIX structure when substituted in myoglobin were examined. The structure of the heme ring from the 1MBO PDB file was isolated, protonated, and the iron atom was removed. For the 1MBO PPIX structure, two inner hydrogens were added; for the 1MBO ZnPPIX structure, the iron atom was replaced with zinc. The structures visually showed some torsion in the macrocycle. After optimization *in vacuo*, which did not produce a planar ring but retained much of its crystal-structure character, transition energies were calculated as before (see Figs. 8.14a-b). The experimental energies are those found for the PPIX-

and ZnPPIX-myoglobin in neutral phosphate buffer, so are subject to some error as above. A similar pattern of accuracy emerged like the planar PPIX structures, with TDDFT / M06-2X producing the most reasonable energies and certainly better than TDDFT / O3LYP results.
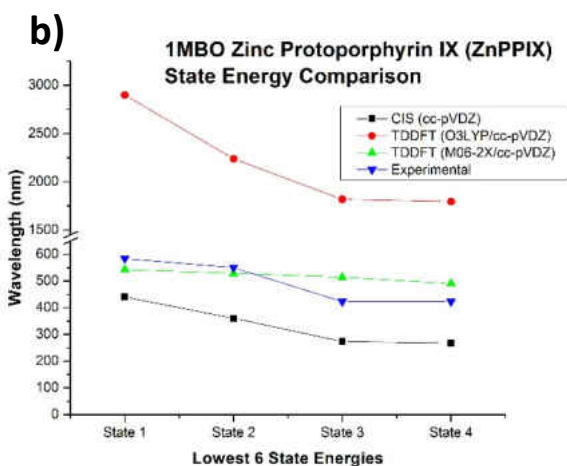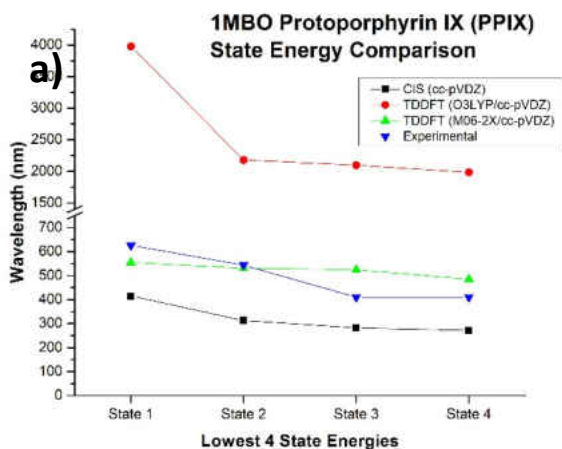
Figure 8.14. The four visible transitions of non-planar, 1MBO protoporphyrin IX calculated with two computational methods compared to experimental values. a) Free-base 1MBO PPIX; b) 1MBO ZnPPIX.

A discussion ensued following these results as to whether the optimization step must be taken for experimentally-obtained structures. Because these 1MBO porphyrin structures were removed from the forces that induced the non-planarity, i.e., the steric and electrostatic forces present inside the protein matrix, to undergo calculations essentially in vacuum, the structure

was likely not at a minimum on the potential energy surface, which is the criteria for any computational method to return valid results. Thus, an optimization process was undertaken for these structures. Had the protein matrix been modeled along with the porphyrin with experimental parameters, then the optimization step may not have been necessary. For comparison, transition energies for a non-optimized structure of PPIX were calculated. The values presented in Table 8.6 show clearly that the optimization step has a profound effect on returned transition energies.

Table 8.6. Comparison of transition energies calculated for optimized and non-optimized structures of 1MBO PPIX and 1MBO ZnPPIX.

Species / State	Experimental Transition Energies (nm)	Optimized Structure (TDDFT/ M06-2X)	Non-optimized structure (TDDFT/ M06-2X)
1MBO PPIX			
State 1	626	555	841
State 2	544	531	744
State 3	410	524	706
State 4	410	485	699
1MBO ZnPPIX			
State 1	584	543	764
State 2	550	529	698
State 3	423	514	698
State 4	423	491	691

8.4 Results of ground-state calculations under perturbation

Several calculations were made for *ground-state* zinc porphyrins in various environments:

- charged propionate groups vs. neutral dimethyl ester substituents;

- solvated in different dielectric constant values;

- planar porphin ring vs. non-planar conformation found in the 1MBO protein structure;

- with an applied field;

- with point charges at the position of myoglobin atoms.

The purpose of these calculations was to begin to quantify the kind of effect an electric field has on a cofactor molecule like porphyrin and the cursory results only serve to dramatize the influence such electrostatic changes might have on an active site molecule.

First, the value that was examined was electric field generated by the porphyrin itself, i.e., the "self" field. This is not a value that arises experimentally with the QM Stark method because the porphyrin acts as a probe whose own electrostatics is not sensed by the transition dipoles. However, in this context, a qualitative picture of how sensitive this property of the molecule is to its surroundings is interesting. Then, a few electronic properties of zinc porphyrin are tallied under various environments: permanent dipole moment and the partial atomic charge on oxygen, pyrrole nitrogens, and the central metal.

8.4.1 "Self" electric field vector

The electric field was calculated for the molecule of interest in *G09* by including the "Prop=Field" keyword in an energy job. All calculations were performed at the DFT / M06-2X level. Electric field results were returned numerically resolved into *x-, y-,* and *z*-components for the standard orientation with the origin at the metal center and the porphin ring oriented in the *x-y* plane with nitrogen atoms situated on the axes; the vector value was then generated and plotted. The common origin of the three-dimensional plots represent the spatial position of the zinc atom of the porphyrin. The vector labeled ZnPPIX is the value calculated *in vacuo*.

Overall, the results show that the relative magnitude of the field is not nearly as sensitive to various environments as direction. Two comparisons of interest from these calculations are shown below: solvent effect (Fig. 8.15) and applied field vs. point charge effect (Fig. 8.16).

Structures were solvated using the "SCRF" keyword (self-consistent reaction field) for pre-defined solvents in *G09*.[13] This method uses bulk dielectric constant values as well as solvent molecule size to solvate the molecule interest. Three solvents of vastly different dielectric constant values were applied: water (ε = 78.3553); acetonitrile (ε = 35.668); chloroform (ε = 4.7113).
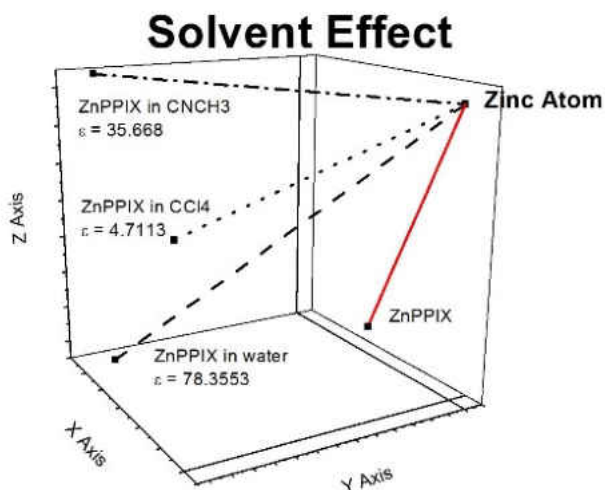
Figure 8.15. Calculated "self" electric-field vector for ZnPPIX in various solvents: water, acetonitrile, and chloroform.

Then, electrostatic influences were added to the environment of 1MBO ZnPPIX without a specific molecular structure to generate them. An applied electric field can be generated with the "Field" keyword and specified in magnitude and direction by loading individual components by adding, for example, "Field=X+10," to apply an electric field in the x-direction of 0.001 a.u., which is about 1 MV/cm. Results in Fig. 8.16 show that fields in the range of 0.1–1.0 MV/cm have

minimal effect on shifting the direction of the "self" electric field vector, even if applied opposing

to the direction of that vector. Increasing the magnitude of the field to about 10 MV/cm,

however, did cause a noticeable shift.

In addition, 133 point charges representing myoglobin's active site atoms (as defined in

Chapter 3) were added to the environment using the "Charge" keyword, coordinates from 1MBO,

and partial atomic charges given by the AMBER-based results in Ref. [14]. This representation of

the protein atom nearest to the active site had an influence that was highly significant in terms

of magnitude and direction, especially in the *z*-direction, as illustrated in Fig. 8.16.
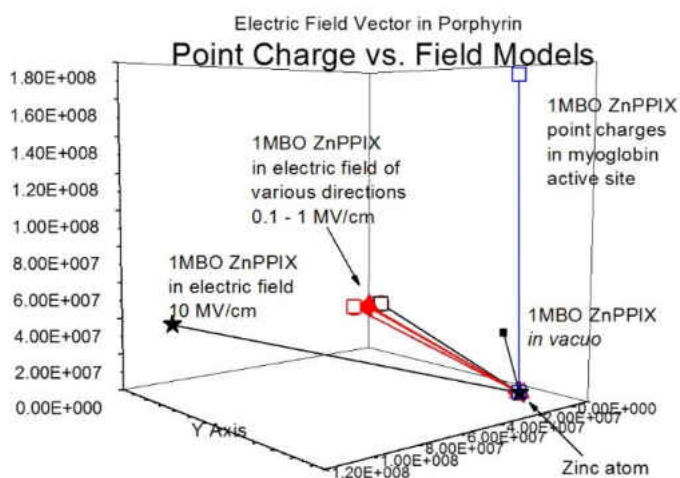
Figure 8.16. Calculated "self" electric-field vector for 1MBO ZnPPIX in various electrostatic environments. The point charge model was the most dramatically influential, a visual inspection placing it on par with the application of 10 MV/cm of an external electric field.

8.4.2   Electronic properties of zinc porphyrin in various environments

The following Table 8.7 tabulates dipole moment and partial atomic charge at the oxygen,

nitrogen, and metal atoms in various species of ZnPPIX as described. The most dramatic

perturbation in this set was to cap the propionate groups as dimethyl ester substituents; the

magnitude of the dipole moment almost disappears without the two full charges of the

propionate groups, which reaffirm s earlier conclusions that the charged propionates are the

main contributor the in-plane internal electric field.

Table 8.7. A comparison of electrostatic properties of zinc porphyrins under various environments.

Species	Dipole Moment (D)	Average Charge at Oxygen	Average Charge at Pyrrole Nitrogen	Charge on Zinc
ZnPPIX, *in vacuo*	39.9177	-0.44115	-0.46872	0.78157
ZnPPIX in water	49.9926	-0.46532	-0.46060	0.85748
ZnPPIX in electric field	31.8672	-0.44118	-0.46851	0.78099
ZnPPIX DME	3.2476	-0.27830	-0.46774	0.79993
ZnPPIX DME in water	4.8083	-0.29610	-0.45856	0.87109
1MBO ZnPPIX	44.8045	-0.42915	-0.50204	0.81265
1MBO ZnPPIX in electric field	44.5744	-0.42889	-0.50207	0.81263

8.4.3   Effect of high energy on porphin structure

Finally, when reporting the number and energies of influential high-lying states in the QM Stark analysis, the question of impact to the porphin structure is often posed. It is important to clarify that the analysis takes into account the *influence* of these high-energy states is included in the analysis as part of the diagonalization but a *transition* to the influential states is never undergone, neither experimentally nor theoretically. Nonetheless, the energy of various

processes that alter molecular structure were calculated for free-base porphin with TDDFT/O3LYP/cc-pVDZ: single ionization energy, double ionization energy, proton affinity energy, and total dissociation energy. This exercise serves to better define the energy range described by the QM Stark analysis but does not represent a real event to cause concern in the experimental process.

The basic computational procedure to calculate the energy of such processes is to optimize and calculate energy for the molecular species "before" and "after," i.e., for ionization energy, the "before" is neutral porphin and "after" is the porphin monocation. Then the difference in energy represents the process energy. In addition, the proper procedure not only includes the DFT (analogous to the SCRF energy in HF) but also the zero-point energy (ZPE) as obtained from a frequency calculation. The magnitude of energy differences between species in a particular process can be on the order of the ZPE and its inclusion is clearly necessary. The process energy is reported in a.u. and also converted to wavenumber in cm^{-1} to place the energy in terms of transition frequency.

8.4.3.1 Ionization energy

Table 8.8. The single ionization energy of porphin, i.e., the loss of an electron, which aligns energetically with the energy of the electronic transition to higher-lying state, $p = 14$.

Molecule	Energy (a.u.)	Zero-Point Energy (a.u.)	Total Energy (a.u.)
PH_2 (neutral porphin)	-989.072159269	0.29285430	-988.7793050
PH_2^+ (porphin cation)	-988.830794035	0.29175528	-988.5390388
Ionization Energy	--	--	0.2402662 a.u. 52732 cm^{-1}

Table 8.9. The double ionization energy of porphin, i.e., the loss of two electrons, which aligns energetically with the energy of the electronic transition of a higher-lying state $p \gg 252$.

Molecule	Energy (a.u.)	Zero-Point Energy (a.u.)	Total Energy (a.u.)
PH_2 (neutral porphin)	-989.072159269	0.29285430	-988.7793050
PH_x^{2+} (porphin dication)	-988.44777	0.292807	-988.1549628
Ionization Energy	--	--	1.3282055 a.u. 291 507 cm^{-1}

8.4.3.2 Proton affinity

Table 8.10. The proton affinity energy of porphin, i.e., a single deprotonation to leave the molecule negatively charged, which aligns energetically with the energy of the electronic transition of a higher-lying state $p > 252$.

Molecule	Energy (a.u.)	Zero-Point Energy (a.u.)	Total Energy (a.u.)
PH_2 (neutral porphin)	-989.072159269	0.29285430	-988.7793050
PH^- (porphin anion)	-988.480598125	0.27644740	-988.2041508
Proton Affinity Energy	--	--	0.5751542 a.u. 126 231 cm^{-1}

8.4.3.3 Total dissociation energy

Table 8.11. The total dissociation energy of porphin, i.e., the obliteration of the molecule into its constituent atoms, which is far beyond any energy within reasonable range to calculate, approximately 20 times the transition to $p = 100$ and 16 times the transition to $p = 252$.

Molecule	Energy (a.u.)	Zero-Point Energy (a.u.)	Total Energy (a.u.)
PH$_2$ (neutral porphin)	-989.072159269	0.29285430	-988.7793050
Atom	**Energy (a.u.)**	**Number of Atom Type**	**Total Energy (a.u.)**
H	-0.496095884369	14	-6.945342388
C	-37.7689193571	20	-755.3783871
N	-54.4672590676	4	-217.8690363
Total Atomic Energy			-980.1927658
Total Dissociation Energy	--	--	8.5806845 a.u. 883 243 cm^{-1}

8.5 Conclusions

Methods to calculate the excited-state energies and transition dipole moment values required for successful QM Stark analysis were explored. Two distinct excited-state methods were calibrated: CIS and TDDFT. It was determined that TDDFT returned more accurate transition energies for porphin, zinc porphin, protoporphyrin IX, and zinc protoporphyrin IX (both planar and the conformation found in the 1MBO coordinates). For the porphin structures, the O3LYP functional with the cc-pVDZ basis set performed the best compared to experimental values. For the protoporphyrin structures, the M06-2X functional was superior to the O3LYP functional. However, the CIS method with the sdd basis set produced more reasonable transition dipole

moment values for free-base porphin. Additionally, because it was determined in Chapter 7 that excited states coupled to both the ground and first-excited state are necessary for analysis, CIS was the only method available to produce transition dipole moments originating from the first-excited state. Thus, it was determined that the values calculated with CIS were currently the best to be used with QM Stark analysis. It is clear, however, that the accuracy of calculations with this method can be improved, especially because it optimized a porphin that was not totally symmetric as the DFT method did.

In addition, the available computational resources were used to calculate some ground-state properties of porphyrins in various environments to begin to quantify the effect of electric fields on porphyrin electronic structure. The "self" electric field vector of zinc porphyrin was visualized in solvents with different dielectric constants as well as with applied field fields (0.1 – 10 MV/cm) and point charges representing the myoglobin matrix. It was clear that the direction of the self field was the most sensitive to surroundings. Calculations that compared charged versus neutral propionate groups in zinc protoporphyrin IX showed a substantial change in the permanent dipole moment of the molecule. Finally, after the discussion of very high-energy electronic states in Chapter 7, the ionization energies, proton affinity, and total dissociation energy of porphin were calculated. Although it must be emphasized that the analysis including such states is not equivalent to the molecule physically undergoing that transition, it was determined that within the energies of essential states, a single ionization might occur but no structural alteration would be expected.

8.6 References

1.	Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.02*, Gaussian, Inc.: Wallingford, CT, 2009.

2.	Sundholm, D., Interpretation of the electronic absorption spectrum of free-base porphin using time-dependent density-functional theory. *Physical Chemistry Chemical Physics* **2000,** *2* (10).

3.	(a) Foresman, J. B.; Frisch, A., *Exploring Chemistry with Electronic Structure Methods*. 2nd ed.; Gaussian, Inc.: Pittsburgh, PA, 1996; (b) MacQuarrie, D.; Simon, J. D., *Physical Chemistry: A Molecular Approach*. University Science Books: Sausalito, CA, 1997.

4.	(a) Cramer, C. J., *Essential of Computational Chemistry: Theories & Models*. John Wiley & Sons: West Sussex, England, 2002; (b) Fiolhais, C.; Nogueira, F.; Marques, M., *A Primer in Density Functional Theory*. Springer: Berlin, 2003.

5.	Gaussian09 User's Reference, DensityFit keyword. (accessed 5 August 2015).

6.	Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J., Toward a systematic molecular orbital theory for excited states. *The Journal of Physical Chemistry* **1992,** *96* (1), 135-149.

7.	(a) Ginsbergen, S. J. A. v.; Rosa, A.; Ricciardi, G.; Baerends, E. J., Time-dependent density functional calculations on the electronic absorption spectrum of free base porphin. *Journal of Chemical Physics* **1999,** *111* (6), 2499-2506; (b) Palummo, M.; Hogan, C.; Sottile, F.; Bagalá, P.; Rubio, A., Ab initio electronic and optical spectra of free-base porphyrins: The role of electronic correlation *Journal of Chemical Physics* **2009,** *131*, 084102; (c) Bauernschmitt, R.; Ahlrichs, R., Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory. *Chemical Physics Letters* **1996,** *256*, 454-464; (d) VanCaillie, C.; Amos, R. D., Geometric derivatives of excitation energies using SCF and DFT. *Chemical Physics Letters* **1999,** *308*, 249-255; (e) VanCaillie, C.; Amos, R. D., Geometric derivatives of density functional theory excitation energies using gradient-corrected functionals. *Chemical Physics Letters* **2000,** *317*, 159-164; (f) Furche, F.; Ahlrichs, R., Adiabatic time-dependent density functional methods

for excited state properties. *Journal of Chemical Physics* **2002,** *117*, 7433-7447; (g) Scalmani, G.; Frisch, M. J.; Mennucci, B.; Tomasi, J.; Cammi, R.; Barone, V., Geometries and properties of excited states in the gas phase and in solution: Theory and application of a time-dependent density functional theory polarizable continuum model. *Journal of Chemical Physics* **2006,** *124*, 094107; (h) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J., An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules. *Journal of Chemical Physics* **1998,** *109*, 8218-8224.

8.   Edwards, L.; Dolphin, D. H.; Gouterman, M.; Adler, A. D., Porphyrins XVII. Vapor absorption spectra and redox reactions: Tetraphenylporphins and porphin. *Journal of Molecular Spectroscopy* **1971,** *38* (1), 16-32.

9.   (a) Geissinger, P., Woehl, Jorg C., Prince, Barry J., A quantum-mechanical model for the determination of internal electric fields at protein active sites from the Stark effect on persistent spectral holes. *Journal of Luminescence* **2004,** *107*, 220-229; (b) Geissinger, P.; DuPrey, M. R.; Schwabacher, I. J.; Woehl, J. C.; Prince, B. J., Evaluation and Properties of a Model for the Determination of Internal Electric Fields in Proteins from the Stark Effect of Spectral Holes. *Optics & Spectroscopy* **2005,** *98* (5), 669-674.

10.   (a) Spassova, M.; Monev, V.; Kanevat, I.; Champagne, B.; Mosley, D. H.; Andréb, J.-M., Ab Initio Summation Over States/SCI for Static and Dynamic Hyperpolarizabilities of Small Molecules. In *Quantum Systems in Chemistry and Physics*, Hernández-Laguna, A.; Maruani, J.; McWeeny, R.; Wilson, S., Eds. Springer: 2000; Vol. 1, pp 101-125; (b) Champagne, B. t.; Kirtman, B., Evaluation of alternative sum-over-states expressions for the first hyperpolarizability of push-pull conjugated systems. *The Journal of Chemical Physics* **2006,** *125* (2), 024101; (c) Panja, N.; Ghanty, T.; Nandi, P., A sum-over-state scheme of analysis of hyperpolarizabilities and its application to spiroconjugated molecular system. *Theoretical Chemistry Accounts* **2010,** *126* (5-6), 323-337.

