Computational studies of noncovalent interactions in ligand-gated ion channels

- and -

Synthesis and characterization of red and near infrared cyanine dyes

Thesis by

Matthew Robert Davis

In Partial Fulfillment of the Requirements for the

degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2018

(Defended 10 August, 2017)

Matthew Robert Davis
ORCID: 0000-0002-6374-4555

Acknowledgements

I have been immeasurably fortunate to have not only spent five years performing state-of-the-art basic research, but to have stumbled into many valuable relationships, friendships, and partnerships that have infused my life with happiness and meaning. There are many individuals that have made my experience at Caltech what it is today, but several deserve special mention;

Dennis Dougherty, for being an excellent mentor and teacher. From Dennis I have learned the value of precise experimental design, the guiding principles of physical organic chemistry, and a healthy respect for effective pedagogy. He has given me the freedom to investigate a huge variety of different problems. This is reflected in my thesis: I have had the opportunity to study computational chemistry, ion channel electrophysiology, and organic synthesis of dye molecules. Not many advisors would allow such a broad focus, and I feel that it has allowed me to reach my potential as a graduate student. The chance to work closely on some of the computational work described in this thesis has been a dream come true. Perhaps most importantly, however, Dennis has imparted to our group the importance of a healthy work-life balance. I admire his relationship with Ellen, his knowledge of current events, and his Dodger fandom. I always enjoyed the opportunities to discuss the merits of OPS+ versus wRC+ in his office;

My thesis committee, for providing valuable insight and direction to my doctoral work. Linda Hsieh-Wilson, Long Cai, and Harry Gray all have brought fresh perspectives to my work. I enjoyed the opportunity to teach alongside Professor Hsieh-Wilson: the experience in pedagogy I gained was extremely valuable. It would be remiss to not mention Linda Syme, whose tireless work keeps the lab functioning day-to-day, and with whom I

have enjoyed many conversation. Agnes Tong (formerly) and Allison Ross (currently) also work tirelessly in the division office to make us feel welcome and to make the graduate student experience relatively painless;

Scott Virgil, David Vander Velde, Jay Winkler, and Mona Shahgholi, the tireless experts and administrators of the chemistry department's facilities. Scott Virgil, who in addition to being an incredibly brilliant scientist, is a life saver for making sure the Church LC-MS run on time and for tolerating perhaps hundreds of e-mails from our lab. David Vander Velde, who was always quick to lend a hand in the NMR facility, Jay Winker, the unflappable curator of the Beckman Institute Laser Resource Center (BILRC), and Mona Shahgholi, who always had time to re-explain how the MALDI works;

Joshua Kritzer, my undergraduate advisor at Tufts University, for kindling my passion for science and research and for still being a role model and advisor today. His approach to scientific research and inquiry forms the backbone of my approach to this day. Justin Quartararo, Santosh Kuman Chourdary, Amanda Aldous, and Jason Gavenonis all served as early role models as graduate students and postdoctoral scholars at Tufts, and their advice and friendship are cherished. I am excited to see what the future holds in their young careers;

Carolyn Mordas and the team working on oral care at Johnson & Johnson helped shape my opinions about scientific industry, and helped to make for a very memorable internship experience. I expect to be talking about making and using my own toothpaste for a long time to come;

The Dougherty lab, in all its various iterations, has been a wonderful place to work and be a budding scientist. The group has always been full of talented people, and I am

lucky that they have all be fun as well. Driving to Pasadena was never a chore. Kristina Daeffler was always a voice of reason in my first year and a half in the lab, and I knew that she would be there in any way that anyone in the lab needed her. Ethan Van Arnam, was a ray of sunshine every day he was there (in spite of his perpetual state of injury). His scientific acumen was impressive, and I admired his ability to tackle difficult problems in innovative ways. Noah Duffy is a loyal friend and always finds a way to fix anything. I am glad that he has stayed in the area: getting a chance to hang out over the past few years has been welcome. Tim Miles, in addition to being the catalyst for Chapter 2 and 3 (and therefore a large portion of my scientific work over the past four years), is an incredibly intelligent person, and I have enjoyed our conversations. Ximena da Silva was there if I had a question. Oliver Shafaat showed me the ropes on the fluorimeter in BILRC. Our many conversations during marathon sessions at the fluorimeter are highlights of my doctoral work. I owe much of my work to him. Clint Regan taught me the basics of organic synthesis. Without his early mentorship, I would not have accomplished nearly what I have today. His knowledge of proper scientific inquiry is unmatched, and I am excited to see where his career takes him. Chris Marotta is a dear friend, and I am very glad that he is still around when we need him for scientific or personal advice. I hope that I have left up the mantle of Tryptoman captain as effectively as he gave it. Betty Wong's sunny disposition and optimism were always a welcome sight in the lab: I am excited to see the results of the quantum and molecular mechanical studies we have reinvigorated in the lab. Matt Rienzo, in addition to being a thoughtful scientist, is a loyal friend, and I continually enjoy our camping trips. Kayla Busby, (David) Paul Walton, and I formed a triumvirate in the Dougherty group in our first year. I am excited to watch Kayla's career unfold after her

doctorate at UCSD: she approaches problems in ways I could not, and is one of the kindest people I have met at Caltech. Paul Walton is a man of many talents: organic and photochemist, extremely fast runner, coffee roaster *extraordinaire*, and loyal friend (and sometimes a goofball). He is a brilliant scientist and a thoughtful person. Annet Blom, in addition to being a constant neighbor and companion in the North Bay nation, is a wonderful person and a great friend. Bryce Jarman does anything to solve a scientific problem, a trait I very much admire. His leadership, both in scientific matters and matters of rooftop recreation, will take the group far in the future. Catie Blunt, another member of Team Photoacid™, is an adventurous scientist in and outside of the lab. I have enjoyed getting a chance to talk Doctor Who, tennis, and winter sports. Richard Mosseso, my neighbor in more ways than one, provides a lighthearted aspect that has made long days in lab much more bearable. I strive to live like SHAMASH, his Dragonborn Paladin. Steve Grant is the future of the Dougherty lab, and will continue the amazing environment I have enjoyed for five years;