11.   Catalan, J.; Del Valle, J. C., A Spectroscopic Rule from the Solvatochromism of Aromatic Solutes in Non-Polar Solvents. *The Journal of Physical Chemistry B* **2014**.

12.   (a) Seely, G. R.; Talmadge, K., PHOTOREDUCTION OF ZINC PORPHIN BY ASCORBIC ACID*. *Photochemistry and Photobiology* **1964,** *3* (3), 195-206; (b) Canters, G. W.; Jansen, G.; Noort, M.; Van der Waals, J. H., High resolution Zeeman experiments on singlet, triplet, and quartet states of metalloporphines. *The Journal of Physical Chemistry* **1976,** *80* (20), 2253-2259.

13.   Gaussian09 User's Reference, SCRF keyword. (accessed 7 August).

14.   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Kenneth M. Merz, J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995,** *117*, 5179-5197.

Chapter 9

Conclusions & Future Directions

9.1 Conclusions

Having established the functional implications of a net internal electric field, $\vec{E}_{int}$, in heme proteins, this work contributed towards the goal of quantifying such a field. Three viewpoints for modeling the overall electrostatics of a protein matrix were proposed: a bulk dielectric constant, a net electric field vector, and a point-charge model. The simplistic Coulomb's law method explored in Chapter 3 for estimating $\vec{E}_{int}$, has usefulness in terms of relative values. The field values can be expected to change with such structural perturbations such as the "two-level system" statistical range of positions for individual amino acids, mutations, and global dynamics (protein "breathing"). This approach is unique in that it treats the protein matrix as an "electrostatic structure" with an overall effect at a single point in the system, e.g., at the active site.

Porphyrin as a probe of $\vec{E}_{int}$ has the advantage of being a native reporter in this approach as a heme analog that can sense the electrostatic environment's effects as a whole on the cofactor. This means that the sum of the field can be seen on the porphyrin's electronic properties, not just a small portion of it. Methods for substituting porphyrins that are good spectroscopic probes for heme have been streamlined and a variety of porphyrin species have been prepared for that purpose, allowing for metallated and free-base options as well as highly-charged (e.g. with propionic and neutral (i.e., dimethyl ester derivatives) options. In addition, porphyrins under a

number of environmental conditions have been characterized spectrally. The increase in resolution afforded by fluorescence measurements at low temperature also revealed the opportunity to assure sample purity, which is essential to the success of the method proposed to measure $\vec{E}_{int}$, especially as the hole-burning Stark spectral resolution improves. In addition, the unexpected splitting in those low-temperature spectra qualitatively supports the notion that the protein matrix provides a consistent energetic environment for its cofactor, because only porphyrins substituted into protein exhibited the dramatic decrease in inhomogeneous broadening. This idea that a particular protein species provides a reproducible environment (within a dynamic range) is an important foundation for the very postulate that an internal electric field exists in proteins and can be measured.

Experimentally, it had been established that Stark spectroscopy can be used to measure $\vec{E}_{int}$. Spectral hole-burning at liquid-helium temperatures had also been shown to be an indispensable technique for improving resolution for samples like porphyrins that suffer from inhomogeneous line broadening. An updated experimental set-up has been constructed and preliminary data collected on PPIX–myoglobin. As the instrumentation becomes streamlined, the opportunity for a number of perturbations to the basis system becomes possible (see Future Directions). These experiments can be expected to capture some of the dynamic nature of the protein matrix suggested by calculations in Chapter 3.

The classical approach to Stark spectroscopy analysis has limitations for systems that do not exhibit a strictly linear shift. In addition, it does not allow for the deduction of a directional vector. A quantum-mechanical Stark analysis had been proposed that would not require arbitrarily

excluding higher-order terms in the analysis and would also allow for the resolution of the calculated $\vec{E}_{int}$ value into spatial components (*x,y,z*). However, some serious practical barriers prevented implementation because the method theoretically called for an *infinite number* of electronic excited states to be calculated and used as input for the method. This work offered relief from this burdensome criteria by showing that electronic states with energies higher than some predictable cutoff point did not appreciably influence the analysis, i.e., the internal electric field value would converge with the inclusion of a finite number of states. Additional support for the expectation of convergence was gathered from theoretical arguments based on oscillator strength and the *f*-sum rule. Then, with a now finite set of electronic states that energetically lay within range of the transition of interest, a method to deduce the most influential states within that set was developed using perturbation theory. The internal electric field value was indeed shown to converge using this new protocol (approximately 1.7 MV/cm for each of the two in-plane components of the field), essentially moving the QM Stark analysis from a purely theoretical to a much more practical realm in terms of usefulness. This analysis also strongly suggested that a planar porphin model was not sufficient for sensing the *z*-component of $\vec{E}_{int}$ but could possibly be included with a domed porphin structure.

Now with a protocol for QM Stark analysis in hand, methods for generating a large number electronic excited-state calculations were explored and compared, especially CIS and TDDFT. A proposal for "calibrating" a computational method was developed and described. Each method had advantages, CIS with transition dipole moment accuracy and TDDFT with accuracy for state energies, but current resources pointed to CIS as the only option for generating the excited-state to excited-state transition dipole moment values that were deemed necessary for the improved

QM Stark analysis protocol. Finally, computational methods with application to ground-state porphyrins were also explored with the intent of characterizing the electronic properties of the probe molecule. Environmental perturbations to the structure were applied to begin to quantify the effect that a net electric field of such large magnitude might have on an active site's electron distribution.

The current work supports future directions from both experimental and theoretical perspectives as described below.

9.2 Future directions

9.2.1 Experimental

Previous hole-burning Stark experiments used free-base PPIX–myoglobin to measure $\vec{E}_{int}$. However, as described in Chapter 4, the lack of a metal center in the free-base porphyrin is problematic both because it is not covalently linked to the protein and because a reaction field does not arise in the same way as with heme. Experiments with zinc PPIX can address these concerns as well as provide a point of comparison with previous measurements to test the hypothesis of a "well-defined field" (i.e., is the same $\vec{E}_{int}$ value deduced with a different probe). ZnPPIX–myoglobin samples have been prepared, characterized, and stored for this future experiment.

Another issue related to porphyrin structure (again described in Chapter 4) is the influence of highly-charged propionate groups in proximity to the porphin ring on the measured $\vec{E}_{int}$ because it is the protein matrix of interest in this measurement. A possibility for mitigating the influence of such an interference is to use dimethyl ester porphyrin species to cap the charged

substituents. Calculations in Chapter 8 show a dramatic change to electronic properties of zinc porphyrin when propionate groups are capped with methyl groups (see Table 8.7). PPIX DME–myoglobin has also been prepared, characterized, and stored for such an experiment.

The accuracy of an $\vec{E}_{int}$ calculation depends on an accurate local field factor, which, in turn, depends on an accurate dielectric constant value for the sample solvent. Experimentally, measuring the dielectric constant of 3:1 glycerol:water at low temperatures as well as that same solvent with dissolved protein would assist in improving its accuracy (or possibly confirming the current value). The effect of protein concentration on dielectric constant could inform the overall local field factor as well as the effect of various tertiary structures (globular vs. beta barrel, etc.). The challenges of measuring dielectric constant of a frozen liquid sample in a cryostat need to be overcome, with a capacitance method most likely to succeed but an optical method employing Brewster's angle might also be useful if an optical cryostat is employed.

With calculations pointing to the importance of protein dynamics on $\vec{E}_{int}$, intentionally altering the protein structure for different replicates of the measurement could reveal the sensitivity of the hole-burning Stark spectroscopy method. This could be accomplished with temperature cycling, solvents with different pH values, and solvents with different polarities. Temperature cycling capabilities are available with the current experimental set-up. In preparation for pH experiments, this work described explorations of the effect of pH on the myoglobin structure as well as a titration of a glycerol–water mixture (see Chapter 3). Information about using non-aqueous solvents in experiments with proteins can be found in Ref. [1].

The current experimental set-up also has the potential to carry out single-molecule experiments. Single-molecule Stark spectroscopy in comparison to the ensemble experiments with hole-burning also would be an important piece of information to support or refute the idea of a conserved $\vec{E}_{int}$ value for a particular protein species as well as establish the dynamic range of the value under some environmental conditions.

Finally, the ultimate test for this method's capabilities to quantify $\vec{E}_{int}$ is to use a series of protein species with heme cavities that are known to have different electrostatic qualities based on crystal structures. For example, while myoglobin has a neutral heme cavity, nitrophorin's has an overall negative charge while cytochrome *c* has an overall positive charge. If differences in $\vec{E}_{int}$ values for such a series differs appreciably, the method's reliability and resolution might be demonstrated.

9.2.2 Theoretical

For improving calculations of internal electric field using the Coulomb's law method, choosing an accurate partial atomic charge scheme is key. Although it has been mentioned that such partitioning schemes are ultimately arbitrary, there are certainly some that are more physically accurate than others. Bader's Atoms-in-Molecules (AIM) scheme[2] is suggested because it is available in *G09* along with exploring the possibility of bond-centered charges in addition to atom-centered charges (which could be accomplished with dummy atoms in *G09)*.

Beyond the Coulomb's law method, calculating the internal electric field with a quantum-mechanical method (including the "prop" keyword in an energy job – see Appendix A) could be

improved using a more extensive basis set beyond the STO-3G and 3-21G basis sets used for such a calculation in this work (see Chapter 3). In combination with the ONIOM (also known as QM/MM) approach, resource limitations would not be a concern. The calculations to determine influential residues (also Chapter 3) showed that the atoms with the most electrostatic contribution lie within about 10 Å of the heme iron. Using these results, the ONIOM layers could be established, i.e., the most influential residues would be modeled with an *ab initio* method while residues further away would be modeled with a molecular mechanics method. This should improve values for internal electric field because the influence of residues on each other in addition to the porphyrin would be accounted for. "Calibrating" basis sets for a method like this might be possible if visualized and compared to experimental electron densities reported by x-ray crystallography techniques. The dynamic nature of the internal electric field might also be captured with a molecular mechanics simulation under physiological conditions to obtain a time-resolved trajectory of atomic coordinates. The changes to the field at different points in time could be modeled with a vector that fluctuates in intensity and direction as the protein matrix evolves over the cycle of its "breathing" motion.

Further work on the effect of different electrostatic environments on the heme cofactor's electronic properties might also reveal some correlation between internal electric field and function. For example, how does the electron density change on the iron atom if a point-charge models of different proteins like myoglobin and nitrophorin are applied to heme? What about the direction of the field within the heme cavity to encourage or oppose the optimal binding angle of a ligand like CO or O_2 with respect to the metal?

Different porphyrin probes are desired for future experimental work and a complete set of excited-state energies and transition dipole moment values for different metallo-porphin probes, e.g., zinc porphin, are needed. Further work on the accuracy of excited-state energies and transition dipole moments is necessary to yield a reliable value for internal electric field calculated with the QM Stark analysis method. A higher-level excited-state method, CASSCF, has the ability to calculate the required excited-state to excited-state transition dipole moment values in *Gaussian09, Revision D.01*. This method should calibrated for calculating porphyrin excited states.

Improving the accuracy of excited-states calculations will also occur with a better model of the porphyrin itself. The "unperturbed" energy values required by the QM Stark analysis method actually should include all of the factors in energy *except* for the internal electric field itself. Thus, a planar porphin molecule *in vacuo* is missing factors such as substituents, structural strain on the ring (such as a domed configuration), and other environmental factors like solvation that could be included in a model of porphyrin for excited-state calculations.

Finally, beyond biological applications, the concept of net electrostatics does not seem to have been extended to inorganic catalysts. In particular, the increasing number of reports about metal–organic frameworks (MOF) seems applicable to this work, especially those that encapsulate heme to control the metalloporphyrin's binding capabilities. [3] In this way, these macro-structures imitate heme proteins but with a framework that is far simplified compared to even a small protein matrix like myoglobin. These MOFs could be utilized to study internal electric fields both experimentally and computationally to manipulate a structure systematically while

avoiding the specific conditions that a biological system needs for mutation. In addition, once the

role of internal electric field is better understood with respect to function, the consideration of

these net electrostatics would become a part of the engineering process for catalysts like MOFs.

9.3 References

1.   Mattos, C.; Ringe, D., Proteins in organic solvents. *Current Opinion in Structural Biology* **2001,** *11*, 761-764.

2.   Bader, R. F. W., A quantum theory of molecular structure and its applications. *Chemical Reviews* **1991,** *91* (5), 893-928.

3.   (a) Anderson, J. S.; Gallagher, A. T.; Mason, J. A.; Harris, T. D., A Five-Coordinate Heme Dioxygen Adduct Isolated within a Metal-Organic Framework. *Journal of the American Chemical Society* **2014**; (b) Alkordi, M. H.; Liu, Y.; Larsen, R. W.; Eubank, J. F.; Eddaoudi, M., Zeolite-like Metal-Organic Frameworks as Platforms for Applications: On Metalloporphyrin-Based Catalysts. *Journal of the American Chemical Society* **2008,** *130* (38), 12639-12641; (c) Durot, S.; Taesch, J.; Heitz, V., Multiporphyrinic Cages: Architectures and Functions. *Chemical Reviews* **2014**; (d) Juan-Alcaniz, J.; Gascon, J.; Kapteijn, F., Metal-organic frameworks as scaffolds for the encapsulation of active species: state of the art and future perspectives. *Journal of Materials Chemistry* **2012,** *22* (20), 10102-10118; (e) Odoh, S. O.; Cramer, C. J.; Truhlar, D. G.; Gagliardi, L., Quantum-Chemical Characterization of the Properties and Reactivities of Metal–Organic Frameworks. *Chemical Reviews* **2015,** *115* (12), 6051-6111; (f) Wang, K.; Feng, D.; Liu, T.-F.; Su, J.; Yuan, S.; Chen, Y.-P.; Bosch, M.; Zou, X.; Zhou, H.-C., A series of Highly Stable Mesoporous Metalloporphyrin Fe-MOFs. *Journal of the American Chemical Society* **2014**.

Appendix A

Gaussian09 Input

A brief introduction to the mechanics of running calculations with *Gaussian09* is given here. A more thorough introduction for a beginner can be found in Ref. [1], *Exploring Chemistry with Electronic Structure Methods*, by James Foresman and Aileen Frisch. A number of test jobs that correspond with the book's examples can be downloaded at http://www.gaussian.com/g_tech/1.htm (under "Exploring Chemistry examples and exercises"). The *G09 User's Manual* is not an instruction manual for a beginner like the above reference is intended to be but can certainly be helpful to determine options and to clarify keywords.

An example of an input file for running an energy job on water at the Hartree-Fock level with the 6-31G* basis set is given here:

%NProcShared=12

%Mem=47GB

%chk=water_test.chk

hf/6-31G(d) test

My test program: Water -- single point energy

0 1

O -0.464 0.177 0.0

H -0.646 1.137 0.0

H   0.441 -0.143 0.0

Input for *Gaussian09* to run on a computational cluster such as Cleve typically requires the following sections given in the example above. In general, the input is not case-sensitive but is sensitive to extra spaces and extra lines. A terminal empty line should be included at the end of an input file. (A program like Notepad++ that displays non-character keystrokes is helpful for determining whether subtle input mistakes are causing errors.) Input files have a **.com** extension for use with a Linux operating system (often the case for computational clusters) and a **.gjf** extension for use with a Windows operating system.

- Number of processors (cores) to be used for the job (**%NProcShared=**)

- Amount of memory designated for the job with units, e.g., GB, etc. (**%Mem=**)

- The name of a checkpoint file for information to be accessed and deposited; if a .chk file

  exists in the same folder as the input file, the input file can call on the. chk for information

(e.g., optimized geometry using keyword **geom=allcheck**) but will also overwrite this file with information from the current job (**%chk=[filename].chk**)

- A route section to specify the job parameters, including options for geometry coordinates and environments, e.g., solvation, electrostatic fields, or point charges (begins with **#**)

- Optional title card (no format requirements)

- Charge and multiplicity of the structure (two numbers separated by a space; negative charge specified by a negative sign) (not necessary if the **geom=allcheck** option is included that references a .chk file with existing information)

- Coordinates for structure geometry (unless an option such as **geom=allcheck** is used); default is Cartesian coordinates but z-matrix and symbolic Cartesian are two of several other options

As described in Chapter 8, a typical procedure for calculating energy will be a geometry optimization followed by a frequency job to check for imaginary frequencies, and finally an energy job (either ground-state or excited-state). The route sections corresponding to such jobs for a Hartree–Fock (**HF**) level calculation using the **sdd** basis set are:

- **Optimization**: # opt hf/sdd

- **Frequency**: # freq hf/sdd

- **Ground-state energy:** # hf/sdd

- **Excited-state energy to request 10 excited states:** # cis=(nstates=10)/sdd

The same types of jobs using DFT require a specified functional and basis set. Examples of route sections using the **O3LYP** functional and **cc-pVDZ** basis set are given here. In addition, an

optimization followed immediately by a frequency job can be requested for any method as shown here. An excited-state optimization route section is also given here, where **"root=2"** specifies that the second excited state be optimized, and can also be used with any excited-state method.

- **Optimization immediately followed by Frequency**: # opt freq o3lyp/cc-pvdz

- **Ground-state energy:** # o3lyp/cc-pvdz

- **Excited-state energy to request 10 excited states:** # td=(nstates=10) o3lyp/cc-pvdz

- **Optimization / Frequency of excited state:** # td=(nstates=10,root=2) opt freq o3lyp/cc-pvdz

Finally, some further options relevant to this work are:

- **Requesting excited-state to excited-state transition dipole moments (for CIS method *only*):** #p cis(nstate=252)/sdd density=alltransition **iop(9/44=3)** pop=none

- **Requesting calculation of potential and electric field *for* a system of interest:** include "**prop**" keyword in the route section of an energy job

- **Solvation:** include "**scrf=[solventname]**" in route section of any job type (list of available solvents at http://www.gaussian.com/g_tech/g_ur/k_scrf.htm)

- ***Applied* electric field *to* system of interest:** include "**field=[direction: *x,y*, or *z*]+[magnitude in 0.001 a.u.]**" to route section of any job type

- **Including point charges in system of interest:** include "**charge**" keyword to route section of any job type and following coordinate list of system of interest with position and charge of point charges as *x y z charge* (one line per charge); note that the coordinates of point

charges must be entered in the *standard orientation* of the system of interest (may need to check this with a preliminary job)

- **Using symbolic Cartesian coordinates for freezing one component of coordinates:** input file for a domed porphin in its entirety given in subsequent pages

Optimization input file for a non-planar porphin using symbolic Cartesian coordinates:

%nprocshared=12

%mem=47GB

%chk=Domed_Porphin_opt_hf_sdd_symcart.chk

#p opt=(calcfc,z-matrix) freq hf/sdd

Domed Porphin Opt, 50% displacement with starting structure

0 1

N	0	x0001	y0001	zdisp
H	0	x0002	y0002	z0002
N	0	x0003	y0003	zdisp
H	0	x0004	y0004	z0004
N	0	x0005	y0005	zdisp
N	0	x0006	y0006	zdisp
C	0	x0007	y0007	z0007
C	0	x0008	y0008	z0008
C	0	x0009	y0009	z0009
C	0	x0010	y0010	z0010
C	0	x0011	y0011	z0011
H	0	x0012	y0012	zzero
H	0	x0013	y0013	zzero
H	0	x0014	y0014	z0014
C	0	x0015	y0015	z0015
C	0	x0016	y0016	z0016
C	0	x0017	y0017	z0017
C	0	x0018	y0018	z0018

C	0	x0019	y0019	z0019
H	0	x0020	y0020	zzero
H	0	x0021	y0021	zzero
H	0	x0022	y0022	z0022
C	0	x0023	y0023	z0023
C	0	x0024	y0024	z0024
C	0	x0025	y0025	z0025
C	0	x0026	y0026	z0026
H	0	x0027	y0027	zzero
H	0	x0028	y0028	zzero
C	0	x0029	y0029	z0029
H	0	x0030	y0030	z0030
C	0	x0031	y0031	z0031
C	0	x0032	y0032	z0032
C	0	x0033	y0033	z0033
C	0	x0034	y0034	z0034
C	0	x0035	y0035	z0035
H	0	x0036	y0036	zzero
H	0	x0037	y0037	zzero
H	0	x0038	y0038	z0038

Variables:

y0001 =   0.000000

x0001 =   2.130183

z0002 =   0.000000

y0002 =   0.000000

x0002 =   1.115693

y0003 =   0.000000

x0003 =   -2.130183

z0004 =   0.000000

y0004 =   0.000000

x0004 =   -1.115693

y0005 =   2.036917

x0005 =   0.000000

y0006 =   -2.036917

x0006 =   0.000000

z0007 =   0.000000

y0007 =   -1.129296

x0007 =   2.903830

z0008 =   0.000000

y0008 =   1.129296

x0008 =   2.903830

z0009 =   0.000000

y0009 =   -0.687482

x0009 =   4.269410

z0010 =   0.000000

y0010 =   0.687482

x0010 =   4.269410

z0011 =   0.000000

y0011 =   -2.442227

x0011 =   2.423865

y0012 =   -1.351377

x0012 =   5.130435

y0013 =   1.351377

x0013 =   5.130435

z0014 =   0.000000

y0014 =   -3.223296

x0014 =   3.185443

z0015 =   0.000000

y0015 =   -2.861996

x0015 =   1.083860

z0016 =   0.000000

y0016 =   -2.861996

x0016 =   -1.083860

z0017 =   0.000000

y0017 =   -4.264845

x0017 =   0.679888

z0018 =   0.000000

y0018 =   -4.264845

x0018 =   -0.679888

z0019 =   0.000000

y0019 =   -2.442227

x0019 =   -2.423865

y0020 =   -5.119182

x0020 =   1.354693

y0021 =   -5.119182

x0021 =   -1.354693

z0022 =   0.000000

y0022 =   -3.223296

x0022 =   -3.185443

z0023 =   0.000000

y0023 =   -1.129296

x0023 =   -2.903830

z0024 =   0.000000

y0024 =   -0.687482

x0024 =   -4.269410

z0025 =   0.000000

y0025 =   0.687482

x0025 =   -4.269410

z0026 =   0.000000

y0026 =   1.129296

x0026 =   -2.903830

y0027 =   -1.351377

x0027 =   -5.130435

y0028 =   1.351377

x0028 =   -5.130435

z0029 =   0.000000

y0029 =   2.442227

x0029 =   -2.423865

z0030 =   0.000000

y0030 =   3.223296

x0030 =   -3.185443

z0031 =   0.000000

y0031 =   2.861996

x0031 =   -1.083860

z0032 =   0.000000

y0032 =   4.264845

x0032 =   -0.679888

z0033 =   0.000000

y0033 =   4.264845

x0033 =   0.679888

z0034 =   0.000000

y0034 =   2.861996

x0034 =   1.083860

z0035 =   0.000000

y0035 =   2.442227

x0035 =   2.423865

y0036 =   5.119182

x0036 =   -1.354693

y0037 =   5.119182

x0037 =   1.354693

z0038 =   0.000000

y0038 =   3.223296

x0038 =   3.185443

Constants:

zdisp = 0.600000

zzero = 0.000000

References

1.   J. B. Foresman and A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, 2nd ed. (Gaussian, Inc., Pittsburgh, PA, 1996).

Appendix B

Stark06 Input Files

      Three types of input files are required for QM Stark analysis with *Stark06*: Hamiltonian states file (.HST), spectral data file (.DAT), and spectral parameter file (.SPL). Examples of the .HST file and the .SPL file are given here; the .DAT file contains two columns with individual points collected for Stark spectra: energy in cm^{-1} and an intensity value. In addition a screen shot is also included of the pop-up box where the remainder of the input is entered: Lorentz local field factor, steps for orientational averaging, states to include in the Hamiltonian (out of the maximum uploaded in a .HST file), observed transition (i.e., excited-state in the experimental spectra), factor for first amoeba vertex set (fitting routine step size), initial values for all three components of the internal electric field ($\vec{E}_{int}$), and check boxes to designate which components are to be calculated ("allowed to vary").

Table B.1. .HST file for free-base porphin (CIS / sdd) with directly coupled states only (2nd order terms) as described in Chapter 7.

Wavenumbers	Transition to State	Transition Dipole (D)	Transition Polarization	Transition to State	Transition Dipole (D)	Transition Polarization
23207	0	0.8464	Y			
37361	0	9.5648	X	1	1.8557	Y
37342	0	9.4876	Y	1	2.1239	X
51567	0	0.6761	Y	1	6.2888	X
41380	0	-7.4905	Y	1	0.6212	X
49783	0	-4.1575	X	1	6.2957	Y
55393	0	-0.9605	Y	1	5.9477	X
42360	0	7.3319	X	1	1.9729	Y
55546	0	-0.6776	X	1	-5.4696	Y
54605	0	-3.7326	Y	1	-0.1634	X
83022	0	4.249	Y	1	0.0366	X
66752	0	3.5419	X	1	-0.7305	Y
63796	0	0.9237	Y	1	-2.4088	X
72734	0	3.0107	X	1	0.1182	Y
52351	0	3.1317	X	1	1.3591	Y
78826	0	-2.7817	X	1	0.093	Y
75536	0	-2.7446	Y	1	0.5775	X
85147	0	2.4401	X	1	-0.0475	Y
87699	0	-2.4469	Y	1	-0.1001	X
63421	0	2.0715	X	1	0.5099	Y
80292	0	2.1081	Y	1	-0.3022	X
82000	0	-2.1038	Y	1	0.4046	X
83172	0	-2.0021	X	1	0.231	Y
90096	0	-2.0621	X	1	0.1395	Y
66683	0	1.7261	Y	1	0.2687	X

Wavenumbers	Transition to State	Transition Dipole (D)	Transition Polarization	Transition to State	Transition Dipole (D)	Transition Polarization
51300	0	-1.7612	X	1	1.6898	Y
52549	0	0.6365	Y	1	-1.1339	X
81244	0	-1.6227	Z	1	0	0
27575	0	0.9811	X	1	0.518	Y
89135	0	1.5883	X	1	0.349	Y
77412	0	0.7971	X	1	1.3164	Y
72109	0	-0.5195	X	1	-1.1415	Y
77332	0	-0.7933	Y	1	-1.2691	X
83947	0	-1.3499	X	1	0.1645	Y
85802	0	1.312	X	1	0.3665	Y
70176	0	-0.2267	X	1	0.929	Y
71349	0	-0.3497	Y	1	-0.9254	X
76084	0	0.3101	X	1	0.8006	Y
79847	0	-1.0551	Y	1	0.4194	X
82990	0	0.9308	Z	1	0	0
65330	0	-1.8461	X	1	1.6028	Y
82596	0	-0.8878	Y	1	0.2537	X
85421	0	0.0206	Y	1	0.6784	X
84278	0	-0.7361	Z	1	0	0
83500	0	-0.1632	X	1	0.6189	Y
46864	0	0.4979	Z	1	0	0
73944	0	0.5635	Z	1	0	0
86046	0	-0.6001	Z	1	0	0
87347	0	-0.6464	X	1	-0.2227	Y
73373	0	0.3874	Y	1	0.548	X
84806	0	0.5737	Z	1	0	0

Wavenumbers	Transition to State	Transition Dipole (D)	Transition Polarization	Transition to State	Transition Dipole (D)	Transition Polarization
87742	0	0.5587	Z	1	0	0
86757	0	-0.5617	Y	1	0.1238	X
60911	0	0.3691	X	1	-0.4293	Y
78842	0	0.4392	Y	1	0.0829	X
49595	0	-1.4999	Y	1	1.1166	X
85090	0	0	0	1	-0.3391	Z
90337	0	-0.4095	Z	1	0	0
60396	0	0.3111	Z	1	0	0
46777	0	0	0	1	-0.1845	Z
84854	0	0.3487	Z	1	0	0
87378	0	0.3086	Z	1	0	0
90550	0	-0.3007	Z	1	0	0
83692	0	0	0	1	-0.2336	Z
83947	0	0	0	1	-0.2323	Z
82311	0	0	0	1	0.2262	Z
88201	0	0	0	1	-0.2077	Z
89025	0	-0.2323	Z	1	0	0
61283	0	-1.0236	Y	1	-0.794	X
75754	0	0	0	1	-0.1645	Z
88250	0	0.2049	Z	1	0	0
80052	0	0	0	1	-0.1586	Z
80031	0	0.1617	Z	1	0	0
89902	0	-0.1723	Y	1	-0.0325	X
74581	0	-0.151	Z	1	0	0
88907	0	0	0	1	0.1301	Z
82031	0	0	0	1	-0.12	Z
69834	0	0.2313	Y	1	0.1591	X

Wavenumbers	Transition to State	Transition Dipole (D)	Transition Polarization	Transition to State	Transition Dipole (D)	Transition Polarization
88068	0	0	0	1	-0.12	Z
85253	0	0	0	1	-0.1159	Z
85852	0	0	0	1	0.1037	Z
60246	0	0	0	1	-0.0735	Z
87930	0	0.0813	Z	1	0	0
70853	0	0	0	1	-0.0524	Z
90003	0	0	0	1	-0.0587	Z
76236	0	0	0	1	-0.0485	Z
84257	0	0	0	1	-0.0508	Z
71088	0	-0.0519	Z	1	0	0
88700	0	-0.0463	Z	1	0	0
74117	0	0	0	1	-0.0287	Z
82457	0	0.0308	Z	1	0	0
75523	0	-0.0244	Z	1	0	0
74956	0	0	0	1	0.0165	Z
72734	0	-0.0188	Z	1	0	0
80214	0	-0.018	Z	1	0	0
86557	0	0	0	1	0.0145	Z
83196	0	0	0	1	0.0117	Z
76506	0	0.0122	Z	1	0	0
90700	0	0	0	1	-0.0053	Z
72108	0	0	0	1	0.002	Z

Table B.2. .SPL file for PPIX-substituted myoglobin weighted spectral data set (Myo A 16131) used for analysis in Chapter 7.

Filename	Applied field (V/cm)	Angle between laser polarisation and applied field (radians)	Zero-field lorentzian hole width (FWHM)
PP9MB218.DAT	11051	0	0.014001
PP9MB219.DAT	8288.4	0	0.014001
PP9MB220.DAT	5525.6	0	0.014001
PP9MB221.DAT	2762.8	0	0.014001
PP9MB918.DAT	11051	0	0.014001
PP9MB919.DAT	8288.4	0	0.014001
PP9MB920.DAT	5525.6	0	0.014001
PP9MB921.DAT	2762.8	0	0.014001
PP9MB265.DAT	11051	1.570796	0.012087
PP9MB266.DAT	8288.4	1.570796	0.012087
PP9MB267.DAT	5525.4	1.570796	0.012087
PP9MB268.DAT	2762.8	1.570796	0.012087

Figure B.1. Screen shot of the pop-up box for entry of other input parameters for a *Stark06* calculation with a background of the .SPL file displayed.
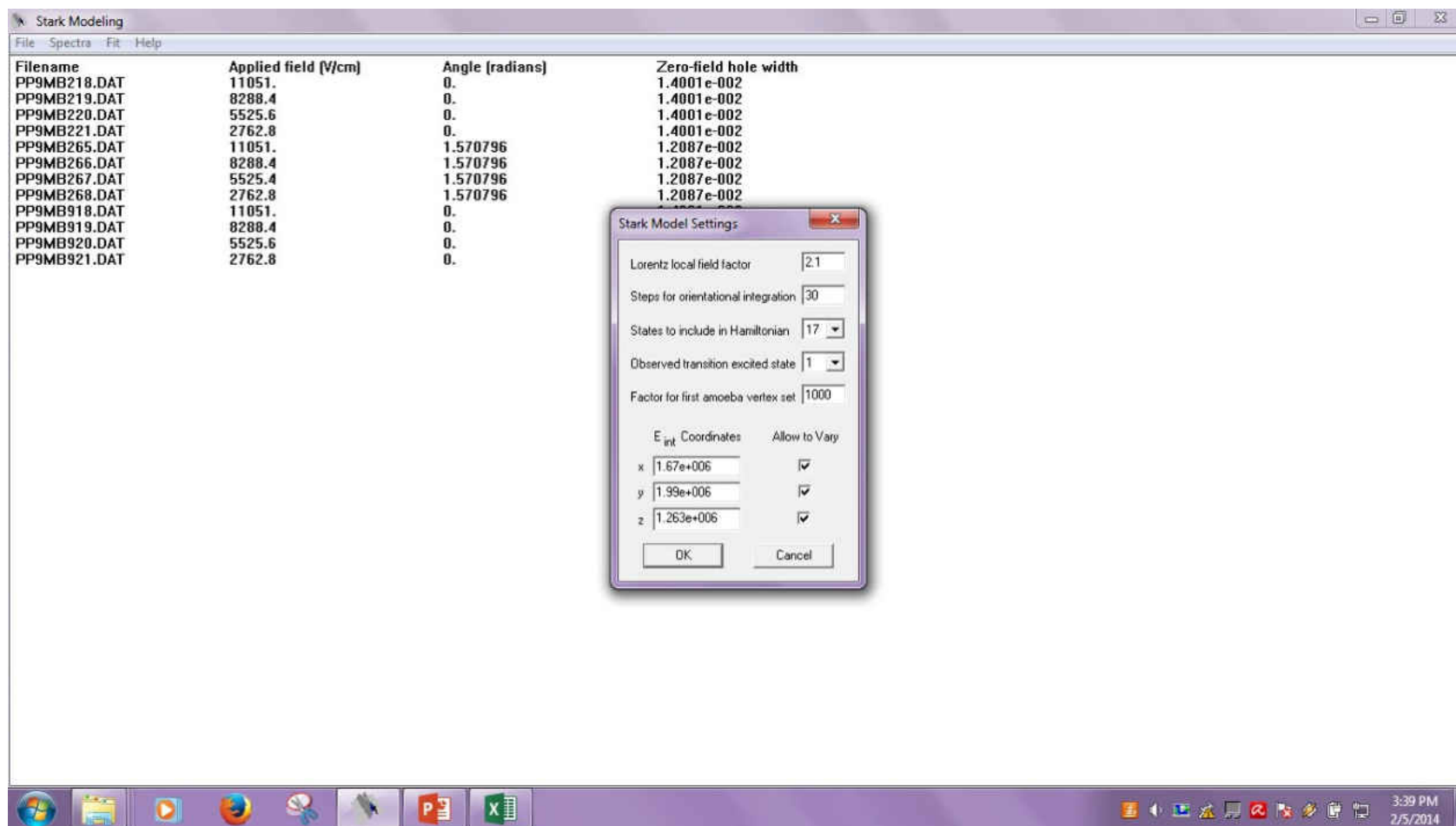
Appendix C

Amino Acid Abbreviation and Charge

Amino Acid	Three-letter abbreviation	One-letter abbreviation	Charged residue at pH 7?
Alanine	ALA	A	No
Aspartate	ASP	D	Yes
Cysteine	CYS	C	No
Glutamate	GLU	E	Yes
Phenylalanine	PHE	F	No
Glycine	GLY	G	No
Histidine	HIS	H	Sometimes
Isoleucine	ISE	I	No
Lysine	LYS	K	Yes
Leucine	LEU	L	No
Methionine	MET	M	No
Asparagine	ASN	N	No
Proline	PRO	P	No
Glutamine	GLN	Q	No
Arginine	ARG	R	Yes
Serine	SER	S	No
Threonine	THR	T	No
Valine	VAL	V	No
Tryptophan	TRP	W	No
Tyrosine	TYR	Y	No

Appendix D

Rotating Protein Data Bank Files for Heme Proteins into Standard Orientation

1. From the Protein Data Bank (www.rcsb.org/pdb/), search for the desired biological macromolecule. Be aware of parameters like species of origin, ligands, mutations, and environmental conditions (such as temperature, pressure, and pH) of the crystallization of that particular study. This information should be contained in the file header as REMARK 200 under EXPERIMENTAL DETAILS. For example, 1MBO is of sperm whale origin and oxygenated at the heme center; water and one sulfate ion also appear as heteroatoms (HETATM) in this file. One the file's summary page, there may also be "related structures" linked, i.e., the same or similar proteins under different conditions.

2. On the upper right corner of the summary page, go to "Download Files" → PDB File (Text) → Open with → Notepad. Save with a .pdb extension, not .txt.

3. Open the .pdb file (preferably in a program like Notepad++ that will retain the space-delimited format, which is required for input files for *G09*). Delete all HEADER, REMARKS, CONECT, etc., leaving only ATOM and desired HETATM (such as heme [HEM]). In addition, any solvent molecules in the file to be included or not should be considered. Save again as a .pdb under a new file name that is clear it is not the original downloaded file.

4. Structures that have residues with more than one probable location can be identified by an entry in Column 17, the Alternate Location Indicator (an A, B, etc. preceding the

287

Residue Name). All of the possible structures should be evaluated (e.g., 1MBO has four residues that have two equally probable locations). Inspect each possible structure in a visualization program like PyMOL (which displays both positions simultaneously) to verify that it is sterically sound.

5. Open one version file in GaussView 5. Right click on the structure → Edit → PDB Residues → Rows → Select All → click one button under the Highlight column (all should turn an aquamarine color → unhighlight HETATM residues) → Edit → Add Hydrogens to Selected Residues. (The "Add Hydrogens" in PyMOL labels the added atoms ambiguously.)

6. The structure should be oriented in a way that vector quantities have a reference (e.g., electric field components, transition dipole moments, etc.).For 1MBO, the heme iron should be the origin of the coordinate file, and the four pyrrole nitrogens should establish the *x–y* plane:

   a. Right click on the structure in GaussView 5 → View →Cartesian Axes, which allow visualization of the Cartesian axes.

   b. Right click → Edit → Atom List, which allows access to both Cartesian and Z-matrix coordinates (bond/angle/dihedral)

   c. Find the atom you want to be the origin (or create a dummy atom, X) and change its Tag to 1. The other atoms will re-Tag automatically.

   d. Create three dummy atoms to establish axes:

Table D.1. Position of dummy atoms in rotation process.

Atom Type	Tag	Description	X	Y	Z
X	2	*x*-axis unit	0.000	0.000	1.000
X	3	*y*-axis unit	1.000	0.000	0.000
X	4	*z*-axis unit	0.000	1.000	0.000

e. For heme, give pyrrole nitrogens tags as follows:

Table D.2. Tag for pyrrole nitrogens in rotation process.

Atom	Tag
NC	5
NB	6
NA	7
ND	8

f. Right click → File → Save → .gjf → uncheck "Write Cartesians → Check "New Molecule Group"

g. Reload the .gjf file in GaussView. Tag 1 is now at the Cartesian origin (view the axes as before).

h. In the Atom List Editor, establish the reference system so that the dummy atoms are reference by only two key atoms. For heme, NC (Tag 5) and NB (Tag 6) are the key atoms. Scan the file to be sure no other atoms reference the dummy atoms. Dummy atoms reference only the origin and each other. Save as a .gjf as before.

i. Instead of loading in GaussView 5, open the .gjf in Notepad. Change the Z-matrix components for the key atoms. If a variable (e.g., D2, A3, etc.) is altered, the line that defines the variable at the end of the file must be deleted. (If Z-matrix values

in the GaussView 5 Atom List Editor, only the key atoms will move and the rest of the structure will not follow because the program may re-Tag or re-Reference automatically.)

j. Reload the oriented .gjf file into GaussView 5. Right click → File → Save→ .pdb file → Check "Write Cartesians."

k. To use with the *proteinfield.c* program (described in Chapter 3 for calculating internal electric field with Coulomb's law), delete HEADER, CONECT, TER, and all HETATM (leaving only ATOM and END lines). If the atoms are not grouped by residue number, load into GaussView 5, open the Atom List Editor:

    i. Columns → PDB Data → Check Residue Number

    ii. Rows → Select All → Hit one Highlight button so they are all highlighted

    iii. Click on the Residue Number column → Rows → Sort Selected → Ascending by Residue Number

    iv. Edit → Reorder → All Atoms by Row

l. To create a *G09* input file, open a .pdb file with GaussView 5:

    i. Right click → Calculate → Gaussian Calculation Setup

    ii. The program will default to an energy calculation and method, basis set, charge, multiplicity, etc. will need to be chosen.

    iii. Click "Submit" button → Save (Gaussian input file) → .gjf

    iv. Resave as a .com file by opening .gjf with Notepad++ and giving it a different extension while selecting file type = "All Files."

Appendix E

Derivation of Sum-Over-States Polarizability Term and Difference

Term Used in QM Stark Analysis

E.1 Stark effect: classical description

Starting with a classical description of the Stark effect, let $\vec{F}$ be the electric field vector. The energy of the state *n* in the presence of an electric field is

$$E_n = E_n^{(0)} - \vec{F} \cdot \vec{\mu}_n - \frac{1}{2}\vec{F} \cdot \overleftrightarrow{\alpha}_n \cdot \vec{F} \tag{E.1}$$

where $\vec{\mu}_n$ is the permanent dipole moment vector in state *n* and $\overleftrightarrow{\alpha}_n$ is the polarizability tensor of state *n*. The transition energy of states *n* & *m* is then

$$E_m - E_n = E_m^{(0)} - E_n^{(0)} - \vec{F} \cdot \mu_m + \vec{F} \cdot \mu_n - \frac{1}{2}\vec{F} \cdot \overleftrightarrow{\alpha}_m \cdot \vec{F} + \vec{F} \cdot \overleftrightarrow{\alpha}_n \cdot \vec{F} = E_m^{(0)} - E_n^{(0)} - \vec{F} \cdot$$

$$(\vec{\mu}_m - \vec{\mu}_n) - \frac{1}{2}\vec{F} \cdot (\overleftrightarrow{\alpha}_m - \overleftrightarrow{\alpha}_n) \cdot \vec{F} \tag{E.2}$$

where $(\vec{\mu}_m - \vec{\mu}_n) =: \Delta\vec{\mu}_{mn}$ and $(\overleftrightarrow{\alpha}_m - \overleftrightarrow{\alpha}_n) =: \Delta\overleftrightarrow{\alpha}_{mn}$.