Mike Post, who deserves special mention for several reasons. Our collaboration together was an interesting diversion into the world of hydrogen bonding and 60s-era NMR papers. That particular experiment (Chapter 4, section 4.4.2) remains a highlight of the past five years. Beyond that, Mike has an outsized impact on my life: he introduced me to my wife and was a groomsman at my wedding. Our baseball games, Shabbat dinners, and frequent exhortations to go and have fun are a highlight of grad school. If he is reading this, I do remind him that we still need to implement our Sabermetrics-only fantasy baseball league. Though he lives in New York City, this thesis could not have been completed without his constant advice and confidence. I am excited to watch his academic career;

The Caltech music department. Though these words seem oxymoronic, I was lucky to have worked with some of the best musicians I have had the fortunate to work with in my life. Bill Bing, the former Director of Bands, always had time for his students, and the institute is better for it. Barb Catlin, the director of the Jazz Band, is a talented director who is not afraid to put some serious charts in front of rocket scientists. My Mondays spent playing with the jazz band have been some of the most memorable at Caltech. Glenn Price, the new Wind Ensemble director, is a similarly talented director. The future of Caltech's music program is in excellent hands, and I look forward to playing a role in it. Deserving special mention are Lynne and Steve Snyder. They have at times been mentors, confidantes, and occasional clarinet teachers. I am lucky to have such amazing people as role models;

My friends at Caltech. There are too many to mention, but the folks at the Lake house always made me feel welcome and were amazing companions for both good and tough times. Nick Cowper and Adam Boynton deserve special mention. Nick has an infectious personality. He is a brilliant scientist and, more importantly, a genuinely kind person. Adam Boynton, who I had the great fortune to live with for my first several years of graduate school, is a similarly rare kind of person. He is a loyal friend, up for anything, even if it requires a healthy investment and a closet covered in proto-Belgian Trippel Ale;

My friends outside of Caltech. Due to my fortune of meeting and marrying an incredible woman who does not go to Caltech, I had the opportunity to meet some amazing new friends. Ryan Willis, Emily Rothstein, Charlie Moran, Sasha Silver, Amy Pijanowski, Laura Caccavo, Justin Yates, and Sara Dolin have at times been a welcome distraction from the Caltech community, and have brought balance and a good deal of fun to my life;

The November Project has been a life changing experience. In more than 40 cities across the globe, people get together at ungodly hours of the morning to run around and do a bunch of silly exercises. The November Project community in LA has enriched Rachel and my life, and provided us with a great group of friends and an opportunity to see Los Angeles in ways that we could never have imagined. The leaders of NP_LAX while we have lived here: Steve Selnick, Ben Otto, Tara Wall, Maggie Nolting, and Jesse Buck-Brennen work tirelessly to provide a positive experience and keep me in shape. Special thanks should go to Steve, who generously let us adopt our dog Kairi, a new and exciting experience for us;

My new extended family: Melissa Sparr and Maxwell Sparr. They have been supportive and positive in even the toughest times. Melissa is always there with advice and encouragement. Maxwell has been like a brother from the moment we met, and his support through the past few years has been invaluable.

Seth and Albert Davis, my brothers. Being a triplet would be an interesting experience if I had known anything else. What I do know is that I have two amazing brothers who have been there before day 1. Albert, with whom I share a love of cars, beer, and football, has been there at moments of levity and sadness, and is a bulwark and a confidante. His perspective has shaped mine. Seth, with whom I could talk for hours, has served as a companion in times of leisure but always has time to lend an ear or his time. I am so excited that he and Michelle are moving to Texas. While still not quite California, a step closer is a step closer;

My parents, for encouraging a love of science and inquiry. In my highest and lowest times, they have been there with an open ear and words of encouragement. I have enjoyed

getting a chance to explore the various nooks and crannies of the San Gabriel, San Fernando, and Santa Monica mountains together. They serve as role models not just for my career, but for how I treat other people. My mother's endless ability to forgive and see the best in people, and my dad's loyalty and pragmatism shape my decision making today. I am who I am today because of them, and for that I will be forever grateful;

My wife Rachel. Coming to Caltech five years ago, I did not expect to meet a spouse. She has turned what could have been a difficult experience into the best five years of my life. My life is richer because of her: our trips to baseball games, breweries, and various locales in LA are highlights of my time here. Our new dog is pretty great too. I expected to move to California and get a degree. Instead, I got that and a soulmate.