Here $\vec{F}$ is assumed to be a homogeneous field. Particle (atoms or molecule) is assumed to be neutral (and polarizable).

E.2 Stark effect: quantum-mechanical description

From time-independent perturbation theory, the Hamiltonian is

$$\hat{H} = \hat{H}_0 + \hat{H}_1 \tag{E.3}$$

where $\widehat{H}_1$ is the perturbation term. Under the influence of the perturbation, the energy levels of

the system, $\widehat{H}_0$, change to

$$E_n = E_n^{(0)} + \langle n|\widehat{H}_1|n\rangle + \sum_{m \neq n} \frac{\left|\langle n|\widehat{H}_1|m\rangle\right|^2}{E_n^{(0)} - E_m^{(0)}} + \dots \tag{E.4}$$

Let $\widehat{H}_1 = -\vec{F} \cdot \hat{\vec{\mu}}$. Then the first-order correction is

$$\langle n|\widehat{H}_1|n\rangle = -\langle n|\hat{\vec{\mu}}|n\rangle \cdot \vec{F} \tag{E.5}$$

where the expectation value of permanent dipole moment vector of state n is $\langle n|\hat{\vec{\mu}}|n\rangle$. If state n

is non-polar, the first-order correction vanishes.

The second-order correction is

$$\sum_{m \neq n} \frac{\left|\langle n|\widehat{H}_1|m\rangle\right|^2}{E_n^{(0)} - E_m^{(0)}} = \sum_{m \neq n} \left|\langle n|-\vec{F} \cdot \hat{\vec{\mu}}|m\rangle\right|^2 \tag{E.6}$$

Taking a look at matrix elements with spatial components:

$$\left|\langle n|-\vec{F} \cdot \hat{\vec{\mu}}|m\rangle\right|^2 = \langle n|-\vec{F} \cdot \hat{\vec{\mu}}|m\rangle\langle m|-\vec{F} \cdot \hat{\vec{\mu}}|n\rangle$$

$$= \langle n|F_x\hat{\mu}_x + F_y\hat{\mu}_y + F_z\hat{\mu}_z|m\rangle(-1)\langle m|F_x\hat{\mu}_x + F_y\hat{\mu}_y + F_z\hat{\mu}_z|n\rangle(-1) = \{\langle n|F_x\hat{\mu}_x|m\rangle +$$

$$\langle n|F_y\hat{\mu}_y|m\rangle + \langle n|F_z\hat{\mu}_z|m\rangle\} \times \{\langle m|F_x\hat{\mu}_x|n\rangle + \langle m|F_y\hat{\mu}_y|n\rangle + \langle m|F_z\hat{\mu}_z|n\rangle\} =$$

$$\langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_x|n\rangle F_x^2 + \langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_y|n\rangle F_x F_y + \langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_z|n\rangle F_x F_z +$$

$$\langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_x|n\rangle F_y F_x + \langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_y|n\rangle F_y^2 + \langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_z|n\rangle F_y F_z +$$

$$\langle n|\hat{\mu}_z|m\rangle\langle m|\hat{\mu}_x|n\rangle F_z F_x + \langle n|\hat{\mu}_z|m\rangle\langle m|\hat{\mu}_y|n\rangle F_z F_y + \langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_x|n\rangle F_z^2 = |\langle n|\hat{\mu}_x|m\rangle|^2 F_x^2 +$$

$$\left|\langle n|\hat{\mu}_y|m\rangle\right|^2 F_y^2 + |\langle n|\hat{\mu}_z|m\rangle|^2 F_z^2 + \{\langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_y|n\rangle + \langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_x|n\rangle\} F_x F_y +$$

$$\{\langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_z|n\rangle + \langle n|\hat{\mu}_z|m\rangle\langle m|\hat{\mu}_x|n\rangle\}F_xF_z + \{\langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_z|n\rangle +$$

$$\langle n|\hat{\mu}_z|m\rangle\langle m|\hat{\mu}_y|n\rangle\}F_yF_z \qquad\qquad\qquad (E.7)$$

Now writing out the classical 2nd-order term:

$$-\frac{1}{2}\vec{F}\cdot\vec{\alpha}_n\cdot\vec{F} = -\frac{1}{2}\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix}\begin{pmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{pmatrix}\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} \qquad (E.8a)$$

Because the tensor is symmetric, i.e., $\alpha_{ij} = \alpha_{ji}$,

$$= -\frac{1}{2}\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix}\begin{pmatrix} \alpha_{xx}\,F_x & \alpha_{xy}\,F_y & \alpha_{xz}\,F_z \\ \alpha_{yx}\,F_x & \alpha_{yy}\,F_y & \alpha_{yz}\,F_z \\ \alpha_{zx}\,F_x & \alpha_{zy}\,F_y & \alpha_{zz}\,F_z \end{pmatrix} \qquad (E.8b)$$

$$= -\frac{1}{2}\{\alpha_{xx}F_x^2 + \alpha_{xy}F_xF_y + \alpha_{xz}F_xF_z + \alpha_{yx}F_yF_x + \alpha_{yy}F_y^2 + \alpha_{yz}F_yF_z + \alpha_{zx}F_zF_x + \alpha_{zy}F_zF_y +$$

$$\alpha_{zz}F_z^2\} \qquad\qquad\qquad (E.8c)$$

The quantum-mechanical 2nd-order term is

$$\sum_{m\neq n}\frac{1}{E_n^{(0)}-E_m^{(0)}}\Big\{|\langle n|\hat{\mu}_x|m\rangle|^2 F_x^2 + \big|\langle n|\hat{\mu}_y|m\rangle\big|^2 F_y^2 + |\langle n|\hat{\mu}_z|m\rangle|^2 F_z^2 + \{\langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_y|n\rangle +$$

$$\langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_x|n\rangle\}F_xF_y + \{\langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_z|n\rangle + \langle n|\hat{\mu}_z|m\rangle\langle m|\hat{\mu}_x|n\rangle\}F_xF_z +$$

$$\{\langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_z|n\rangle + \langle n|\hat{\mu}_z|m\rangle\langle m|\hat{\mu}_y|n\rangle\}F_yF_z\Big\} \qquad (E.9)$$

Comparing the classical and quantum-mechanical:

$$\sum_{m\neq n}\frac{1}{E_n^{(0)}-E_m^{(0)}}|\langle n|\hat{\mu}_x|m\rangle|^2 F_x^2 == -\frac{1}{2}\alpha_{xx}F_x^2 \qquad (E.10)$$

Rearranging

$$\alpha_{xx} = -2\sum_{m \neq n} \frac{\left|\langle n|\hat{\mu}_x|m\rangle\right|^2}{E_n^{(0)} - E_m^{(0)}} \qquad \text{(E.11)}$$

Likewise

$$\sum_{m \neq n} \frac{1}{E_n^{(0)} - E_m^{(0)}} \langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_x|n\rangle F_x F_y = -\frac{1}{2}\alpha_{xy}F_x F_y \qquad \text{(E.12)}$$

Rearranging

$$\alpha_{xy} = -2\sum_{m \neq n} \frac{1}{E_n^{(0)} - E_m^{(0)}} \langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_x|n\rangle \qquad \text{(E.13)}$$

The remaining components are

$$\alpha_{xz} = -2\sum_{m \neq n} \frac{\langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_z|n\rangle}{E_n^{(0)} - E_m^{(0)}} \qquad \text{(E.14)}$$

$$\alpha_{yx} = -2\sum_{m \neq n} \frac{\langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_x|n\rangle}{E_n^{(0)} - E_m^{(0)}} \qquad \text{(E.15)}$$

By symmetry, $\alpha_{yx} = \alpha_{xy}$, so then $\langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_x|n\rangle = \langle n|\hat{\mu}_x|m\rangle\langle m|\hat{\mu}_y|n\rangle$ and the remaining y-

and z-containing components are then

$$\alpha_{yy} = -2\sum_{m \neq n} \frac{\left|\langle n|\hat{\mu}_y|m\rangle\right|^2}{E_n^{(0)} - E_m^{(0)}} \qquad \text{(E.16)}$$

$$\alpha_{yz} = -2\sum_{m \neq n} \frac{\langle n|\hat{\mu}_y|m\rangle\langle m|\hat{\mu}_z|n\rangle}{E_n^{(0)} - E_m^{(0)}} = \alpha_{zy} \qquad \text{(E.17)}$$

$$\alpha_{zx} = -2\sum_{m \neq n} \frac{\langle n|\hat{\mu}_z|m\rangle\langle m|\hat{\mu}_x|n\rangle}{E_n^{(0)} - E_m^{(0)}} = \alpha_{xz} \qquad \text{(E.18)}$$

$$\alpha_{zy} = -2\sum_{m \neq n} \frac{\langle n|\hat{\mu}_z|m\rangle\langle m|\hat{\mu}_y|n\rangle}{E_n^{(0)} - E_m^{(0)}} = \alpha_{yz} \qquad \text{(E.19)}$$

$$\alpha_{zz} = -2\sum_{m \neq n} \frac{|\langle n|\hat{\mu}_z|m\rangle|^2}{E_n^{(0)} - E_m^{(0)}} \tag{E.20}$$

In a generalized form

$$\alpha_{pr,n} = -2\sum_{m \neq n} \frac{\langle n|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|n\rangle}{E_n^{(0)} - E_m^{(0)}} \tag{E.21}$$

where $p, r = x, y, z$.

The difference between the polarizability of two states, *n* and *l*, is

$$\Delta\alpha_{pr,ln} = \alpha_{pr,l} - \alpha_{pr,n} = -2\sum_{m \neq l} \frac{\langle l|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|l\rangle}{E_l^{(0)} - E_m^{(0)}} + 2\sum_{m \neq n} \frac{\langle n|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|n\rangle}{E_n^{(0)} - E_m^{(0)}} =$$

$$-2\sum_{m \neq l,n} \frac{\langle l|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|l\rangle}{E_l^{(0)} - E_m^{(0)}} - 2\frac{\langle l|\hat{\mu}_p|n\rangle\langle n|\hat{\mu}_r|l\rangle}{E_l^{(0)} - E_n^{(0)}} + 2\sum_{m \neq n,l} \frac{\langle n|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|n\rangle}{E_n^{(0)} - E_m^{(0)}} + 2\frac{\langle n|\hat{\mu}_p|l\rangle\langle l|\hat{\mu}_r|n\rangle}{E_n^{(0)} - E_l^{(0)}}$$

$$\tag{E.22a}$$

Rearranging

$$\Delta\alpha_{pr,ln} = -2\sum_{m \neq n,l} \left\{ \frac{\langle l|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|l\rangle}{E_l^{(0)} - E_m^{(0)}} - \frac{\langle n|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|n\rangle}{E_n^{(0)} - E_m^{(0)}} \right\} - 2\frac{\langle l|\hat{\mu}_p|n\rangle\langle n|\hat{\mu}_r|l\rangle}{E_l^{(0)} - E_n^{(0)}} -$$

$$2\frac{\langle n|\hat{\mu}_p|l\rangle\langle l|\hat{\mu}_r|n\rangle}{E_l^{(0)} - E_n^{(0)}} = -2\sum_{m \neq n,l} \left\{ \frac{\langle l|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|l\rangle}{E_l^{(0)} - E_m^{(0)}} - \frac{\langle n|\hat{\mu}_p|m\rangle\langle m|\hat{\mu}_r|n\rangle}{E_n^{(0)} - E_m^{(0)}} \right\} -$$

$$\frac{2}{E_l^{(0)} - E_n^{(0)}} \left\{ \langle l|\hat{\mu}_p|n\rangle\langle n|\hat{\mu}_r|l\rangle + \langle l|\hat{\mu}_p|n\rangle\langle n|\hat{\mu}_r|l\rangle \right\} \tag{E.22b}$$

Now, in the principal axis system, $\overset{\leftrightarrow}{\alpha}$ is diagonal (i.e., set *p=r*) and Eq. E.22b becomes

$$\Delta\alpha_{pr,ln} = -2\sum_{m\neq n,l}\left\{\frac{|\langle l|\hat{\mu}_p|m\rangle|^2}{E_l^{(0)}-E_m^{(0)}} - \frac{|\langle n|\hat{\mu}_p|m\rangle|^2}{E_n^{(0)}-E_m^{(0)}}\right\} - \frac{2}{E_l^{(0)}-E_n^{(0)}}\left\{|\langle l|\hat{\mu}_p|n\rangle|^2 + |\langle n|\hat{\mu}_r|l\rangle|^2\right\} =$$

$$-2\sum_{m\neq n,l}\left\{\frac{|\langle l|\hat{\mu}_p|m\rangle|^2}{E_l^{(0)}-E_m^{(0)}} - \frac{|\langle n|\hat{\mu}_p|m\rangle|^2}{E_n^{(0)}-E_m^{(0)}}\right\} - \frac{4}{E_l^{(0)}-E_n^{(0)}}|\langle l|\hat{\mu}_p|n\rangle|^2 \qquad \text{(E.23)}$$

Breaking up the sum into two sums

$$\sum_{m\neq n,l} = \sum_{\substack{m\geq m_{crit}\\m\neq n,l}} + \sum_{\substack{m<m_{crit}\\m\neq n,l}} \qquad \text{(E.24)}$$

where m_{crit} is the value at which $E_m^{(0)} \gg E_l^{(0)} \wedge E_m^{(0)} \gg E_n^{(0)}$.

$$\Delta\alpha_{pr,ln} = -2\sum_{m\geq m_{crit}}\left\{\frac{|\langle l|\hat{\mu}_p|m\rangle|^2}{E_l^{(0)}-E_m^{(0)}} - \frac{|\langle n|\hat{\mu}_p|m\rangle|^2}{E_n^{(0)}-E_m^{(0)}}\right\} - 2\sum_{\substack{m<m_{crit}\\m\neq n,l}}\left\{\frac{|\langle l|\hat{\mu}_p|m\rangle|^2}{E_l^{(0)}-E_m^{(0)}} - \frac{|\langle n|\hat{\mu}_p|m\rangle|^2}{E_n^{(0)}-E_m^{(0)}}\right\} -$$

$$\frac{4}{E_l^{(0)}-E_n^{(0)}}|\langle l|\hat{\mu}_p|n\rangle|^2 \qquad \text{(E.25)}$$

Under the condition $E_m^{(0)} \gg E_l^{(0)}, E_n^{(0)}$,

$$\Delta\alpha_{pp,ln} \approx -2\sum_{m\geq m_{crit}}\left\{\frac{|\langle l|\hat{\mu}_p|m\rangle|^2}{(-E_m^{(0)})} - \frac{|\langle n|\hat{\mu}_p|m\rangle|^2}{(-E_m^{(0)})}\right\} - 2\sum_{\substack{m<m_{crit}\\m\neq n,l}}\left\{\frac{|\langle l|\hat{\mu}_p|m\rangle|^2}{E_l^{(0)}-E_m^{(0)}} - \frac{|\langle n|\hat{\mu}_p|m\rangle|^2}{E_n^{(0)}-E_m^{(0)}}\right\} -$$

$$\frac{4}{E_l^{(0)}-E_n^{(0)}}|\langle l|\hat{\mu}_p|n\rangle|^2 \qquad \text{(E.26a)}$$

Collecting terms,

$$\Delta\alpha_{pp,ln} \approx \sum_{m\geq m_{crit}}\frac{2}{E_m^{(0)}}\left\{|\langle l|\hat{\mu}_p|m\rangle|^2 - |\langle n|\hat{\mu}_p|m\rangle|^2\right\} - 2\sum_{\substack{m<m_{crit}\\m\neq n,l}}\left\{\frac{|\langle l|\hat{\mu}_p|m\rangle|^2}{E_l^{(0)}-E_m^{(0)}} - \frac{|\langle n|\hat{\mu}_p|m\rangle|^2}{E_n^{(0)}-E_m^{(0)}}\right\} -$$

$$\frac{4}{E_l^{(0)}-E_n^{(0)}}|\langle l|\hat{\mu}_p|n\rangle|^2 \qquad \text{(E.26b)}$$

Thus, the behavior of $\Delta\alpha_{pp,ln}$ for large values of $m \geq m_{crit}$ is governed by the first term,

$$\sum_{m \geq m_{crit}} \frac{2}{E_m^{(0)}} \left\{ \left| \langle l | \hat{\mu}_p | m \rangle \right|^2 - \left| \langle n | \hat{\mu}_p | m \rangle \right|^2 \right\} \tag{E.27a}$$

where $\langle l | \hat{\mu}_p | m \rangle := \mu_{p,lm}$, and Eq. E.27a can be written

$$\sum_{m \geq m_{crit}} \frac{2}{E_m^{(0)}} \left\{ \mu_{p,lm}^2 - \mu_{p,nm}^2 \right\} \tag{E.27b}$$

Appendix F

Applicable Constants and Conversion Factors[1]

The following constants and conversions are applicable to this work.

Constants

Electron charge (e) = $1.602176487 \times 10^{-19}$ Coulombs (C)

Planck's constant (h) = $6.62606896 \times 10^{-34}$ Joule-secs (J-s)

Hartree (E_h; energy atomic unit) = $4.3597439 \times 10^{-18}$ J

Speed of light in vacuum (c) = 2.99792458×10^8 m / s

Conversions

1 Electron volt (eV) = $1.60217649 \times 10^{-19}$ J

1 Hartree = 27.2114 eV = 219474.63 cm^{-1}

1 atomic unit of electric field = 5.142206×10^{11} V / m (volts per meter)

1 atomic unit of electric polarizability = 1.648777×10^{-41} C^2-m^2 / J

1 atomic unit of electric dipole moment = 1 Bohr-electron = 8.478352×10^{-30} C-m = 2.541746 Debye

Transition energies in nm = $1/ cm^{-1}$ x (1×10^7)

References

1.   P. J. Mohr, B. N. Taylor and D. B. Newell, Reviews of Modern Physics **80** (2), 633-730 (2008).

2.   http://www.gaussian.com/g_tech/g_ur/k_constants.htm

Appendix G

Derivation of Perturbation Terms in

Quantum-Mechanical Stark Analysis

The perturbation ranking terms used in Chapter 7 to develop a protocol for determining the most influential electronic states in quantum-mechanical Stark analysis are derived here.

The shift in transition frequency $\Delta\mu_{ij}$ between two states *i* and *j* in the presence of an electric field $\vec{E}$:

$$h\Delta v_{ij} = -\vec{E}\cdot\Delta\vec{\mu}_{ij} - \frac{1}{2}\vec{E}\cdot\Delta\vec{\alpha}_{ij}\cdot\vec{E} - \ldots \tag{G.1}$$

Electric field perturbation in the point dipole approximation

$$\widehat{H} = \widehat{H}^{(0)} - \vec{E}\cdot\widehat{\vec{\mu}} \tag{G.2}$$

and beyond:

$$\widehat{H} = \widehat{H}^{(0)} - \vec{E}\cdot\widehat{\vec{\mu}} - \frac{1}{6}\sum_{i=x,y,z}\sum_{j=x,y,z}\widehat{Q}_{ij}\frac{\partial E_j}{\partial x_i} - \ldots \tag{G.3}$$

In the point dipole approximation, the effect of an electric field on the energy of an unperturbed, non-degenerate energy level $\varepsilon_n^{(0)}$ can be expressed as an infinite sum of (decreasing) correction terms:

$$\varepsilon_n = \varepsilon_n^{(0)} + \varepsilon_n^{(1)} + \varepsilon_n^{(2)} + \varepsilon_n^{(3)} + \varepsilon_n^{(4)} + \rightleftharpoons, \tag{G.4}$$

According to (time-independent) perturbation theory, the first-, second-, third-, and fourth-order energy correction terms are

$$\varepsilon_n^{(1)} = -\vec{E} \cdot \vec{\mu}_n = -E_{\mu_n} \mu_n \tag{G.5}$$

$$\varepsilon_n^{(2)} = \sum_{p \neq n} \frac{\left| -\vec{E} \cdot \vec{\mu}_{np} \right|^2}{\varepsilon_n^{(0)} - \varepsilon_p^{(0)}} = \sum_{p \neq n} E_{\mu_{np}}^{2} \frac{\mu_{np}^{2}}{\varepsilon_n^{(0)} - \varepsilon_p^{(0)}} \tag{G.6}$$

$$\varepsilon_n^{(3)} = \sum_{p \neq n} \sum_{q \neq n} \frac{\left(-\vec{E} \cdot \vec{\mu}_{np} \right)\left(-\vec{E} \cdot \vec{\mu}_{pq} \right)\left(-\vec{E} \cdot \vec{\mu}_{qn} \right)}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)} \right)} - \left(-\vec{E} \cdot \vec{\mu}_n \right) \sum_{p \neq n} \frac{\left| -\vec{E} \cdot \vec{\mu}_{np} \right|^2}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)^2}$$

$$= -\sum_{p \neq n} \sum_{q \neq n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qn}} \frac{\mu_{np} \mu_{pq} \mu_{qn}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)} \right)} + E_{\mu_n} \mu_n \sum_{p \neq n} E_{\mu_{np}}^{2} \frac{\mu_{np}^{2}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)^2} \tag{G.7}$$

$$\varepsilon_n^{(4)} = \sum_{p \neq n} \sum_{q \neq n} \sum_{r \neq n} \frac{\left(-\vec{E} \cdot \vec{\mu}_{np} \right)\left(-\vec{E} \cdot \vec{\mu}_{pq} \right)\left(-\vec{E} \cdot \vec{\mu}_{qr} \right)\left(-\vec{E} \cdot \vec{\mu}_{rn} \right)}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)} \right)\left(\varepsilon_n^{(0)} - \varepsilon_r^{(0)} \right)} - \left(\sum_{p \neq n} \frac{\left| -\vec{E} \cdot \vec{\mu}_{np} \right|^2}{\varepsilon_n^{(0)} - \varepsilon_p^{(0)}} \right) \left(\sum_{p \neq n} \frac{\left| -\vec{E} \cdot \vec{\mu}_{np} \right|^2}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)^2} \right)$$

$$-2\left(-\vec{E} \cdot \vec{\mu}_n \right) \sum_{p \neq n} \sum_{q \neq n} \frac{\left(-\vec{E} \cdot \vec{\mu}_{np} \right)\left(-\vec{E} \cdot \vec{\mu}_{pq} \right)\left(-\vec{E} \cdot \vec{\mu}_{qn} \right)}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)^2 \left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)} \right)} + \left(-\vec{E} \cdot \vec{\mu}_n \right)^2 \sum_{p \neq n} \frac{\left| -\vec{E} \cdot \vec{\mu}_{np} \right|^2}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)^3}$$

$$= \sum_{p \neq n} \sum_{q \neq n} \sum_{r \neq n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qr}} E_{\mu_{rn}} \frac{\mu_{np} \mu_{pq} \mu_{qr} \mu_{rn}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)} \right)\left(\varepsilon_n^{(0)} - \varepsilon_r^{(0)} \right)} \tag{G.8}$$

$$-\sum_{p \neq n} \sum_{q \neq n} E_{\mu_{np}}^{2} E_{\mu_{nq}}^{2} \frac{\mu_{np}^{2} \mu_{nq}^{2}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)} \right)^2}$$

$$-2 E_{\mu_n} \mu_n \sum_{p \neq n} \sum_{q \neq n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qn}} \frac{\mu_{np} \mu_{pq} \mu_{qn}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)^2 \left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)} \right)} + E_{\mu_n}^{2} \mu_n^{2} \sum_{p \neq n} E_{\mu_{np}}^{2} \frac{\mu_{np}^{2}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)} \right)^3}$$

Infinite-order perturbation theory is equivalent to the numerical diagonalization of the perturbed Hamiltonian containing an infinite number of states ($N \to \infty$):

$$\hat{H} = \begin{pmatrix} \varepsilon_0^{(0)} - \vec{E} \cdot \vec{\mu}_0 & -\vec{E} \cdot \vec{\mu}_{01} & -\vec{E} \cdot \vec{\mu}_{02} & \cdots & -\vec{E} \cdot \vec{\mu}_{0N} \\ -\vec{E} \cdot \vec{\mu}_{10} & \varepsilon_1^{(0)} - \vec{E} \cdot \vec{\mu}_1 & -\vec{E} \cdot \vec{\mu}_{12} & & -\vec{E} \cdot \vec{\mu}_{1N} \\ -\vec{E} \cdot \vec{\mu}_{20} & -\vec{E} \cdot \vec{\mu}_{21} & \varepsilon_2^{(0)} - \vec{E} \cdot \vec{\mu}_2 & & -\vec{E} \cdot \vec{\mu}_{2N} \\ \vdots & & & \ddots & \\ -\vec{E} \cdot \vec{\mu}_{N0} & -\vec{E} \cdot \vec{\mu}_{N1} & -\vec{E} \cdot \vec{\mu}_{N2} & & \varepsilon_N^{(0)} - \vec{E} \cdot \vec{\mu}_N \end{pmatrix} \qquad (G.9)$$

For a centrosymmetric system like porphin that exhibits no permanent dipole moments ($\mu_i = 0$), the energy corrections according to perturbation theory reduce to:

$$\varepsilon_n^{(1)} = 0 \qquad (G.10)$$

$$\varepsilon_n^{(2)} = \sum_{p \neq n} E_{\mu_{np}}^2 \frac{\mu_{np}^2}{\varepsilon_n^{(0)} - \varepsilon_p^{(0)}} \qquad (G.11)$$

$$\varepsilon_n^{(3)} = \sum_{p \neq n} \sum_{q \neq n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qn}} \frac{\mu_{np} \mu_{pq} \mu_{qn}}{(\varepsilon_n^{(0)} - \varepsilon_p^{(0)})(\varepsilon_n^{(0)} - \varepsilon_q^{(0)})} \qquad (G.12)$$

$$\varepsilon_n^{(4)} = \sum_{p \neq n} \sum_{q \neq n} \sum_{r \neq n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qr}} E_{\mu_{rn}} \frac{\mu_{np} \mu_{pq} \mu_{qr} \mu_{rn}}{(\varepsilon_n^{(0)} - \varepsilon_p^{(0)})(\varepsilon_n^{(0)} - \varepsilon_q^{(0)})(\varepsilon_n^{(0)} - \varepsilon_r^{(0)})} -$$

$$\sum_{p \neq n} \sum_{q \neq n} E_{\mu_{np}}^2 E_{\mu_{nq}}^2 \frac{\mu_{np}^2 \mu_{nq}^2}{(\varepsilon_n^{(0)} - \varepsilon_p^{(0)})(\varepsilon_n^{(0)} - \varepsilon_q^{(0)})} \qquad (G.13)$$

Using only the first four excited states of porphin, this can be written as

$$\hat{H} = \begin{pmatrix} 0 & -E_x \mu_{Q_x} & -E_y \mu_{Q_y} & -E_x \mu_{B_x} & -E_y \mu_{B_y} \\ -E_x \mu_{Q_x} & \varepsilon_{Q_x}^{(0)} & -\vec{E} \cdot \vec{\mu}_{Q_x Q_y} & -\vec{E} \cdot \vec{\mu}_{Q_x B_x} & -\vec{E} \cdot \vec{\mu}_{Q_x B_y} \\ -E_y \mu_{Q_y} & -\vec{E} \cdot \vec{\mu}_{Q_y Q_x} & \varepsilon_{Q_y}^{(0)} & -\vec{E} \cdot \vec{\mu}_{Q_y B_x} & -\vec{E} \cdot \vec{\mu}_{Q_y B_y} \\ -E_x \mu_{B_x} & -\vec{E} \cdot \vec{\mu}_{B_x Q_x} & -\vec{E} \cdot \vec{\mu}_{B_x Q_y} & \varepsilon_{B_x}^{(0)} & -\vec{E} \cdot \vec{\mu}_{B_x B_y} \\ -E_y \mu_{B_y} & -\vec{E} \cdot \vec{\mu}_{B_y Q_x} & -\vec{E} \cdot \vec{\mu}_{B_y Q_y} & -\vec{E} \cdot \vec{\mu}_{B_y B_x} & \varepsilon_{B_y}^{(0)} \end{pmatrix} \qquad (G.14)$$

First level correction to Hamiltonian: addition of $-\vec{E}\cdot\vec{\mu}$ terms on diagonal according to their contribution to $\left|\varepsilon_1^{(1)}-\varepsilon_0^{(1)}\right|$, all of which are zero (this gives the unperturbed Hamiltonian in our case):

$$\widehat{H}^{(1)} = \begin{pmatrix} \varepsilon_0^{(0)}-\vec{E}\cdot\vec{\mu}_0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \varepsilon_1^{(0)}-\vec{E}\cdot\vec{\mu}_1 & 0 & 0 & & 0 \\ 0 & 0 & \varepsilon_{A_2}^{(0)}-\vec{E}\cdot\vec{\mu}_{A_2} & 0 & & 0 \\ 0 & 0 & 0 & \varepsilon_{A_3}^{(0)}-\vec{E}\cdot\vec{\mu}_{A_3} & & 0 \\ \vdots & & & & \ddots & \\ 0 & 0 & 0 & 0 & & \varepsilon_{A_N}^{(0)}-\vec{E}\cdot\vec{\mu}_{A_N} \end{pmatrix}$$

$$= \begin{pmatrix} \varepsilon_0^{(0)} & 0 & 0 & 0 & \cdots & 0 \\ 0 & \varepsilon_1^{(0)} & 0 & 0 & & 0 \\ 0 & 0 & \varepsilon_{A_2}^{(0)} & 0 & & 0 \\ 0 & 0 & 0 & \varepsilon_{A_3}^{(0)} & & 0 \\ \vdots & & & & \ddots & \\ 0 & 0 & 0 & 0 & & \varepsilon_{A_N}^{(0)} \end{pmatrix}$$

(G.15)

Second level correction to Hamiltonian (direct coupling, or better: **single-state coupling**):

$$\hat{H}^{(2)} = \begin{pmatrix} \varepsilon_0^{(0)} & -E_x\mu_{01} & -\vec{E}\cdot\vec{\mu}_{0A_2} & -\vec{E}\cdot\vec{\mu}_{0A_3} & \cdots & -\vec{E}\cdot\vec{\mu}_{0A_N} \\[6pt] -E_x\mu_{10} & \varepsilon_1^{(0)} & -\vec{E}\cdot\vec{\mu}_{1A_2} & -\vec{E}\cdot\vec{\mu}_{1A_3} & & -\vec{E}\cdot\vec{\mu}_{1A_N} \\[6pt] -\vec{E}\cdot\vec{\mu}_{A_20} & -\vec{E}\cdot\vec{\mu}_{A_21} & \varepsilon_{A_2}^{(0)} & 0 & & 0 \\[6pt] -\vec{E}\cdot\vec{\mu}_{A_30} & -\vec{E}\cdot\vec{\mu}_{A_31} & 0 & \varepsilon_{A_3}^{(0)} & & 0 \\[6pt] \vdots & & & & \ddots & \\[6pt] -\vec{E}\cdot\vec{\mu}_{A_N0} & -\vec{E}\cdot\vec{\mu}_{A_N0} & 0 & 0 & & \varepsilon_{A_N}^{(0)} \end{pmatrix} \qquad \text{(G.16)}$$

For single-state coupling, states A_i are successively added to the matrix according to the contribution *of single sums* to $\left| \varepsilon_1^{(2)} - \varepsilon_0^{(2)} + \varepsilon_1^{(3)} - \varepsilon_0^{(3)} + - \rightleftharpoons \right|$, which scales, for arbitrary electric fields, according to

$$\left| \frac{\mu_{0A_i}^2}{\varepsilon_{A_i}^{(0)} - \varepsilon_0^{(0)}} - \frac{\mu_{1A_i}^2}{\varepsilon_{A_i}^{(0)} - \varepsilon_1^{(0)}} + 2\frac{E_{\mu_{01}}\mu_{01}}{\varepsilon_1^{(0)} - \varepsilon_0^{(0)}}\left(\frac{\mu_{0A_i}\mu_{1A_i}}{\varepsilon_{A_i}^{(0)} - \varepsilon_0^{(0)}} + \frac{\mu_{0A_i}\mu_{1A_i}}{\varepsilon_{A_i}^{(0)} - \varepsilon_1^{(0)}} \right) + \ldots \right| \simeq \left| \frac{\mu_{0A_i}^2}{\varepsilon_{A_i}^{(0)} - \varepsilon_0^{(0)}} - \frac{\mu_{1A_i}^2}{\varepsilon_{A_i}^{(0)} - \varepsilon_1^{(0)}} \right|$$

$$\text{(G.17)}$$

$$\begin{aligned}
\varepsilon_1^{(2)} - \varepsilon_0^{(2)} &= \sum_{p\neq 1} E_{\mu_{1p}}{}^2 \frac{\mu_{1p}^2}{\varepsilon_1^{(0)} - \varepsilon_p^{(0)}} - \sum_{p\neq 0} E_{\mu_{0p}}{}^2 \frac{\mu_{0p}^2}{\varepsilon_0^{(0)} - \varepsilon_p^{(0)}} \\[6pt]
&= E_{\mu_{10}}{}^2 \frac{\mu_{10}^2}{\varepsilon_1^{(0)} - \varepsilon_0^{(0)}} + \sum_{p\neq 0,1} E_{\mu_{1p}}{}^2 \frac{\mu_{1p}^2}{\varepsilon_1^{(0)} - \varepsilon_p^{(0)}} \\[6pt]
&\quad - \left(E_{\mu_{01}}{}^2 \frac{\mu_{01}^2}{\varepsilon_0^{(0)} - \varepsilon_1^{(0)}} + \sum_{p\neq 0,1} E_{\mu_{0p}}{}^2 \frac{\mu_{0p}^2}{\varepsilon_0^{(0)} - \varepsilon_p^{(0)}} \right) \\[6pt]
&= 2E_{\mu_{01}}{}^2 \frac{\mu_{01}^2}{\varepsilon_1^{(0)} - \varepsilon_0^{(0)}} + \sum_{p\neq 0,1} \left(E_{\mu_{1p}}{}^2 \frac{\mu_{1p}^2}{\varepsilon_1^{(0)} - \varepsilon_p^{(0)}} - E_{\mu_{0p}}{}^2 \frac{\mu_{0p}^2}{\varepsilon_0^{(0)} - \varepsilon_p^{(0)}} \right) \\[6pt]
&= 2E_{\mu_{01}}{}^2 \frac{\mu_{01}^2}{\varepsilon_1^{(0)} - \varepsilon_0^{(0)}} + \sum_{p\neq 0,1} \left(E_{\mu_{0p}}{}^2 \frac{\mu_{0p}^2}{\varepsilon_p^{(0)} - \varepsilon_0^{(0)}} - E_{\mu_{1p}}{}^2 \frac{\mu_{1p}^2}{\varepsilon_p^{(0)} - \varepsilon_1^{(0)}} \right)
\end{aligned} \qquad \text{(G.18)}$$

Third level correction to Hamiltonian (indirect coupling between pairs of states A_i and A_j, or

better: **double-state coupling**):

$$
\widehat{H}^{(3)} = \begin{pmatrix}
\varepsilon_0^{(0)} & -E_x\mu_{01} & \cdots & -\vec{E}\cdot\vec{\mu}_{0A_i} & \cdots & -\vec{E}\cdot\vec{\mu}_{0A_j} & \cdots & -\vec{E}\cdot\vec{\mu}_{0A_N} \\
-E_x\mu_{10} & \varepsilon_1^{(0)} & & -\vec{E}\cdot\vec{\mu}_{1A_i} & & -\vec{E}\cdot\vec{\mu}_{1A_j} & & -\vec{E}\cdot\vec{\mu}_{1A_N} \\
\vdots & & \ddots & & & & & \\
-\vec{E}\cdot\vec{\mu}_{A_i0} & -\vec{E}\cdot\vec{\mu}_{A_i1} & & \varepsilon_{A_i}^{(0)} & & -\vec{E}\cdot\vec{\mu}_{A_iA_j} & & 0 \\
\vdots & & & & \ddots & & & \\
-\vec{E}\cdot\vec{\mu}_{A_j0} & -\vec{E}\cdot\vec{\mu}_{A_j1} & & -\vec{E}\cdot\vec{\mu}_{A_jA_i} & & \varepsilon_{A_j}^{(0)} & & 0 \\
\vdots & & & & & & \ddots & \\
-\vec{E}\cdot\vec{\mu}_{A_N0} & -\vec{E}\cdot\vec{\mu}_{A_N1} & & 0 & & 0 & & \varepsilon_{A_N}^{(0)}
\end{pmatrix}
$$

For double-state coupling, pairs $\left(A_i, A_j\right)$ are successively added according to the contribution *of*

double sums to $\left| \varepsilon_1^{(3)} - \varepsilon_0^{(3)} + \varepsilon_1^{(4)} - \varepsilon_0^{(4)} + -\rightleftharpoons \right|$, which scales, for arbitrary electric fields, according

to

$$
\left| \mu_{A_iA_j} \left(\frac{\mu_{0A_i}\mu_{0A_j}}{\left(\varepsilon_{A_i}^{(0)} - \varepsilon_0^{(0)}\right)\left(\varepsilon_{A_j}^{(0)} - \varepsilon_0^{(0)}\right)} - \frac{\mu_{1A_i}\mu_{1A_j}}{\left(\varepsilon_{A_i}^{(0)} - \varepsilon_1^{(0)}\right)\left(\varepsilon_{A_j}^{(0)} - \varepsilon_1^{(0)}\right)} \right) + \rightleftharpoons \right| \tag{G.20}
$$

(Double-sum contributions from higher-order perturbation terms are again neglected

compared to the double-sum contribution from the third-order perturbation term given above.)

$$\varepsilon_n^{(3)} = -\sum_{p\neq n}\sum_{q\neq n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qn}} \frac{\mu_{np}\mu_{pq}\mu_{qn}}{\left(\varepsilon_n^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)}-\varepsilon_q^{(0)}\right)}$$

$$= -\sum_{p\neq n}\left(E_{\mu_{np}} E_{\mu_{pm}} E_{\mu_{mn}} \frac{\mu_{np}\mu_{pm}\mu_{mn}}{\left(\varepsilon_n^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)}-\varepsilon_m^{(0)}\right)} + \sum_{q\neq m,n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qn}} \frac{\mu_{np}\mu_{pq}\mu_{qn}}{\left(\varepsilon_n^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)}-\varepsilon_q^{(0)}\right)} \right)$$

$$= -\left(E_{\mu_m}\mu_m E_{\mu_{mn}}{}^2 \frac{\mu_{mn}{}^2}{\left(\varepsilon_n^{(0)}-\varepsilon_m^{(0)}\right)^2} + \sum_{q\neq m,n} E_{\mu_{nm}} E_{\mu_{mq}} E_{\mu_{qn}} \frac{\mu_{nm}\mu_{mq}\mu_{qn}}{\left(\varepsilon_n^{(0)}-\varepsilon_m^{(0)}\right)\left(\varepsilon_n^{(0)}-\varepsilon_q^{(0)}\right)} \right) \qquad \text{(G.21)}$$

$$- \sum_{p\neq m,n}\left(E_{\mu_{np}} E_{\mu_{pm}} E_{\mu_{mn}} \frac{\mu_{np}\mu_{pm}\mu_{mn}}{\left(\varepsilon_n^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)}-\varepsilon_m^{(0)}\right)} + \sum_{q\neq m,n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qn}} \frac{\mu_{np}\mu_{pq}\mu_{qn}}{\left(\varepsilon_n^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)}-\varepsilon_q^{(0)}\right)} \right)$$

$$= -2 E_{\mu_{nm}} \frac{\mu_{nm}}{\varepsilon_n^{(0)}-\varepsilon_m^{(0)}} \sum_{q\neq m,n} E_{\mu_{mq}} E_{\mu_{qn}} \frac{\mu_{mq}\mu_{qn}}{\varepsilon_n^{(0)}-\varepsilon_q^{(0)}} - \sum_{p\neq m,n}\sum_{q\neq m,n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qn}} \frac{\mu_{np}\mu_{pq}\mu_{qn}}{\left(\varepsilon_n^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)}-\varepsilon_q^{(0)}\right)}$$

$$\varepsilon_1^{(3)} - \varepsilon_0^{(3)} = -2 E_{\mu_{10}} \frac{\mu_{10}}{\varepsilon_1^{(0)}-\varepsilon_0^{(0)}} \sum_{q\neq 0,1} E_{\mu_{0q}} E_{\mu_{q1}} \frac{\mu_{0q}\mu_{q1}}{\varepsilon_1^{(0)}-\varepsilon_q^{(0)}} - \sum_{p\neq 0,1}\sum_{q\neq 0,1} E_{\mu_{1p}} E_{\mu_{pq}} E_{\mu_{q1}} \frac{\mu_{1p}\mu_{pq}\mu_{q1}}{\left(\varepsilon_1^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_1^{(0)}-\varepsilon_q^{(0)}\right)}$$

$$+ 2 E_{\mu_{01}} \frac{\mu_{01}}{\varepsilon_0^{(0)}-\varepsilon_1^{(0)}} \sum_{q\neq 1,0} E_{\mu_{1q}} E_{\mu_{q0}} \frac{\mu_{1q}\mu_{q0}}{\varepsilon_0^{(0)}-\varepsilon_q^{(0)}} + \sum_{p\neq 1,0}\sum_{q\neq 1,0} E_{\mu_{0p}} E_{\mu_{pq}} E_{\mu_{q0}} \frac{\mu_{0p}\mu_{pq}\mu_{q0}}{\left(\varepsilon_0^{(0)}-\varepsilon_p^{(0)}\right)\left(\varepsilon_0^{(0)}-\varepsilon_q^{(0)}\right)}$$

$$= +2 E_{\mu_{01}} \frac{\mu_{01}}{\varepsilon_1^{(0)}-\varepsilon_0^{(0)}} \sum_{q\neq 0,1} E_{\mu_{0q}} E_{\mu_{1q}} \frac{\mu_{0q}\mu_{1q}}{\varepsilon_q^{(0)}-\varepsilon_1^{(0)}} - \sum_{p\neq 0,1}\sum_{q\neq 0,1} E_{\mu_{1p}} E_{\mu_{pq}} E_{\mu_{q1}} \frac{\mu_{1p}\mu_{pq}\mu_{q1}}{\left(\varepsilon_p^{(0)}-\varepsilon_1^{(0)}\right)\left(\varepsilon_q^{(0)}-\varepsilon_1^{(0)}\right)} \qquad \text{(G.22)}$$

$$+ 2 E_{\mu_{01}} \frac{\mu_{01}}{\varepsilon_1^{(0)}-\varepsilon_0^{(0)}} \sum_{q\neq 0,1} E_{\mu_{0q}} E_{\mu_{1q}} \frac{\mu_{q0}\mu_{1q}}{\varepsilon_q^{(0)}-\varepsilon_0^{(0)}} + \sum_{p\neq 0,1}\sum_{q\neq 0,1} E_{\mu_{0p}} E_{\mu_{pq}} E_{\mu_{q0}} \frac{\mu_{0p}\mu_{pq}\mu_{q0}}{\left(\varepsilon_p^{(0)}-\varepsilon_0^{(0)}\right)\left(\varepsilon_q^{(0)}-\varepsilon_0^{(0)}\right)}$$

$$= 2 E_{\mu_{01}} \frac{\mu_{01}}{\varepsilon_1^{(0)}-\varepsilon_0^{(0)}} \sum_{p\neq 0,1}\left[E_{\mu_{0p}} E_{\mu_{1p}} \left(\frac{\mu_{0p}\mu_{1p}}{\varepsilon_p^{(0)}-\varepsilon_0^{(0)}} + \frac{\mu_{0p}\mu_{1p}}{\varepsilon_p^{(0)}-\varepsilon_1^{(0)}} \right) \right]$$

$$+ \sum_{p\neq 0,1}\sum_{q\neq 0,1}\left[E_{\mu_{pq}}\mu_{pq} \left(E_{\mu_{0p}} E_{\mu_{0q}} \frac{\mu_{0p}\mu_{0q}}{\left(\varepsilon_p^{(0)}-\varepsilon_0^{(0)}\right)\left(\varepsilon_q^{(0)}-\varepsilon_0^{(0)}\right)} - E_{\mu_{1p}} E_{\mu_{1q}} \frac{\mu_{1p}\mu_{1q}}{\left(\varepsilon_p^{(0)}-\varepsilon_1^{(0)}\right)\left(\varepsilon_q^{(0)}-\varepsilon_1^{(0)}\right)} \right) \right]$$