Abstract

This thesis is presented in two parts. The first part, Chapter 2, 3, and 4, offers a series of studies on noncovalent interactions in Ligand Gation Ion Channels (LGICs). The second part describes a series of studies involving the synthesis and characterization of cyanine dyes. The common thread in this work is the use of Density Functional Theory (DFT) to study chemical-scale phenomenon. Chapter 1 offers a brief introduction to DFT and a comparison with other traditional computational chemistry methodology; Hartree-Fock (HF). Summaries of the use of DFT to study both noncovalent interations and electronically excited states are also presented. In addition, the author comments on the correct application of DFT.

Chapter 2 details a computational study of the cation-π interaction of complex cations to substituted benzenes and indoles. The cation-π interaction is the electrostatic interaction between a cation and the negative electrostatic potential on an aromatic ring originating from its large permanent quadrupole moment. This chapter, in addition to establishing the correct computational parameters, establishes a large set of substituent effects with which to study cation-π interactions *in vivo*. These binding energy values are compared to previous applications of cation-π binding energies from our lab, and it was found that the derived binding energies are sufficiently accurate.

Chapter 3 applies the foundational knowledge from the previous chapter to study cation-π interactions of cationic ligands to multiple aromatics. This is a common motif *in vivo* known as the aromatic box. Using this methodology, it is established that cation binding in this form is cooperative. Further, many aromatic boxes from crystal structures were evaluated energetically.

Chapter 4 describes work to develop a new amino acid to study hydrogen bonds in *Xenopus laevis* oocytes. These fluorinated aliphatic amino acids inductively attenuate the hydrogen bond accepting ability of the carbonyl. This new strategy was used to probe for a hydrogen bond between the indole NH α4 TrpB and a backbone carbonyl associated with L119 on the β2 subunit of the α4β2 nicotinic acetylcholine receptor (nAChR). The fluorinated amino acids were validated computationally and with NMR studies. This new strategy showed that the α4-β2 interfacial hydrogen prediction was false.

Chapter 5 describes the synthesis and characterization of a series of *meso*-aromatic-acetylene cyanine dyes which feature a very large Stokes shift. Synthesis of the dyes features a key Sonagashira reaction. These dyes are investigated photophysically and computationally using time dependent DFT (TDDFT). The mechanism for this Stokes shift is an excitation to the S2 state, relaxation to the S1 state, and normal cyanine fluorescence.

Chapter 6 describes three separate strategies to construct a cyanine-based photocage to release drugs *in vivo* using an *ortho*-quinone methide strategy. One strategy utilized an acetylene-aromatic cyanine dye much like those described in Chapter 5, the second utilized an ethynyl-trimethylphenyl cation dye, and the third a photoinduced electron transfer cyanine dye. None of these strategies produced a usable photocage. The failure of these strategies are ascribed to both the short excited state lifetime of cyanine dyes and the direction of the transition dipole moment.

Finally, three appendices are presented. Appendix A describes early work to synthesize and characterize a *meso*-hydroxy substituted Cy5 dye. Appendix B offers many of the same computations as Chapters 2 and 3 using HF instead of DFT. Appendix C describes orbital mixing of cyanine dyes from Chapter 5 using HF instead of DFT.

Published Content and Contributions

Matthew R. Davis and Dennis A. Dougherty. "Cation-π Interactions: Computational Analyses of the Aromatic Box Motif and the Fluorination Strategy for Experimental Evaluation." *Physical Chemistry Chemical Physics*, 2015, **17**, 29262 – 29270 doi: 10.1039/c5cp04668h

Personal Contribution: The author performed all the computation and data analysis for this paper and contributed to research direction.

Table of Contents

Chapter 1: Application of density functional theory

1.1: Computational chemistry and density functional theory

1.1.1: Introduction to computational chemistry and the Hartree-Fock equation

The accurate prediction of molecular structure and properties is an ongoing goal of the computational community. Further, computational accuracy has the potential to be a major boon to experimentalists. Computational prediction of relevant molecular phenomena can enhance experimental evidence. For instance, computational results can be related directly to experimentally-derived values. A correlation between an experimental measurement and a computation can establish or support the identity of the molecular property being measured. This thesis demonstrates a number of uses for computational chemistry to elucidate physically observed phenomena. This includes the use of noncovalent binding energies in the study of protein structure and ligand binding as well as photophysical and photochemical phenomenon.

It is crucial to first understand the underlying concepts of the computational methods used. This subsection will provide a brief summary of the principles of one of the first and most established *ab initio* computational methodologies: Hartree-Fock. This topic has been considered in detail elsewhere,[1-3] so only a brief summary is presented here. Further sections will go into brief detail in the two main computational applications used in this work: DFT in noncovalent interactions and TDDFT for excited states.

Computational chemistry techniques discussed in this chapter are derived from a common methodology. The Hartree-Fock method in particular is an *ab initio* method, or

from first principles. The goal is to use various means to solve for the structure and energy of a molecule using the Schrödinger Equation:

(1) $H(\mathbf{\Psi}) = E(\mathbf{\Psi})$

where *H* is the Hamiltonian operator acting on the total wavefunction of the system $\mathbf{\Psi}$, $E(\mathbf{\Psi})$ the wavefunction multiplied by a constant *E*, the eigenvalue. The wavefunction $\mathbf{\Psi}$ in a multi-electron system (with *n* electrons) can be represented instead by a series of single-electron wavefunctions by the equation:

(2) $\mathbf{\Psi} = \psi_1 \times \psi_2 \times \ldots \psi_n$

where each single electron wavefunction ψ_n is orthogonal. Solving equation 1 is a computationally expensive task. Modern computational methods make a number of approximations in order to gain molecular insights. The Hamiltonian operator can be written as:

(3) $H = T + V$

where *T* the kinetic energy operator and *V* the potential energy operator. The kinetic energy term can be written as a function of the mass and momentum of the particle and is generated by the potential field in which the molecule inhabits. Mathematically, This can be represented by equation 4:

(4) $T = \left(\frac{1}{2m}\right) p^2$

where the momentum operator can be written as:

(5) $p = \left(\frac{\hbar}{i}\right) \nabla \;\; ; \;\; \nabla = \frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k$

and therefore, the kinetic energy operator is:

(6) $T = \left(\frac{-\hbar^2}{2m}\right) \nabla^2$

The kinetic energy of atoms and electrons are considered separately in the Hamiltonian operator. The other consideration is potential energy, given solely by variations on Coulomb's law. Potential interactions are considered between atoms, electrons and atoms, and between electrons. With these five terms, the Hamiltonian operator is:

$$(7) \qquad H = -\frac{\hbar^2}{2}\sum_A^N M_A^{-1}\nabla_A^2 + \sum_{A<B} e^2 Z_A\, Z_B\, r_{AB}^{-1} - \frac{\hbar^2}{2m}\sum_A^N \nabla^2 \; - \sum_A \sum_i e^2 Z_A\, r_{Ai}^{-1} + \sum_{i<j} e^2 r_{ij}^{-1}$$

The first term is the kinetic energy of the nuclei, the second is nuclear-nuclear repulsions, the third is the kinetic energy of the electrons, the fourth is nuclear-electron attractions, and the fifth is electron-electron repulsions. Stabilizing energies are negative, while destabilizing ones are positive. However, many terms of this equation are too complex to evaluate for systems larger than diatomic hydrogen. Therefore, a number of common approximations have been used.

First, the Born-Oppenheimer approximation assumes that the electrons instantly conform to any movement of the nuclei, as they are roughly three orders of magnitude lighter. Therefore, the nuclei are essentially frozen in the calculation, so the first term (the kinetic energy of the nuclei) can be ignored. Second, the orbital approximation defines the wavefunction $\mathbf{\Psi}$ as a product of one electron wavefunctions ψ_n. This by necessity assumes that all ψ are both normalized and orthogonal, which ignores overlap between orbitals and interaction between electrons, but does simplify the computation. The final simplification is to apply the Pauli Exclusion Principle: in most systems (and all systems studied in this work), having two sets of degenerate wavefunctions differentiated only by spin results in

a restricted wavefunction, which decreases the number of orbitals needed to compute the wavefunction by half.

One important feature of Hartree-Fock and other *ab initio* methods (and an important difference between these methods and Density Functional Theory) is the variational theorem, which states that the energy of any approximate solution to the Schrödinger equation will always be higher than the true energy. Application of this theorem results in the Hartree-Fock equation, which calculates the energy of a system by splitting up the Hamiltonian into three terms:

(8) $$F(\psi_{ij}) = H_{ij} + \sum(2J_{ij} - K_{ij})$$

where F is the Fock operator and H_{ij} is the core Hamiltonian that just contains terms for the Coulombic attraction between electrons and nuclei and the kinetic energy of electrons (the third and fourth terms in equation 7). The J and K operators are the electron coulomb and exchange integrals respectively. These two integrals attempt to capture the electron-electron interaction term. The exchange term is only nonzero for parallel electron spins. This in part simulates electron correlation, though does not fully account for it. This will mean that Hartree-Fock calculations have a limit as to how accurately they can depict a molecule. More advanced methods that include electron correlation can get around this limitation. However, these methods (such as configuration interactions (CI) or Moller-Plesset (ex: MP2)) are much more computationally expensive. In any event, the total energy for a Hartree-Fock computed system is:

(9) $$E_{TOT} = 2\sum\epsilon_i - \sum\sum(2J_{ij} - K_{ij})$$

In order to find a molecule's molecular orbitals (MOs) and solve for its energy, an iterative method is applied. First, an initial guess is input and the MOs calculated. The

resulting potential field will change the system slightly, which is then computed again. When the overall force on each electron and the distance traveled between iterations is at a minimum (usually an arbitrary cutoff), the calculation is complete. This minimization is known as a self-consistent field (SCF). It is important to note that, due to the nature of the calculation, in particular the Born-Oppenheimer approximation, the nuclei do not move. Equilibrium geometry calculations (discussed in greater length later in this work) involve a number of SCF steps at different geometries until a global geometry/energy minimum is reached. In order to model the molecular orbitals, each single electron molecular orbital is modeled as a sum of atomic orbitals:

(10) $$\psi_i = \sum_k c_{ik}\phi_k$$

where ϕ is an atomic orbital (AO) and c_i is the coefficient on that atomic orbital for one particular molecular orbital. Through the SCF computation, the correct coefficients are eventually agreed upon for each MO. In order to model orbitals, an additional mathematic approximation is used. Commonly, atomic orbitals are represented by a number of gaussian type orbitals.