Fourth level correction to Hamiltonian (indirect coupling between triples $\left(A_i, A_j, A_k\right)$ or better:

triple-state coupling):

For triple-state coupling, triples $\left(A_i, A_j, A_k\right)$ are successively added according to the

contribution *of triple sums* to $\left| \varepsilon_1^{(4)} - \varepsilon_0^{(4)} + \varepsilon_1^{(5)} - \varepsilon_0^{(5)} + - \rightleftharpoons \right|$.

$$\varepsilon_n^{(4)} = \sum_{p \neq n} \sum_{q \neq n} \sum_{r \neq n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qr}} E_{\mu_{rn}} \frac{\mu_{np} \mu_{pq} \mu_{qr} \mu_{rn}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_r^{(0)}\right)}$$

$$- \sum_{p \neq n} \sum_{q \neq n} E_{\mu_{np}}^{\ 2} E_{\mu_{nq}}^{\ 2} \frac{\mu_{np}^{\ 2} \mu_{nq}^{\ 2}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)}\right)^2}$$

$$= \sum_{p \neq n} \sum_{q \neq n} \Bigg(E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qm}} E_{\mu_{mn}} \frac{\mu_{np} \mu_{pq} \mu_{qm} \mu_{mn}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_m^{(0)}\right)} \qquad\text{(G.23)}$$

$$+ \sum_{r \neq m,n} E_{\mu_{np}} E_{\mu_{pq}} E_{\mu_{qr}} E_{\mu_{rn}} \frac{\mu_{np} \mu_{pq} \mu_{qr} \mu_{rn}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_r^{(0)}\right)} \Bigg)$$

$$- \sum_{p \neq n} \Bigg(E_{\mu_{np}}^{\ 2} E_{\mu_{nm}}^{\ 2} \frac{\mu_{np}^{\ 2} \mu_{nm}^{\ 2}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_m^{(0)}\right)^2} + \sum_{q \neq m,n} E_{\mu_{np}}^{\ 2} E_{\mu_{nq}}^{\ 2} \frac{\mu_{np}^{\ 2} \mu_{nq}^{\ 2}}{\left(\varepsilon_n^{(0)} - \varepsilon_p^{(0)}\right)\left(\varepsilon_n^{(0)} - \varepsilon_q^{(0)}\right)^2} \Bigg)$$

Appendix H

35 Essential States for QM Stark Analysis of Porphin (CIS / sdd)

Table H.1. List of 35 essential states for quantum-mechanical Stark analysis of free-base porphin determined in Chapter 7 calculated with CIS / sdd.

Transition to State	Delta alpha	Wavenumbers	Transition from State	Transition Dipole (D)	Transition Polarization	Transition from State	Transition Dipole (D)	Transition Polarization
4	74.9232	37361	0	9.5648	X	1	1.8557	Y
3	71.0511	37342	0	9.4876	Y	1	2.1239	X
12	47.0750	51567	0	0.6761	Y	1	6.2888	X
5	45.3433	41380	0	-7.4905	Y	1	0.6212	X
10	38.8699	49783	0	-4.1575	X	1	6.2957	Y
16	36.7725	55393	0	-0.9605	Y	1	5.9477	X
6	36.2095	42360	0	7.3319	X	1	1.9729	Y
17	31.1462	55546	0	-0.6776	X	1	-5.4696	Y
15	8.6389	54605	0	-3.7326	Y	1	-0.1634	X
63	7.3871	83022	0	4.249	Y	1	0.0366	X
26	5.9684	66752	0	3.5419	X	1	-0.7305	Y
23	4.4022	63796	0	0.9237	Y	1	-2.4088	X
34	4.2242	72734	0	3.0107	X	1	0.1182	Y
13	4.2114	52351	0	3.1317	X	1	1.3591	Y
49	3.3296	78826	0	-2.7817	X	1	0.093	Y
42	3.1714	75536	0	-2.7446	Y	1	0.5775	X
75	2.3743	85147	0	2.4401	X	1	-0.0475	Y
85	2.3142	87699	0	-2.4469	Y	1	-0.1001	X
22	2.0791	63421	0	2.0715	X	1	0.5099	Y
55	1.8260	80292	0	2.1081	Y	1	-0.3022	X
57	1.7391	82000	0	-2.1038	Y	1	0.4046	X
64	1.6071	83172	0	-2.0021	X	1	0.231	Y
97	1.5936	90096	0	-2.0621	X	1	0.1395	Y
25	1.4615	66683	0	1.7261	Y	1	0.2687	X
11	1.3987	51300	0	-1.7612	X	1	1.6898	Y
14	1.2267	52549	0	0.6365	Y	1	-1.1339	X
56	1.1010	81244	0	-1.6227	Z	1	0	0
2	0.9008	27575	0	0.9811	X	1	0.518	Y
94	0.8988	89135	0	1.5883	X	1	0.349	Y
48	0.8072	77412	0	0.7971	X	1	1.3164	Y
33	0.7780	72109	0	-0.5195	X	1	-1.1415	Y
47	0.7345	77332	0	-0.7933	Y	1	-1.2691	X
68	0.7223	83947	0	-1.3499	X	1	0.1645	Y
78	0.6087	85802	0	1.312	X	1	0.3665	Y

Appendix I

Derivation of Oscillator Strength with Perturbation Theory

Starting with time-dependent perturbation theory, the perturbation continues with adiabatic turn-on as

$$C_m^{(1)}(t) = \delta_{ml} + \frac{A_{ml}}{i\hbar}\frac{e^{i(\omega_{ml}-\omega)t}}{i(\omega_{ml}-\omega)} + \frac{A_{ml}^*}{i\hbar}\frac{e^{i(\omega_{ml}-\omega)t}}{i(\omega_{ml}-\omega)} \qquad (I.1)$$

where the initial condition is $C_m(t = -\infty) = \delta_{ml}$.

The time-dependent wavefunction,

$$|\psi(t)\rangle = \sum_m C_m(t)e^{-\frac{i}{\hbar}E_m t}|m\rangle \qquad (I.2)$$

with substitution from Eq. I.1 becomes

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}E_l t}\left\{|l\rangle + \sum_{m\neq l}\left[\frac{A_{ml}}{i\hbar}\frac{e^{-i\omega t}}{i(\omega_{ml}-\omega)} + \frac{A_{ml}^*}{i\hbar}\frac{e^{i\omega t}}{i(\omega_{ml}+\omega)}\right]|m\rangle\right\} \qquad (I.3)$$

with the term in square brackets representing the perturbed part of the wavefunction oscillating with the frequency of the perturbation.

If $|\omega| \ll |\omega_{ml}|$ (where $\omega = 2\pi\nu$ and $\omega_{ml} = \frac{(E_m^{(0)}-E_l^{(0)})}{\hbar}$)

$$|\psi(t)\rangle > \cong e^{-\frac{i}{\hbar}E_l t}\left\{|l\rangle + \sum_{m\neq l}\frac{H_{1,ml}(t)}{E_l - E_m}|m\rangle\right\} \qquad (I.4)$$

in the limit $\langle m|H_1(t)|l\rangle \cong \langle m|Ae^{-i\omega t}|l\rangle$, where $H_1(t) \equiv e^{-i\omega t}A_{ml}$.

I.1 Classical Theory of Dispersion

Electrodynamics gives the Maxwell relation

$$n^2 = \varepsilon \tag{I.5}$$

where n is refractive index and ε is dielectric constant. Then, the electric displacement, $\vec{D}$, is

$$\vec{D} = \varepsilon\vec{E} = \vec{E} + 4\pi\vec{p} = (1 + 4\pi\chi)\vec{E} \tag{I.6}$$

where $\vec{E}$ is an electric field, $\vec{p}$ is the polarization, and χ is the electric susceptibility.

Thus

$$n^2 = 1 + 4\pi\chi \tag{I.7}$$

and polarization is

$$\vec{p} = \chi\vec{E} = N\alpha\vec{E} \tag{I.8}$$

Then

$$\vec{\rho} = \alpha\vec{E} \tag{I.9}$$

which is the dipole moment, $\vec{\rho}$, of an atom induced by the field $\vec{E}$ and where α is the polarizability and N is the number density.

Now

$$n^2 = 1 + 4\pi N\alpha \tag{I.10}$$

Representing an atom with a series of harmonic oscillators (classical model), the equation of motion of an oscillator, μ, when driven with no damping is

$$m\ddot{x}_\mu + m\omega_\mu^2 x_\mu = eE_0 e^{i\omega t} \qquad (I.11)$$

Then rearranging for the solution to the oscillator

$$x_\mu = \frac{e}{m(\omega_\mu^2 - \omega^2)} E_0 e^{i\omega t} \qquad (I.12)$$

The dipole moment for the oscillator is

$$\vec{p}_\mu = e\vec{x}_\mu = \alpha_\mu E \qquad (I.13)$$

where e is the elementary charge and x is position. Substituting Eq. I.12 for $\vec{x}_\mu$ into that for dipole moment in Eq. I.13,

$$e\frac{eE_0 e^{i\omega t}}{m(\omega_\mu^2 - \omega^2)} = \alpha_\mu E = \alpha_\mu E_0 e^{i\omega t} \qquad (I.14)$$

Polarizability, α, is then defined as

$$\alpha_\mu = \frac{e^2}{m(\omega_\mu^2 - \omega^2)} \qquad (I.15)$$

Polarizability for a set of oscillators is then

$$\alpha = \sum_\mu \alpha_\mu f_\mu = \sum_\mu \frac{e^2}{m(\omega_\mu^2 - \omega^2)} f_\mu \qquad (I.16)$$

where an important factor arises, f_μ, the oscillator strength of oscillator μ as a weighting factor for the set.

I.2 Quantum-Mechanical Theory of Dispersion

Starting with the time-dependent, quantum-mechanical definition of dipole moment,

$$p(t) = \langle e \cdot x(t) \rangle = \langle \psi(t) | e \cdot \vec{x} | \psi(t) \rangle \qquad (I.17)$$

and

$$\psi(t) = e^{-\frac{i}{\hbar}E_l t}\{|l\rangle + \sum_{m \neq l} C_{ml}(t) |m\rangle\} \tag{I.18}$$

and

$$C_{ml}(t) = \frac{A_{ml}}{i\hbar} \frac{e^{-i\omega t}}{i(\omega_{ml}-\omega)} + \frac{A_{lm}^*}{i\hbar} \frac{e^{i\omega t}}{i(\omega_{ml}+\omega)} \tag{I.19}$$

and

$$A_{ml} = -F\langle m|e\vec{x}\cdot\vec{a}|l\rangle \tag{I.20}$$

where F is the amplitude of the electromagnetic field and $\vec{a}$ is the polarization vector. The perturbing field in the x-direction is

$$A_{ml} = -F\langle m|e\cdot x|l\rangle = -Fd_{ml} \tag{I.21}$$

where d is the dipole matrix element. Substituting the wavefunction into the QM definition of the dipole moment

$$p(t) = \langle|l\rangle + \sum_{m \neq l} C_{ml}^*(t) |m\rangle|e\cdot x||l\rangle + \sum_{m \neq l} C_{ml}(t) |m\rangle\rangle = \langle l|e\cdot x|l\rangle +$$

$$\sum_{m \neq l} C_{ml}(t)\langle l|e\cdot x|m\rangle + \sum_{m \neq l} C_{ml}^*(t)\langle m|e\cdot x|l\rangle \tag{I.22}$$

If the molecular system has no permanent dipole moment due to symmetry, then $\langle l|e\cdot x|l\rangle = 0$. Making substitutions from Eq. I.17–I.20:

$$p(t) = \sum_{m \neq l}\left\{\frac{d_{ml}}{\hbar}\frac{Fe^{-i\omega t}}{(\omega_{ml}-\omega)} + \frac{d_{lm}^*}{\hbar}\frac{F^*e^{i\omega t}}{(\omega_{ml}+\omega)}\right\}d_{lm} + \sum_{m \neq l}\left\{\frac{d_{ml}^*}{\hbar}\frac{F^*e^{i\omega t}}{(\omega_{ml}-\omega)} + \frac{d_{lm}}{\hbar}\frac{Fe^{i\omega t}}{(\omega_{ml}+\omega)}\right\}d_{ml} =$$

$$\sum_{m \neq l}\frac{|d_{ml}|^2}{\hbar}\left\{\frac{Fe^{-i\omega t}}{(\omega_{ml}-\omega)} + \frac{F^*e^{i\omega t}}{(\omega_{ml}+\omega)} + \frac{F^*e^{i\omega t}}{(\omega_{ml}-\omega)} + \frac{Fe^{i\omega t}}{(\omega_{ml}+\omega)}\right\} = \sum_{m \neq l}\frac{|d_{ml}|^2}{\hbar}\left\{Fe^{-i\omega t}\left(\frac{1}{(\omega_{ml}-\omega)} + \right.\right.$$

$$\frac{1}{(\omega_{ml}+\omega)} + F^* e^{i\omega t}\left(\frac{1}{(\omega_{ml}+\omega)} + \frac{1}{(\omega_{ml}-\omega)}\right) = \sum_{m \neq l} \frac{|d_{ml}|^2}{\hbar} \frac{2\omega_{m'l}}{\omega_{ml}^2 - \omega^2}\left\{Fe^{-i\omega t} + F^* e^{i\omega t}\right\} =$$

$$\sum_{m \neq l} \frac{e^2}{m(\omega_{ml}^2 - \omega^2)} \frac{2\omega_{m'l}}{\hbar}\left|x_{m'l}\right|^2 F(t) \tag{I.23}$$

because it is recognized that $F(t) \equiv Fe^{-i\omega t} + F^* e^{i\omega t}$ and $|d_{ml}|^2 = e^2\left|x_{m'l}\right|^2$.

In addition, because Eq. I.9 gives $\vec{\rho} = \alpha\vec{E}$, it is identified that

$$\alpha \equiv \sum_{m \neq l} \frac{e^2}{m(\omega_{ml}^2 - \omega^2)} \frac{2\omega_{m'l}}{\hbar}\left|x_{m'l}\right|^2 \tag{I.24}$$

Classically, we found that polarizability is also

$$\alpha = \sum_{\mu} \frac{e^2}{m(\omega_{\mu}^2 - \omega^2)} f_{\mu} \tag{I.25}$$

Comparing the two Eq. I.24 and I.25 and seeing that $\mu = m' l$,

$$\sum_{m \neq l} \frac{e^2}{m(\omega_{ml}^2 - \omega^2)} \frac{2\omega_{m'l}}{\hbar}\left|x_{m'l}\right|^2 = \sum_{m \neq l} \frac{e^2}{m(\omega_{m'l}^2 - \omega^2)} f_{m'l} \tag{I.26}$$

A definition for oscillator strength arises:

$$f_{m'l} = \frac{2m\omega_{m'l}}{\hbar}\left|x_{m'l}\right|^2 \tag{I.27}$$

Thus

$$\alpha = \frac{e^2}{m}\sum_{m' \neq l} \frac{f_{m'l}}{(\omega_{m'l}^2 - \omega^2)} \tag{I.28}$$

Then a definition of dipole moment

$$p(t) = \alpha F(t) = \frac{e^2}{m}\sum_{m' \neq l} \frac{f_{m'l}}{\left(\omega_{m'l}^2 - \omega^2\right)} F(t) \tag{I.29}$$

and one for refractive index

$$n^2 = 1 + 4\pi N\alpha = 1 + \frac{4\pi N}{m}e^2 \sum_{m' \neq l} \frac{f_{m'l}}{\left(\omega_{m'l}^2 - \omega^2\right)} \tag{I.30}$$

Now assuming $\omega \gg \omega_{m'l}$, i.e., $\omega_{m'l}^2 - \omega^2 \approx -\omega^2$, then

$$n^2 = 1 - \frac{4\pi Ne^2}{m\omega^2} \sum_{m' \neq l} f_{m'l} \tag{I.31}$$

Then, what is $\sum_{m' \neq l} f_{m'l}$? The *f*-sum rule arises. Starting with

$$f_{m'l} = \frac{2m\omega_{m'l}}{\hbar}\left|x_{m'l}\right|^2 \tag{I.32}$$

and defining $\omega_{m'l} = \frac{1}{\hbar}(E_{m'} - E_l)$,

$$f_{m'l} = 2\frac{m}{\hbar^2}(E_{m'} - E_l)\langle l|x|m'\rangle\langle m'|x|l\rangle = 2\frac{m}{\hbar^2}\{E_{m'}\langle l|x|m'\rangle\langle m'|x|l\rangle -$$

$$E_l\langle l|x|m'\rangle\langle m'|x|l\rangle\} = 2\frac{m}{\hbar^2}\{\langle l|xE_{m'}|m'\rangle\langle m'|x|l\rangle - \langle l|x|m'\rangle\langle m'|xE_l|l\rangle\}$$

$$\tag{I.33}$$

The Schrödinger equation gives us $E_{m'}|m'\rangle = \hat{H}|m'\rangle$ and $E_l|l\rangle = \hat{H}|l\rangle$, so now

$$f_{m'l} = 2\frac{m}{\hbar^2}\{\langle l|x\hat{H}|m'\rangle\langle m'|x|l\rangle - \langle l|x|m'\rangle\langle m'|x\hat{H}|l\rangle\} \tag{I.34}$$

The sum of oscillator strength is now

$$\sum_{m' \neq l} f_{m'l} = \sum_{m'} f_{m'l} = 2\frac{m}{\hbar^2}\{\langle l|x\hat{H}|m'\rangle\langle m'|x|l\rangle - \langle l|x|m'\rangle\langle m'|x\hat{H}|l\rangle\} \tag{I.35}$$

where *l* is not exempt in sum because there is no permanent dipole moment, i.e., $\langle l|p|l\rangle = 0$.