In this work, the most common basis set is 6-31G(d,p). This refers to 6 gaussian orbitals used to model the core orbitals (1s on C, 1s, 2s, and 2p for Si) and the valence orbitals modeled by a split function of a linear combination of three gaussians and a more diffuse gaussian respectively. This gives greater flexibility for polarization. The (d,p) portion of the basis set means that for all heavy (non H or He) atoms, d-orbitals are added to better model polarization. In addition, p orbitals are added to H and He atoms. This basis set has been shown to be sufficient for many applications. However, there exist a number

of extended basis sets that, with a method such as Hartree-Fock, can improve the accuracy of a computation.

1.1.2: Density Functional Theory

While Hartree-Fock and other methods which by definition solve the Schrödinger equation by calculating wavefunctions, Density Functional Theory (DFT) attempts to solve for the energy and geometry of a system by finding its exact electron density (ρ). If one knows the exact electron density of a system, one can determine the energy and other properties of the system precisely. ρ is also much simpler to calculate than the total wavefunction used in the orbital approximation.[4-5] However, the exact function used to describe the electron density is not obvious. Further, because the electron density is itself a function of three dimensional space, the function describing this is actually a functional. Much of the current and early work in DFT revolves around developing better functionals. Solving for the total energy of the system using DFT uses equation 11:

$$E_{TOT} = E_T + E_V + E_J + E_{XC} \quad (11)$$

where E_T, E_V, and E_J are easily solved for. E_T represents the kinetic energy of the noninteracting reference system; similar to the electron kinetic energy of electrons from the previous section. E_V and E_J are both Coulombic energies: E_V is the attraction between the nuclei and electrons and E_J is the classical Coulombic self-energy, evaluating electron-electron repulsions. The key term in DFT calculations is the exchange-correlation term E_{XC}. Having electron correlation considered in the computation from the start is a major advantage over wavefunction-based methods. This is the main reason why DFT calculations are less expensive than advanced methodology that includes correlation.[6] The

E^{XC} term is usually broken into exchange and correlation parts. One early approximation for this is the local spin-density approximation (LSDA):

(12) $$E_{XC}^{LSDA} = \int \epsilon_{XC}\left[\rho_\alpha(\boldsymbol{r}), \rho_\beta(\boldsymbol{r})\right] d^3\boldsymbol{r}$$

where ρ_α and ρ_β are electron densities of α and β spins, and ϵ_{XC} is the exchange-correlation energy density of a uniform electron gas with the spin-densities described above. These exchange-correlation energies have been accurately described by Monte Carlo simulations.[7-8] This has been known to give better molecular geometries and vibrational frequencies than Hartree-Fock, but tends to overestimate the energies of bonds. A further advancement is to add a gradient to the uniform electron gas, which partially solves this issue.[6] However, there was no systematic way of calculating these gradients, and early efforts relied on the knowledge of the system in question. Because exchange is the dominant component of exchange-correlation energy, and it is already adequately calculated by a Hartree-Fock-like methodology, E_{XC} can be split into two components:

(13) $$E_{XC} = E_X + E_C^{LSDA}$$

where E_X is the exchange energy of a Slater determinant of a set of orbitals much like Hartree-Fock orbitals. These orbitals, because they arise from electrons in a potential and not from traditional wavefunctions, are not Hartree-Fock orbitals. The orbitals are defied as:

(14) $$-\frac{1}{2}\nabla^2\psi_i + V_{KS}\psi_i = \epsilon_i\psi_i$$

(15) $$\rho = \sum_i^N |\psi_i|^2$$

These orbitals are Kohn-Sham orbitals.[9-10] V_{KS} is the Kohn-Sham potential, and is defined in much the same way as equation 11:

(16) $$V_{KS} = T_0 + \int \rho V_{nuc} d^3r + \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} d^3r_1 d^3r_2 + E_{XC}$$

Although the Hartree-Fock and Kohn-Sham orbitals derive differently, they are numerically very similar. Splitting of the exchange and correlation integrals in this manner, while numerically sound and logical, results in poor agreement with empirical evidence in practice. This is because exchange and correlation cannot be decoupled: splitting these two integrals does not fully eliminate the error from Hartree-Fock calculations. In order to include some exchange-correlation while maintaining the mathematical simplicity of calculation of the exact exchange energy, a percentage of E_{XC} is calculated by exact exchange and the remainder with a E_{XC} integral that contains both exchange and correlation components. This has proven to be an effective approximation. In fact, even the rough "half and half" approximation made by Becke et al., wherein the E_{XC} was found half by exact exchange and the other half by the LSDA approximation showed major improvement.[11]

Functionals developed more recently have more complex solutions to this problem. Many of these functionals contain several different parameters, and so are called hybrid functionals. One of the most common, and still today one of the most popular, is B3LYP.[12] This functional is calculated by a linear combination of several parameters to model the exchange-correlation integral. These parameters include LSDA terms, exact exchange terms, and generalized gradient (GGA) terms:

$$E_{xc}^{hyb} = E_{xc}^{LSD} + a_0(Ex - E_x^{LSD}) + a_x(E_x^{GGA} - E_x^{LSD}) - a_c(E_c^{LSD} - E_c^{GGA}) \quad (17)$$

The coefficients for each of these parameters were determined by a best fit to a number of both computationally derived and empirically derived values. These functionals have a constant Hartree-Fock (HF) exchange across all distances, and are called global hybrid GGAs. Alternatively, the HF exchange can be varied across interelectron distances. One example of this is the CAM-B3LYP functional, used later in this thesis (Chapters 5

and 6).[13] It includes 19% non-local (non HF) exchange in short interelectron ranges, and 65% in long ranges. This will result in a much more accurate calculation for systems in which electrons interact over long distances in a concerted manner, such as with large aromatic or chromophore systems.

One modern set of functionals of particular relevance to this work is the collection of Minnesota Meta functionals. These functionals are named by the convention M*xy* or M*xy*-suffix. These functionals have been developed in a variety of ways, but generally utilized a very large dataset to parameterize against. Various functionals in this class have been optimized for particular applications.[14-15] For example, M06 is a functional that has been optimized for organometallic chemistry and noncovalent interactions. M06-2X is a functional that has twice as much nonlocal exchange and is parameterized only for nonmetals, whereas M06-HF has only exact exchange. Picking the correct functional for any experimental application is therefore a challenge in understanding the relative value for each application. Later in this work, a particular methodology is chosen by testing several functionals and choosing which matches relevant empirical results the best (Chapter 2).

Because experimental parameters are utilized in the formulation of the most modern and accurate functionals, DFT is not an *ab initio* method and does not conform to the variational theorem. Therefore, reduction in the energy of the system does not necessarily mean that the computation is more chemically accurate. The trade off, however, is that when applied correctly, DFT can result in the prediction of accurate molecular properties at a much lower computational cost. Further caution should be applied with evaluating the resulting orbitals; while they can look similar to orbitals that result from Hartree-Fock

orbitals, their theoretical origin is from the electron density. While they cannot be conflated quantitatively with the molecular orbitals resulting from HF, they are similar in topology. Therefore, qualitative use of these orbitals is theoretically sound.[16] Where possible in this work, orbital analysis from DFT methodology was also conducted using Hartree-Fock. (Chapter 5, Appendix 3).

1.2 Computation of noncovalent interactions using Density Functional Theory

Much of the first half of this thesis is concerned primarily with the quantification of noncovalent interactions using computational chemistry, in particular DFT (Chapters 2, 3, and 4). In addition to offering a brief description of the state of the art in computing noncovalent interactions using DFT, a few comments on the choice of functional and the type of experiments necessary to do so will be presented and discussed. It is the goal of this work to provide a resource to non-computational chemists by which one can utilize computational chemistry in their own work.

Computation of noncovalent interactions using DFT has until recently been a difficult problem, due to both an underestimation of electrostatic interactions and dispersion forces. Dispersion correction terms have been introduced into functionals to better model these interactions, but their performance outside of test datasets has been uneven.[17-18] These improved functionals are known as Grimme dispersion functionals, which add this correction to B3LYP (also noted D-B3LYP or D-DFT, with the D standing for dispersion).[19-20] The dispersion correction functionals derive from empirical parameters, much like the exchange-correlation hybrid functionals discussed in the previous section.

The introduction of much larger datasets for parameterization of specialized functionals (such as the Minnesota Meta functionals) has significantly improved the performance of DFT in the realm of noncovalent interactions.[21] The previous caveat in the above section is still important, however. Because each hybrid functional is constructed differently with different datasets, their performance in each individual system can vary greatly. For an untested system, it is important to relate a number of different functionals with a system of interest.

When using computational chemistry to supplement or explain experimental results, it is important to choose the functional and basis set based on previously validated experimental results that are as close to the computational methodology as possible. This is especially important if the functional has not been expressly parameterized for the experimental parameter being investigated. This work utilized M06 for computations involving cation-π interactions, which are primarily electrostatic. In addition to M06, B3LYP, B3LYP with the Grimme dispersion correction, M06-2X, and several wavefunction theory methods (Hartree-Fock and MP2) were tested with a small set of trial systems. These systems were gas phase binding energies of benzene and various cations; conditions very close to the computations being performed.

In our case, M06 was the best method. This did not conform to expectations based on the literature, where a D-DFT method would be favored. This is could be due to an overcorrection of the Grimme-B3LYP method because of its dispersion function: it may have underestimated the electrostatic portion of the interaction. This is an example of the importance of careful choice of functional. In addition, a smaller basis set was favored over a larger one: this result, while fortuitous in the case of computational expense, was also

surprising, and could have been due to several factors. However, the fact that empirical gas phase energies can be replicated nearly exactly by a computational method represents a large improvement in the state of the art for computation of noncovalent interactions.

1.3: Time-dependent Density Functional Theory

Solving for excited states of molecules computationally is a useful tool in the understanding of dye chromophores and photochemistry. Time-dependent Density Functional Theory (TDDFT) is homologous to time dependent wavefunction theory in that rather than attempting to solve a time-independent Schrödinger equation, a time-dependent variant is used. This significantly increases the complexity of the method. The key concept behind TDDFT is that a time dependent wavefunction can be modeled by a time dependent electron density, much the same as in normal DFT.[22] The electron density would therefore have the form:

(18) $$\rho_s(\boldsymbol{r},t) = \sum_{i=1}^{N_b} f_i(t)|\phi_i(\boldsymbol{r},t)|^2$$

This is homologous to the time-independent version of the density-space relationship. ϕ_i are a set of time dependent non-interacting Kohn-Sham orbitals that can map on to the fully interacting system. These orbitals would obey the equation:

(19) $$\left(-\frac{1}{2}\nabla^2 + V_s(\boldsymbol{r},t)\right)\phi_s(\boldsymbol{r},t) = i\frac{\partial}{\partial t}\phi_s(\boldsymbol{r},t)\ ;\ \phi_s(\boldsymbol{r},0) = \phi_s(\boldsymbol{r})$$