Now

$$\sum_{m'} f_{m'\,l} = 2\frac{m}{\hbar^2}\{\langle l|x\widehat{H}\sum_{m'}|m'\rangle\langle m'|x|l\rangle - \langle l|x\sum_{m'}|m'\rangle\langle m'|x\widehat{H}|l\rangle\}$$

(I.36)

By definition, $\sum_{m'}|m'\rangle\langle m'| = \sum_{m'}\langle m'|m'\rangle = 1$, so then

$$\sum_{m'} f_{m'\,l} = 2\frac{m}{\hbar^2}\{\langle l|x\widehat{H}x|l\rangle - \langle l|x^2\widehat{H}|l\rangle\} = 2\frac{m}{\hbar^2}\{\langle l|x\widehat{H}x - x^2H|l\rangle\}2\frac{m}{\hbar^2}\{\langle l|x(\widehat{H}x - x\widehat{H})|l\rangle\} =$$

$$2\frac{m}{\hbar^2}\{\langle l|x[\widehat{H},x]|l\rangle\}$$

(I.37)

This commutator of the Hamiltonian and position operators, $[\widehat{H},\hat{x}]$, is defined as

$$[\widehat{H},\hat{x}] = \left(\frac{\hat{p}^2}{2m}+V\right)\hat{x} - \hat{x}\left(\frac{\hat{p}^2}{2m}+V\right) = \frac{1}{2m}\{\hat{p}^2\hat{x} + V\hat{x} - \hat{x}\hat{p}^2 - \hat{x}V\} = \frac{1}{2m}[\hat{p}^2,\hat{x}] + \frac{1}{2m}(\hat{p}^2 \times$$

$$\hat{p}^2) - \frac{1}{2m}(\hat{p}^2\times\hat{p}^2) = \frac{1}{2m}\{\hat{p}^2\hat{x} - \hat{x}\hat{p}^2 + (\hat{p}\ \hat{x}\ \hat{p}\ ) - (\hat{p}\ \hat{x}\ \hat{p}\ )\} = \{\hat{p}(\hat{p}\hat{x} - \hat{x}\hat{p}) + (\hat{p}\hat{x} -$$

$$\hat{x}\hat{p})\hat{p}\}\frac{1}{2m} = \frac{1}{2m}\hat{p}[\hat{p},\hat{x}] + \frac{1}{2m}[\hat{p},\hat{x}]\hat{p}$$

(I.38)

The commutator of momentum operator and position is $[\hat{p},\hat{x}] = \frac{\hbar}{i}$, and then

$$[\widehat{H},\hat{x}] = \frac{1}{2m}\hat{p}\frac{\hbar}{i} + \frac{1}{2m}\frac{\hbar}{i}\hat{p} = \frac{1}{m}\frac{\hbar}{i}\hat{p}$$

(I.39)

Back to developing the sum of oscillator strength:

$$\sum_{m'} f_{m'\,l} = 2\frac{m}{\hbar^2}\left\langle l\left|x\cdot\frac{1}{m}\frac{\hbar}{i}\hat{p}\right|l\right\rangle = \frac{2}{i\hbar}\langle l|x\cdot\hat{p}|l\rangle = \frac{1}{i\hbar}\{\langle l|x\cdot\hat{p}|l\rangle + \langle l|x\cdot\hat{p}|l\rangle\} =$$

$$\frac{1}{i\hbar}\{\langle l|x\hat{p} - \hat{p}x + x\hat{p} + \hat{p}x|l\rangle\}$$

(I.40)

Inserting $x\hat{p} - \hat{p}x = -\frac{\hbar}{i}$,

$$\sum_{m'} f_{m'\,l} = \frac{1}{i\hbar}\left\{\left\langle l\left|-\frac{\hbar}{i}\right|l\right\rangle + \langle l|x\hat{p} + \hat{p}x|l\rangle\right\} = \langle l|l\rangle + \frac{1}{i\hbar}\langle l|x\hat{p} + \hat{p}x|l\rangle$$

(I.41)

Again $\langle l|l \rangle = 1$ and $\frac{1}{i\hbar}$ is imaginary so the second term drops out and the *f*-sum rule appears:

$$\sum_{m'} f_{m'\,l} = 1 \tag{I.42}$$

Returning to Eq. I.31,

$$n^2 = 1 - \frac{4\pi Ne^2}{m\omega^2} \sum_{m' \neq l} f_{m'\,l} \tag{I.31}$$

now Eq. I.42 is substituted to arrive at

$$\left\| n^2 = 1 - \frac{4\pi Ne^2}{m\omega^2} \right\| \tag{I.43}$$

Also returning to Eq. I.7,

$$n^2 = 1 + 4\pi\chi \rightarrow n = \sqrt{1 + 4\pi\chi} \tag{I.44}$$

A Taylor expansion about $\chi=0$ gives

$$n \simeq \left(\sqrt{1 + 4\pi\chi}\right)\Big|_{\chi=0} + \left(\frac{4\pi}{2\sqrt{1+4\pi\chi}}\right)\Big|_{\chi=0} \chi + \cdots \tag{I.45}$$

The first term in the series at $\chi = 0$ becomes 1; the second term is $\frac{\partial f}{\partial \chi}\Big|_{\chi=0}$, which approximates

$$n \approx 1 + 2\pi\chi \tag{I.46}$$

Rearranging Eqs. I.8 and I.16,

$$\chi = N\alpha = N \sum_{m' \neq l} \frac{e^2}{m(\omega_{m'\,l}^2 - \omega^2)} f_{m'\,l} \tag{I.47}$$

and then substituting into Eq. I.46,

$$n \approx 1 + \frac{2\pi Ne^2}{m}\sum_{m' \neq l}\frac{f_{m' l}}{(\omega^2_{m' l}-\omega^2)} \tag{I.48}$$

Bringing back the definition of oscillator strength in Eq. I.32,

$$f_{m' l} = \frac{2m\omega_{m' l}}{\hbar}\left|x_{m' l}\right|^2 \tag{I.32}$$

Eq. I.48 becomes

$$n \simeq 1 + \frac{2\pi Ne^2}{m}\sum_{m' \neq l}\frac{1}{(\omega^2_{m' l} - \omega^2)}\frac{2m}{\hbar}\omega_{m' l}\left|x_{m' l}\right|^2 \simeq 1 + \frac{4\pi Ne^2}{\hbar}\sum_{m' \neq l}\frac{\omega_{m' l}\left|x_{m' l}\right|^2}{(\omega^2_{m' l} - \omega^2)} \tag{I.49}$$

As above, $\omega_{m' l} = \frac{1}{\hbar}(E_{m'} - E_l)$, then

$$n \simeq 1 + \frac{4\pi Ne^2}{\hbar}\sum_{m' \neq l}\frac{\frac{1}{\hbar}\left(E_{m'} - E_l\right)}{\left(\left(\frac{1}{\hbar}\left(E_{m'} - E_l\right)\right)^2 - \omega^2\right)}\left|x_{m' l}\right|^2 = 1 +$$

$$4\pi Ne^2\sum_{m' \neq l}\frac{\left(E_{m'} - E_l\right)}{\left(\left(\left(E_{m'} - E_l\right)\right)^2 - \omega^2\right)}\left|x_{m' l}\right|^2 \tag{I.50}$$

If E_l is the energy of the initial state, now the oscillator strength is

$$f_{m' l} = \frac{2m}{\hbar}\frac{1}{\hbar}\left(E_{m'} - E_l\right)\left|x_{m' l}\right|^2 = \frac{2m}{\hbar^2}\left(E_{m'} - E_l\right)\left|x_{m' l}\right|^2 \tag{I.51}$$

If $|l\rangle$ is the ground state ($|l\rangle = |1\rangle$), then $E_{m'} - E_1 > 0$ and $f_{m' l} > 0$. This results in positive dispersion (attenuation) of the radiation field. If $|l\rangle$ is the excited state, e.g., $|l\rangle = |2\rangle$ and $|m'\rangle = |1\rangle$ (i.e., emission), then $E_{m'} - E_2 < 0$ and $f_{12} < 0$. This results in negative dispersion (amplification) of the radiation field.

Appendix J

Optimized Input Files

The following coordinates were optimized for ground-state planar porphin (HF and DFT), first excited-state porphin (HF), zinc porphin (HF and DFT), protoporphyrin IX (HF and DFT), zinc protoporphyrin IX (HF and DFT), and domed porphin (HF). The route section of each are included. Note that the porphin files are rotated centered in the *x*–*y* plane but PPIX files are not.

Hartree–Fock: # opt freq hf / sdd

Planar Ground-State Porphin, HF / sdd

Atom Type	X	Y	Z
N	2.11160	0.00000	0.00000
N	-0.01449	-2.05302	0.00000
N	-0.01449	2.05302	0.00000
N	-2.09603	0.00000	0.00000
H	1.11416	0.00000	0.00000
H	5.08385	-1.34351	0.00000
H	3.17236	-3.22028	0.00000
H	3.17236	3.22028	0.00000
H	5.08385	1.34351	0.00000
H	1.33274	-5.12403	0.00000
H	1.33274	5.12403	0.00000
H	-1.34062	-5.09647	0.00000
H	-1.34062	5.09647	0.00000
H	-3.18785	-3.19867	0.00000
H	-3.18785	3.19867	0.00000
H	-5.10750	-1.32601	0.00000
H	-1.10010	0.00000	0.00000
H	-5.10750	1.32601	0.00000
C	2.89362	-1.12232	0.00000
C	2.89362	1.12232	0.00000
C	4.23330	-0.69844	0.00000
C	2.41212	-2.46213	0.00000
C	2.41212	2.46213	0.00000

C	4.23330	0.69844	0.00000
C	1.11644	-2.87781	0.00000
C	1.11644	2.87781	0.00000
C	0.68019	-4.27781	0.00000
C	0.68019	4.27781	0.00000
C	-0.66998	-4.26505	0.00000
C	-1.07548	-2.85036	0.00000
C	-1.07548	2.85036	0.00000
C	-0.66998	4.26505	0.00000
C	-2.44140	-2.42903	0.00000
C	-2.44140	2.42903	0.00000
C	-2.87231	-1.13960	0.00000
C	-2.87231	1.13960	0.00000
C	-4.26202	-0.67411	0.00000
C	-4.26202	0.67411	0.00000

Porphin, DFT / O3LYP / cc-pVDZ

Atom Type	X	Y	Z
N	2.13446	0.00000	0.00000
N	-2.13446	0.00000	0.00000
N	0.00000	2.04475	0.00000
N	0.00000	-2.04475	0.00000
H	1.11966	0.00000	0.00000
H	-1.11966	0.00000	0.00000
H	5.15174	-1.35042	0.00000
H	5.15174	1.35042	0.00000
H	3.20493	-3.23846	0.00000
H	1.35471	-5.14061	0.00000
H	-1.35471	-5.14061	0.00000
H	-3.20493	-3.23846	0.00000
H	-5.15174	-1.35042	0.00000
H	-5.15174	1.35042	0.00000
H	-3.20493	3.23846	0.00000
H	-1.35471	5.14061	0.00000
H	1.35471	5.14061	0.00000
H	3.20493	3.23846	0.00000
C	2.92022	-1.14012	0.00000
C	2.92022	1.14012	0.00000
C	4.29356	-0.69254	0.00000
C	4.29356	0.69254	0.00000
C	2.44610	-2.46041	0.00000
C	1.10294	-2.88105	0.00000

C	-1.10294	-2.88105	0.00000
C	0.68565	-4.28953	0.00000
C	-0.68565	-4.28953	0.00000
C	-2.44610	-2.46041	0.00000
C	-2.92022	-1.14012	0.00000
C	-4.29356	-0.69254	0.00000
C	-4.29356	0.69254	0.00000
C	-2.92022	1.14012	0.00000
C	-2.44610	2.46041	0.00000
C	-1.10294	2.88105	0.00000
C	-0.68565	4.28953	0.00000
C	0.68565	4.28953	0.00000
C	1.10294	2.88105	0.00000
C	2.44610	2.46041	0.00000

Optimized First Excited-State Porphin

 # rcis(root=1,read)/sdd opt guess=read

Atom Type	X	Y	Z
N	2.10447	0.00000	0.00000
N	-0.00157	-2.05260	0.00000
N	-0.00157	2.05260	0.00000
N	-2.10771	0.00000	0.00000
H	1.10693	0.00000	0.00000
H	5.10124	-1.33357	0.00000
H	3.18366	-3.20918	0.00000
H	3.18366	3.20918	0.00000
H	5.10124	1.33357	0.00000
H	1.33712	-5.10887	0.00000
H	1.33712	5.10887	0.00000
H	-1.34062	-5.10856	0.00000
H	-1.34062	5.10856	0.00000
H	-3.18677	-3.20914	0.00000
H	-3.18677	3.20914	0.00000
H	-5.10505	-1.33304	0.00000
H	-1.11019	0.00000	0.00000
H	-5.10505	1.33304	0.00000
C	2.88410	-1.13210	0.00000
C	2.88410	1.13210	0.00000
C	4.25455	-0.68372	0.00000
C	2.43040	-2.44539	0.00000
C	2.43040	2.44539	0.00000
C	4.25455	0.68372	0.00000
C	1.09400	-2.86745	0.00000
C	1.09400	2.86745	0.00000
C	0.67636	-4.26926	0.00000
C	0.67636	4.26926	0.00000
C	-0.67973	-4.26910	0.00000
C	-1.09739	-2.86721	0.00000
C	-1.09739	2.86721	0.00000
C	-0.67973	4.26910	0.00000
C	-2.43366	-2.44523	0.00000
C	-2.43366	2.44523	0.00000
C	-2.88771	-1.13195	0.00000
C	-2.88771	1.13195	0.00000
C	-4.25790	-0.68372	0.00000
C	-4.25790	0.68372	0.00000

Zinc Porphin, DFT / O3LYP / cc-pVDZ

 # opt o3lyp/cc-pvdz

Atom Type	X	Y	Z
Zn	0.00300	0.00000	0.00000
N	2.06237	0.00000	0.00000
N	-2.05637	0.00000	0.00000
N	0.00300	2.05828	0.00000
N	0.00300	-2.05828	0.00000
H	5.11916	-1.35353	0.00000
H	5.11916	1.35353	0.00000
H	3.20426	-3.20218	0.00000
H	1.35608	-5.11555	0.00000
H	-1.35008	-5.11555	0.00000
H	-3.19826	-3.20218	0.00000
H	-5.11316	-1.35353	0.00000
H	-5.11316	1.35353	0.00000
H	-3.19826	3.20218	0.00000
H	-1.35008	5.11555	0.00000
H	1.35608	5.11555	0.00000
H	3.20426	3.20218	0.00000
C	2.87821	-1.10244	0.00000
C	2.87821	1.10244	0.00000
C	4.26169	-0.68352	0.00000
C	4.26169	0.68352	0.00000
C	2.43339	-2.43011	0.00000
C	1.10551	-2.87453	0.00000
C	-1.09951	-2.87453	0.00000
C	0.68647	-4.25773	0.00000
C	-0.68047	-4.25773	0.00000
C	-2.42739	-2.43011	0.00000
C	-2.87221	-1.10244	0.00000
C	-4.25569	-0.68352	0.00000
C	-4.25569	0.68352	0.00000
C	-2.87221	1.10244	0.00000
C	-2.42739	2.43011	0.00000
C	-1.09951	2.87453	0.00000
C	-0.68047	4.25773	0.00000
C	0.68647	4.25773	0.00000
C	1.10551	2.87453	0.00000
C	2.43339	2.43011	0.00000

Zinc Porphin, HF / cc-pVDZ

 # opt hf/cc-pvdz

Atom Type	X	Y	Z
Zn	0.00300	0.00000	0.00000
N	2.05626	0.00000	0.00000
N	-2.05026	0.00000	0.00000
N	0.00300	2.05304	0.00000
N	0.00300	-2.05304	0.00000
H	5.08783	-1.34060	0.00000
H	5.08783	1.34060	0.00000
H	3.18192	-3.17936	0.00000
H	1.34330	-5.08490	0.00000
H	-1.33730	-5.08490	0.00000
H	-3.17592	-3.17936	0.00000
H	-5.08183	-1.34060	0.00000
H	-5.08183	1.34060	0.00000
H	-3.17592	3.17936	0.00000
H	-1.33730	5.08490	0.00000
H	1.34330	5.08490	0.00000
H	3.18192	3.17936	0.00000
C	2.85480	-1.09332	0.00000
C	2.85480	1.09332	0.00000
C	4.23845	-0.67575	0.00000
C	4.23845	0.67575	0.00000
C	2.41764	-2.41467	0.00000
C	1.09623	-2.85187	0.00000
C	-1.09023	-2.85187	0.00000
C	0.67873	-4.23529	0.00000
C	-0.67273	-4.23529	0.00000
C	-2.41164	-2.41467	0.00000
C	-2.84880	-1.09332	0.00000
C	-4.23245	-0.67575	0.00000
C	-4.23245	0.67575	0.00000
C	-2.84880	1.09332	0.00000
C	-2.41164	2.41467	0.00000
C	-1.09023	2.85187	0.00000
C	-0.67273	4.23529	0.00000
C	0.67873	4.23529	0.00000
C	1.09623	2.85187	0.00000
C	2.41764	2.41467	0.00000

Protoporphyrin IX, DFT / O3LYP / cc-pVDZ

 # opt o3lyp/cc-pvdz, -2 charge

Atom Type	X	Y	Z
O	-5.31723	2.06927	5.47096
O	-5.69553	0.03395	4.53863
O	4.74973	-1.97901	6.11482
O	5.14110	0.07065	5.21995
N	-1.39977	-0.16724	0.26317
N	1.63215	0.13146	-2.47478
N	-1.27381	-0.19556	-2.72790
N	1.47265	0.13270	0.53406
H	-0.23500	-0.01300	3.44273
H	4.56068	0.47567	-0.78941
H	0.44394	-0.01660	-5.64314
H	-4.35520	-0.48289	-1.35885
H	2.40136	-0.30670	-6.47268
H	-5.17292	-1.03916	-3.24696
H	-2.13643	-0.61411	4.28684
H	-3.82486	-0.98123	3.89254
H	-3.49925	1.86127	3.48154
H	-2.87073	1.45850	5.06869
H	3.30366	1.01398	4.38524
H	1.59578	0.56282	4.55987
H	2.33429	-1.45683	5.45443
H	3.14003	-1.84958	3.94812
H	-6.43011	-0.56675	-5.21196
H	-4.94169	0.10479	-6.09806
H	4.94567	1.38293	-6.07718
H	4.23418	0.67373	-7.64230
H	-2.50420	0.53784	-6.75454
H	-1.28495	-0.74742	-6.72920
H	-3.00700	-1.15279	-6.59873
H	5.66733	-0.02898	-4.32522
H	5.60717	1.49976	-3.43216
H	5.72711	-0.02977	-2.55084
H	5.10393	0.45078	3.01553
H	5.47674	-0.22458	1.40937
H	5.23816	1.52112	1.58785
H	-5.54095	0.25407	0.64449
H	-5.36388	-1.49582	0.85005
H	-5.40600	-0.42432	2.28241
H	-0.60284	-0.09127	-1.97313

Protoporphyrin IX, DFT / O3LYP / cc-pVDZ (cont'd)

Atom Type	X	Y	Z
H	0.77877	0.05182	-0.20046
C	-1.33585	-0.14083	1.62335
C	-2.72921	-0.30548	-0.02143
C	-2.67323	-0.27316	2.24570
C	-3.54834	-0.37132	1.19416
C	-0.14129	-0.01559	2.35884
C	1.15962	0.11075	1.86567
C	2.82227	0.28299	0.35795
C	2.40153	0.25780	2.60985
C	3.41979	0.36337	1.67340
C	3.48243	0.34765	-0.86761
C	2.95106	0.27526	-2.17322
C	3.77003	0.34769	-3.38035
C	2.88757	0.24451	-4.44259
C	1.55730	0.09892	-3.83123
C	0.36338	-0.03846	-4.55813
C	-0.93788	-0.18123	-4.05915
C	-2.15591	-0.33581	-4.79952
C	-3.20341	-0.43188	-3.87365
C	-2.61803	-0.34623	-2.55094
C	-3.27230	-0.38536	-1.31205
C	3.14694	0.22896	-5.87386
C	-4.62801	-0.62328	-4.09997
C	-5.35905	-0.35266	-5.20102
C	4.16862	0.78692	-6.55728
C	-3.00516	-0.26905	3.70574
C	-3.52132	1.08304	4.26256
C	2.54870	0.26376	4.09967
C	3.05447	-1.06515	4.71852
C	-5.00213	1.04900	4.81957
C	4.46380	-0.97361	5.43000
C	-2.23958	-0.42800	-6.28964
C	5.26128	0.45072	-3.42282
C	4.88190	0.53714	1.94063
C	-5.03707	-0.51885	1.25011

324

Protoporphyrin IX, HF / cc-pVDZ

 # opt hf / cc-pvdz, -2 charge

Atom Type	X	Y	Z
O	-1.16602	-1.63252	9.05416
O	0.92913	-1.65041	8.28762
O	6.23896	-6.55895	3.68446
O	4.06921	-6.09699	3.92664
N	0.63466	0.32183	2.31092
N	2.93504	1.50900	-0.83869
N	0.26486	2.30005	0.09305
N	3.20773	-0.48025	1.32439
H	2.14744	-2.10739	4.06683
H	5.79863	-0.24451	-0.79283
H	1.37981	3.83816	-2.71409
H	-2.22032	2.11451	2.36748
H	3.44794	4.52787	-3.67082
H	-3.50974	3.21093	1.07397
H	1.11127	-1.17739	5.98265
H	0.01643	-2.39140	5.40093
H	-1.82718	-1.38983	6.69235
H	-0.91161	0.09245	6.92041
H	4.62560	-3.01759	4.29044
H	3.48477	-3.91219	3.33552
H	5.46291	-4.66151	1.85270
H	6.49734	-4.16202	3.18259
H	-4.64466	5.15655	0.38865
H	-3.21842	5.72386	-0.62825
H	4.48146	2.23431	-5.40535
H	4.16784	3.98949	-5.89277
H	-1.15607	5.83042	-1.82797
H	-0.55496	4.70029	-3.03692
H	-2.23361	4.62116	-2.51967
H	6.13654	0.46268	-3.32095
H	6.33325	2.19190	-3.62255
H	6.81333	1.49857	-2.07099
H	6.92028	-2.85795	1.60385
H	6.42231	-2.66326	-0.07445
H	7.10449	-1.29788	0.80844
H	-2.36692	1.72735	4.90383
H	-3.17709	0.61323	3.80287
H	-2.63906	0.05503	5.38226
H	0.98258	1.66998	0.38580

Protoporphyrin IX, HF / cc-pVDZ (cont'd)

Atom Type	X	Y	Z
H	2.52276	0.20724	1.09801
C	0.92397	-0.51432	3.26436
C	-0.59294	0.87617	2.65505
C	-0.10077	-0.53097	4.36095
C	-1.06123	0.30886	3.93979
C	2.08333	-1.38118	3.27394
C	3.08883	-1.35786	2.37138
C	4.38287	-0.65389	0.67604
C	4.29920	-2.22455	2.34274
C	5.08887	-1.75572	1.35413
C	4.83090	0.03354	-0.41094
C	4.12216	1.05601	-1.12258
C	4.69337	1.71457	-2.32277
C	3.75143	2.57845	-2.74623
C	2.62847	2.45727	-1.80026
C	1.47225	3.15045	-1.88751
C	0.33672	3.09824	-1.01179
C	-0.84474	3.81981	-1.09168
C	-1.64186	3.43167	0.01851
C	-0.90583	2.47538	0.73076
C	-1.27503	1.80397	1.95399
C	3.78470	3.51767	-3.88262
C	-2.98692	3.87587	0.39645
C	-3.64157	4.97339	0.02387
C	4.16857	3.23077	-5.11924
C	0.08847	-1.32325	5.63524
C	-0.81781	-0.99502	6.82089
C	4.44495	-3.40096	3.28062
C	5.48105	-4.46932	2.93139
C	-0.28909	-1.49525	8.19870
C	5.23146	-5.85423	3.60294
C	-1.21296	4.79410	-2.17386
C	6.06661	1.45506	-2.86659
C	6.45955	-2.17497	0.90319
C	-2.38089	0.69339	4.54763

Zinc Protoporphyrin IX, DFT / O3LYP / cc-pVDZ

 # opt o3lyp/cc-pvdz, -2 charge

Atom Type	X	Y	Z
Zn	1.75602	0.84315	0.75520
O	-0.75317	-0.53830	9.03430
O	-2.22244	-1.23839	7.45120
O	5.73526	-6.63604	2.86855
O	7.01119	-4.77429	3.11346
N	0.59960	0.28741	2.34755
N	2.91749	1.43001	-0.87376
N	0.27017	2.15801	0.12409
N	3.22041	-0.45733	1.34721
H	2.23687	-2.05253	4.16897
H	5.84577	-0.29370	-0.78271
H	1.27380	3.77517	-2.68594
H	-2.32056	1.99672	2.30906
H	2.64241	3.61633	-4.30011
H	-3.60468	3.07259	0.98653
H	0.47738	-2.47880	5.36860
H	-1.21786	-2.02922	5.63171
H	0.36833	0.15325	6.66917
H	0.98347	-1.34424	7.34284
H	5.84576	-3.00723	3.81264
H	4.20701	-2.89941	4.48250
H	3.90085	-5.09960	3.78324
H	4.12315	-4.61468	2.11345
H	-4.69203	5.12118	0.43504
H	-3.19694	5.75682	-0.46862
H	5.71930	3.71242	-4.38897
H	4.48070	4.50704	-5.52455
H	-0.99428	5.72446	-1.94001
H	-0.74857	4.42621	-3.11933
H	-2.34683	4.67397	-2.38921
H	5.89384	1.37141	-4.13065
H	6.77575	1.96706	-2.71492
H	6.31487	0.26089	-2.81206
H	6.87394	-3.00360	1.71867
H	6.59504	-2.49998	0.03393
H	7.31804	-1.35673	1.17643
H	-2.60986	1.47271	4.71267
H	-3.30416	0.07707	3.87310
H	-2.55443	-0.13024	5.48462

Atom Type	X	Y	Z
C	0.92101	-0.63076	3.31512
C	-0.64982	0.76267	2.66412
C	-0.16708	-0.75055	4.29061
C	-1.14476	0.12889	3.87505
C	2.12039	-1.35344	3.34367
C	3.18999	-1.28471	2.43982
C	4.43077	-0.67341	0.73130
C	4.42991	-2.06216	2.53090
C	5.19986	-1.67182	1.45547
C	4.85545	-0.01793	-0.42467
C	4.16736	0.95149	-1.17622
C	4.66772	1.58542	-2.37489
C	3.68037	2.47232	-2.79432
C	2.58831	2.34502	-1.83665
C	1.38308	3.05740	-1.87546
C	0.30615	2.97359	-0.98244
C	-0.92015	3.73187	-1.08210
C	-1.70431	3.35708	0.00407
C	-0.93200	2.36065	0.73886
C	-1.34453	1.71686	1.91740
C	3.64706	3.34706	-3.95635
C	-3.04284	3.78922	0.37861
C	-3.67120	4.94701	0.08686
C	4.67451	3.87132	-4.65717
C	-0.20081	-1.62378	5.50651
C	0.10504	-0.90324	6.84516
C	4.77626	-3.09247	3.56125
C	4.59080	-4.56470	3.11091
C	-1.08652	-0.90179	7.88555
C	5.92477	-5.41076	3.02838
C	-1.26636	4.68215	-2.18412
C	5.97328	1.28023	-3.03747
C	6.56499	-2.15970	1.08292
C	-2.46830	0.39543	4.52146

Zinc Protoporphyrin IX, HF / cc-pVDZ

 # opt hf / cc-pvdz, -2 charge

Atom Type	X	Y	Z
Zn	1.77630	0.90475	0.71450
O	-1.20822	-1.56022	9.04493
O	0.89703	-1.53797	8.30597
O	6.20919	-6.49875	3.80792
O	4.04428	-6.01129	4.03889
N	0.63228	0.36197	2.31303
N	2.93362	1.49173	-0.89861
N	0.29360	2.21675	0.11154
N	3.20533	-0.43460	1.31034
H	2.23195	-1.98447	4.12566
H	5.79517	-0.30850	-0.83214
H	1.26570	3.82247	-2.68671
H	-2.26094	2.07331	2.28737
H	3.06996	4.35979	-3.78385
H	-3.66173	3.05474	0.73221
H	1.09574	-1.17915	5.99973
H	0.01292	-2.39811	5.40374
H	-1.84668	-1.40995	6.67187
H	-0.97127	0.10198	6.86059
H	4.64996	-2.92978	4.32522
H	3.48621	-3.82519	3.39789
H	5.43585	-4.64478	1.92216
H	6.49202	-4.12535	3.22625
H	-4.51230	5.28519	0.61445
H	-2.88646	5.92036	0.01089
H	4.95499	2.49987	-5.29625
H	4.25972	4.12699	-5.81526
H	-0.77456	5.72929	-2.01332
H	-0.91072	4.38692	-3.14660
H	-2.32387	4.91333	-2.22856
H	5.98616	0.70107	-3.79718
H	6.42414	2.35415	-3.37422
H	6.73585	1.04568	-2.24296
H	6.95191	-2.83511	1.63655
H	6.40646	-2.79574	-0.03607
H	7.11906	-1.35741	0.69466
H	-2.45510	1.61803	4.81700
H	-3.17750	0.47543	3.68569
H	-2.67413	-0.06595	5.28319

Zinc Protoporphyrin IX, HF / cc-pVDZ (cont'd)

Atom Type	X	Y	Z
C	0.98135	-0.52069	3.29652
C	-0.58765	0.84812	2.62815
C	-0.07981	-0.54973	4.34841
C	-1.05925	0.25646	3.89637
C	2.10751	-1.29858	3.30470
C	3.14542	-1.27887	2.32655
C	4.40313	-0.66620	0.65370
C	4.32686	-2.17584	2.35035
C	5.11545	-1.75178	1.33750
C	4.83215	-0.01027	-0.44864
C	4.16141	1.01611	-1.20055
C	4.66940	1.64848	-2.35558
C	3.68107	2.54615	-2.76743
C	2.62179	2.42157	-1.83208
C	1.38718	3.13023	-1.86732
C	0.34159	3.03887	-1.00081
C	-0.90025	3.79061	-1.08399
C	-1.65338	3.41342	-0.01845
C	-0.86923	2.41170	0.71329
C	-1.30091	1.77649	1.90023
C	3.68402	3.47333	-3.90596
C	-3.00870	3.84079	0.36704
C	-3.49064	5.07689	0.32092
C	4.33863	3.35585	-5.05827
C	0.08093	-1.32932	5.63363
C	-0.84936	-0.98529	6.79701
C	4.45317	-3.32895	3.32403
C	5.46829	-4.42427	2.99541
C	-0.32390	-1.42461	8.19726
C	5.20811	-5.78646	3.70665
C	-1.24764	4.75707	-2.17757
C	6.02098	1.42252	-2.97409
C	6.47287	-2.21659	0.88893
C	-2.41350	0.58530	4.45960

Non-planar, Domed Ground-State Porphin, HF / sdd

 # opt=(calcfc,z-matrix) hf/sdd

Atom Type	X	Y	Z
N	2.13788	0.04961	0.60000
N	-2.09981	-0.01527	0.60000
N	-0.03444	2.08685	0.60000
N	0.02892	-2.05317	0.60000
H	1.14918	0.03448	0.73216
H	-1.11157	-0.00014	0.72496
H	5.07144	-1.24794	0.00000
H	5.03035	1.43633	0.00000
H	3.22365	-3.14473	0.25143
H	1.42065	-5.04716	0.00000
H	-1.25101	-5.05868	0.00000
H	-3.11814	-3.22085	0.25766
H	-5.03348	-1.38523	0.00000
H	-5.07403	1.26428	0.00000
H	-3.21576	3.15766	0.25766
H	-1.40578	5.05177	0.00000
H	1.26499	5.12204	0.00000
H	3.12538	3.27568	0.25143
C	2.92244	-1.06064	0.43849
C	2.88809	1.18335	0.43849
C	4.23264	-0.61685	0.19574
C	4.21126	0.77986	0.19574
C	2.45829	-2.40635	0.39934
C	1.17164	-2.84673	0.43988
C	-1.01912	-2.85137	0.43756
C	0.75671	-4.23180	0.19244
C	-0.59318	-4.23799	0.18753
C	-2.38987	-2.44707	0.40102
C	-2.84323	-1.16606	0.43256
C	-4.22006	-0.72167	0.19655
C	-4.24069	0.62592	0.19655
C	-2.87810	1.11224	0.43256
C	-2.46416	2.40652	0.40102
C	-1.10642	2.85259	0.43756
C	-0.72313	4.25160	0.18752
C	0.62632	4.28673	0.19243
C	1.08345	2.91502	0.43988
C	2.38297	2.51422	0.39934

331

CURRICULUM VITAE

Hannah E. Wagie

<u>Education</u>   University of Wisconsin – Milwaukee, Milwaukee, WI, January 2010–Fall 2015), Ph.D., Physical Chemistry.

University of Wisconsin – Milwaukee, Milwaukee, WI, September 2007–December 2009, Bachelor of Science, Chemistry with Biochemical Option, *summa cum laude.*

<u>Experience</u>   University of Wisconsin – Milwaukee, Department of Chemistry & Biochemistry, Milwaukee, WI. **Lecturer**. Spring 2015.

- **Course taught**: CHEM 100 ("Chemical Science," prerequisite to General Chemistry I)
- Emphasis on developing problem-solving skills along with basic chemistry concepts
- Course management programs utilized:
    - **TopHat** attendance system
    - **Connect** (McGraw-Hill) online homework system
    - **D2L** (Desire2Learn), the University of Wisconsin system's online course management system
- Supervision of two graduate-student teaching assistants
- Incorporation of visual images and videos into a traditional blackboard presentation
- Incorporated **Classroom Salon** as part of chemical education research

University of Wisconsin – Milwaukee, Department of Chemistry & Biochemistry, Milwaukee, WI. **Teaching assistant**. September 2009– present.

- Writing an interferometry lab experiment for CHEM 563: Physical Chemistry laboratory
- Lectures in support of laboratory experiments
- Techniques used in teaching labs:
    - UV–VIS spectrometry
    - Titration: pH, iodometry, colorimetric, complexation
    - Paper chromatography
    - Flame photometry
    - Fluorometry
    - Gas chromatography
    - Ion exchange chromatography
    - Cyclic voltammetry
- Introduction & review of chemistry concepts including mathematical models
- Teaching problem solving strategies
- Writing weekly quizzes & worksheets
- Grading quizzes, exams, homework sets & lab write-ups
- One-on-one tutoring for introductory chemistry in the MacFarland Reading Room
- Mentor for new teaching assistants, Fall 2011, Fall 2012, Fall 2014
- Supervision of four undergraduate researchers, Spring 2012, Spring 2013, Spring 2014, Summer 2014
- **Courses taught**:
    - CHEM 100: Chemical Science discussion
    - CHEM 102: General Chemistry discussion & laboratory
    - CHEM 104: General Chemistry II discussion & laboratory
    - CHEM 105: General Chemistry for Engineering laboratory
    - CHEM 221: Elementary Quantitative Analysis laboratory

University of Wisconsin – Milwaukee, Department of Chemistry & Biochemistry, Milwaukee, WI. **Research assistant**. January 2010– present.

- Conducting **original biophysics** research with the goal to measure internal electric fields in heme proteins

- **GAUSSIAN09 + GaussView5** computational package to predict excitation energies and oscillator strengths of various porphyrins; attended **Introduction to Gaussian** workshop, July 2012, Ohio State University
- **UNIX-based cluster** for computational work
- **Protein modification** of heme cofactor in myoglobin to substitute various porphyrins
- **Protein purification** with chromatography & centrifugation methods
- **Metallation reactions** to produce commercially unavailable porphyrins
- **UV-VIS absorbance & emission spectroscopy** to characterize various porphyrins & modified proteins
- **Low-temperature laser spectroscopy** using liquid nitrogen cryostat (77 K) to characterize porphyrins & proteins in solid matrices
- **Class IV lasers:** 532-nm solid state, tunable dye laser
- **Scanning electron microscopy** to characterize polystyrene microspheres
- **Chemical inventory** with CISPro management system
- **Software**: OriginPro 7.0, GAUSSIAN09, GaussView 5.0, Microsoft Office, EndNote, MATLAB, CorelDraw, PyMOL, ImageJ

Publications

Wagie, H. E. and P. Geissinger (2012). "Hole-Burning Spectroscopy as a Probe of Nano-environments and Processes in Biomolecules: A Review." *Applied Spectroscopy* **66**(6): 609-627.

Wagie, H. E., J. C. Woehl, and P. Geissinger (2015). "A Systematic, Finite-Order Approach for the Quantitative Determination of Internal Electric Fields in Proteins from Stark Spectroscopy." *In preparation.*

Awards

- UWM Graduate School *Distinguished Graduate Student Award*, 2012–2013 academic year
- UWM Office of Undergraduate Research (OUR) *SURF Award for undergraduate research funding* with Peter Geissinger's group, Spring & Summer 2009
- SCIX Conference, **Three student poster awards** at the 2009 (as an undergraduate), 2011, 2013
- UWM Department of Chemistry & Biochemistry 2014 Awards Day, **1st Place poster**

Presentations

- UWM Department of Chemistry & Biochemistry Awards Day, ***poster presentation***, April 2015, for graduate research entitled "Spectroscopic Sensitivity of Porphyrins to Protein Environment"
- Biophysical Society Annual Meeting, ***poster presentation***, February 2015, Baltimore, MD, for graduate research entitled "Development of a Quantum-Mechanical Analysis of Stark Effects of Porphyrins Employed as Sensors of Internal Electric Fields in Biological Systems"
- SCIX conference, ***oral presentation***, October 2014, Reno, NV, for graduate research entitled "Spectroscopic Method Using Porphyrins as a Probe of Local Electrostatic Environment"
- UWM Research Foundation Breakfast Research Talk, oral presentation, May 2014, UWM campus, for graduate research entitled "Porphyrin as a Spectroscopic Probe of Net Electric Fields in Heme Proteins"
- UWM Department of Chemistry & Biochemistry Awards Day, ***poster award***, April 2014, for graduate research entitled "Predicting the First Electronic Transition Energy in Porphin with an Applied Electric Field (Stark Spectroscopy)"
- SCIX conference, ***poster award***, October 2013, Milwaukee, WI, for graduate research entitled "Effects of Local Environment on UV-VIS Spectra of Zinc Porphyrins: Protein & Solvent"
- UWM Department of Chemistry & Biochemistry Awards Day, ***poster presentation***, April 2013, for graduate research entitled "Towards an Electrostatic Model of Heme Proteins: Environment & Electronic Properties of Zinc Porphyrins"
- SCIX conference (formerly FACSS), ***invited talk***, October 2012, Kansas City, MO, for recent review paper published in *Applied Spectroscopy* entitled "What Lies Beneath: Broadened Absorption Spectra of Biomolecules Exposed"
- Biophysical Society Annual Meeting, ***poster presentation***, February 2012, San Diego, CA, for graduate research entitled "Porphyrins as Detectors of Internal Electric Field in Heme Proteins"
- FACSS conference SAS student session, ***poster award & oral presentation***, October 2011, Reno, NV, for graduate research entitled "Molecular Electric Field Determination in Large Proteins with Hole-Burning Spectroscopy"
- UWM Department of Chemistry & Biochemistry Awards Day, ***poster presentation***, April 2012, for graduate research entitled "Describing Myoglobin's Electrostatic Structure"
- UWM Department of Chemistry & Biochemistry Awards Day, ***poster presentation***, May 2011, for graduate research entitled "Exploring the Electrostatic Effects of Point Mutations in the Myoglobin Heme Pocket"

- UWM Department of Chemistry & Biochemistry Awards Day, *poster presentation*, May 2010, for graduate research entitled "Determination of Internal Electric Fields in Protoporphyrin IX with Hole-Burning Spectroscopy"
- Federation of Analytical Chemistry & Spectroscopy Societies (FACSS) conference, *poster award*, October 2009, Louisville, KY, for undergraduate research entitled "Optimizing a Luminescent Dye Mixture for pH-Sensing in Optical Fiber Systems for Water Quality Monitoring"
- National Conference on Undergraduate Research (NCUR), *poster presentation*, April 2009, LaCrosse, WI, undergraduate research, "Remote Monitoring with Luminescent Optical Fibers: Enhancing Reproducibility and Dynamic Range"
- American Water Resources Association (AWRA) conference, *poster presentation*, March 2009, Stevens Point, WI, undergraduate research, "Enhancing Reproducibility and Dynamic Range in Remote Monitoring with Luminescent Optical Fiber Sensors"

Volunteer & Service

- Wisconsin Science Olympiad, *chemistry competition judge*, hosted by UWM, April 2015
- SMART (Students Modeling a Research Topic) Team *mentor* for Milwaukee School of Engineering's BioMolecular Modeling Center's *program to expose high-school students to biochemistry research*, 2013-2014 schoolyear, 20142015 schoolyear
- Badger State Science & Engineering Fair, *process judge*, hosted by Marquette University, April 2013
- NSF-funded scale project outreach (Murphy, Geissinger, Owen), *remote SEM operator*, September 2012,
- Society for Applied Spectroscopy (SAS), *community education outreach*, May 2011 & May 2012 conception & production of original forensic science spectroscopy lab for high-school students, presentation to 500 high school students at four Milwaukee-area high schools
- Chemistry & Biochemistry Graduate Student Council, *founding president*, service organization with influence in department policies, June 2011 – May 2012
- *Applied Spectroscopy* international journal, *volunteer copyediting & proofing* for journal's managing editor, October 2011 – August 2012
- *Mentor* for new teaching assistants in the Department of Chemistry & Biochemistry, Fall 2011, Fall 2012, Fall 2014
- Society for Applied Spectroscopy (SAS), *founding officer* of student section at UWM, 2009 – 2010