These TD-Kohn Sham orbitals can be therefore found in a similar manner to time-independent DFT, where the functional is split in to kinetic energy, Coulombic (between nuclei and electrons) and an exchange-correlation terms.[23] Of course, inclusion of time-dependence does significantly increase the cost of the computation. However, this does allow the computation of perturbations to an established electron density. One important

application of this is computation of excited states. Calculation of an excited state is an example of a perturbation to the ground state electron density. This is represented by the application of an external potential δV^{ext}. This can be represented as:

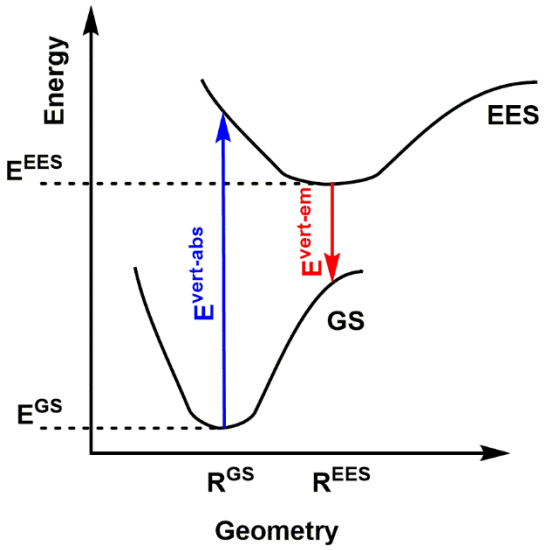


Figure 1: Simplified Jablonski diagram representing two singlet states free of intersection. One is the ground state (GS) and the other is an electronically excited state (EES)

$$V'(t) = V + \delta V^{ext}(t) \quad (20)$$

Where V is the unperturbed time dependent functional, and δV^{ext} is the external perturbation. The mathematics of the application of this external perturbation to calculation of excited states have been covered extensively elsewhere. While this calculation adds significant computational cost, the calculation of excitation energies follows relatively easily, as it requires no geometry minimization in the excited state (electronically excited state, EES) (Figure 1). The excitation energy is calculated using the equation:[24]

$$E^{vert-abs} = E^{EES}(R^{GS}) - E^{GS}(R^{GS}) \quad (21)$$

where $E^{vert-abs}$ is the excitation energy, $E^{EES}(R^{GS})$ is the energy of the electronically excited state (EES) at the ground state geometry, and $E^{GS}(R^{GS})$ is the energy of the ground state at the ground state geometry. Calculation of an emission spectrum requires geometry minimization on the excited state followed by the calculation of the energy difference between the two states.

$$E^{vert-em} = E^{EES}(R^{EES}) - E^{GS}(R^{EES}) \quad (22)$$

where $E^{vert\text{-}em}$ is the energy of emission, $E^{EES}(R^{EES})$ is the energy of the excited state at the excited state equilibrium geometry, and $E^{GS}(R^{EES})$ is the ground state energy at the excited state geometry. It is also possible to add the calculation of vibronic coupling into the perturbation to the excited state. In so doing, the oscillator strengths of each transition can be calculated. Using this method, along with appropriate solvation modeling, it is possible to recreate the spectra of well-known chromophores with relative accuracy. In addition to vibronic coupling, dipole moments of the excited and ground state can be calculated with relative ease directly from the computed electron density. This will yield the valuable measurements of transition dipole moment and its derivatives, which are a preliminary measure of the vector quantity of excited state charge transfer.

Both of these measurements rely on the solvation of the chromophore under investigation. Because excitation into a higher excited state can greatly change the electronic properties of the molecule, modeling an excitation or emission spectrum depends greatly on how it is solvated. For this reason, accurate comparison between a computed excited state and an experimentally derived one cannot be accomplished when the computation is performed in a vacuum.[24] In this thesis, however, solvation was not utilized. Instead, the properties of the computed dye molecules were compared to each other and dyes of known properties, and values such as the transition dipole moment and oscillator strength were used as the output for comparison with empirical observations. These quantities will not change substantially relative to each other when solvation parameters are added.[25-26]

As with noncovalent interactions, the choice of functional and basis set is especially important. Because B3LYP has a tendency to overestimate polarization, methods that will

vary the amount of nonlocal exchange-correlation by distance are normally used. The two most commonly used functionals for TDDFT are CAM-B3LYP (discussed above and used in this work) and ωB97X-D. Basis set choice should follow from the optimal basis set for ground state geometry. Addition of diffuse functions to a basis set is normally performed. Further, for excited state geometry minimizations, a large basis set is advisable due to the 'flatness' of the excited state energy surface.

1.4: Comments on DFT and wavefunction theory

This chapter has briefly explained the theory and applications of both DFT and wavefunction theory. Wavefunction theory, unlike DFT, is variationally correct and not based on empirical evidence, so can be considered *ab initio*. While this does make it more theoretically sound, DFT has become a much more popular methodology mainly because of its ability to recreate accurate geometries, energies, and molecular properties at a much smaller cost than wavefunction methods.

However, a recent study has shown that more complex hybrid density functionals – the most commonly used in the literature – do not accurately predict the electron density of a number of elemental cations.[27] This study has started a lively discussion of what the overall goal of DFT research is.[28-29] There is certainly an argument to be made for accurate prediction of electron density. However, in addition to several technical issues with this particular study, practical users of DFT must consider which is important: results that align with experimental observations or theoretical accuracy of electron density.

To that end, wherever possible, the Hartree-Fock calculated energies and orbitals are compared to those calculated by DFT. These energies and orbitals are shown in Appendices 2 and 3. Comparison to a variationally correct methodology, though it may not

quantitatively reproduce the energies found in the literature in training sets, is a valuable tool to verify that the chosen DFT method has some solid theoretical basis.

1.5: References

1. Anslyn, E. V.; Dougherty, D. A., *Modern Physical Organic Chemistry*. University Science Books: Saucalito, CA, 2006.
2. Cramer, C. J., *Essentials of Computational Chemistry: Theories and Models (2nd Edition)*. John Wiley & Sons: 2004.
3. Lewars, E. G., *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Springer Netherlands: 2010.
4. Geerlings, P.; De Proft, F.; Langenaeker, W., Conceptual Density Functional Theory. *Chemical Reviews* **2003,** *103* (5), 1793-1874.
5. Parr, R. G., Density Functional Theory. *Annual Review of Physical Chemistry* **1983,** *34* (1), 631-656.
6. Becke, A. D., Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *The Journal of Chemical Physics* **1992,** *96* (3), 2155-2160.
7. Ceperley, D. M.; Alder, B. J., Ground State of the Electron Gas by a Stochastic Method. *Physical Review Letters* **1980,** *45* (7), 566-569.
8. Perdew, J. P.; Wang, Y., Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **1992,** *45* (23), 13244-13249.
9. Hohenberg, P.; Kohn, W., Inhomogeneous Electron Gas. *Physical Review* **1964,** *136* (3B), B864-B871.
10. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965,** *140* (4A), A1133-A1138.
11. Becke, A. D., A new mixing of Hartree–Fock and local density-functional theories. *The Journal of Chemical Physics* **1993,** *98* (2), 1372-1377.
12. Becke, A. D., Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics* **1993,** *98* (7), 5648-5652.
13. Yanai, T.; Tew, D. P.; Handy, N. C., A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **2004,** *393* (1), 51-57.
14. Zhao, Y.; Truhlar, D. G., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and

transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts* **2008,** *120* (1), 215-241.

15. Peverati, R.; Truhlar, D. G., Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2014,** *372* (2011).
16. Stowasser, R.; Hoffmann, R., What Do the Kohn−Sham Orbitals and Eigenvalues Mean? *Journal of the American Chemical Society* **1999,** *121* (14), 3414-3420.
17. Cho, Y.; Cho, W. J.; Youn, I. S.; Lee, G.; Singh, N. J.; Kim, K. S., Density Functional Theory Based Study of Molecular Interactions, Recognition, Engineering, and Quantum Transport in π Molecular Systems. *Accounts of Chemical Research* **2014,** *47* (11), 3321-3330.
18. Hobza, P., Calculations on Noncovalent Interactions and Databases of Benchmark Interaction Energies. *Accounts of Chemical Research* **2012,** *45* (4), 663-672.
19. Ehrlich, S.; Moellmann, J.; Grimme, S., Dispersion-Corrected Density Functional Theory for Aromatic Interactions in Complex Systems. *Accounts of Chemical Research* **2013,** *46* (4), 916-926.
20. Grimme, S., Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry* **2004,** *25* (12), 1463-1473.
21. Mardirossian, N.; Head-Gordon, M., How Accurate Are the Minnesota Density Functionals for Noncovalent Interactions, Isomerization Energies, Thermochemistry, and Barrier Heights Involving Molecules Composed of Main-Group Elements? 2016.
22. Runge, E.; Gross, E. K. U., Density-Functional Theory for Time-Dependent Systems. *Physical Review Letters* **1984,** *52* (12), 997-1000.
23. Elliott, P.; Furche, F.; Burke, K., Excited States from Time-Dependent Density Functional Theory. In *Reviews in Computational Chemistry*, John Wiley & Sons, Inc.: 2009; pp 91-165.
24. Adamo, C.; Jacquemin, D., The calculations of excited-state properties with Time-Dependent Density Functional Theory. *Chemical Society Reviews* **2013,** *42* (3), 845-856.
25. Mohanty, J.; Nau, W. M., Refractive index effects on the oscillator strength and radiative decay rate of 2,3-diazabicyclo[2.2.2]oct-2-ene. *Photochemical & Photobiological Sciences* **2004,** *3* (11-12), 1026-1031.
26. Renger, T.; Grundkötter, B.; Madjet, M. E.-A.; Müh, F., Theory of solvatochromic shifts in nonpolar solvents reveals a new spectroscopic rule. *Proceedings of the National Academy of Sciences* **2008,** *105* (36), 13235-13240.

27. Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A., Density functional theory is straying from the path toward the exact functional. *Science* **2017,** *355* (6320), 49-52.
28. Kepp, K. P., Comment on “Density functional theory is straying from the path toward the exact functional”. *Science* **2017,** *356* (6337), 496-496.
29. Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A., Response to Comment on “Density functional theory is straying from the path toward the exact functional”. *Science* **2017,** *356* (6337), 496-496.

Chapter 2: Quantum mechanical investigation of the cation-π interaction: properties and substituent effects.

2.1: Abstract

The cation-π interaction is a noncovalent interaction that is essential in the pharmacology of neurotransmitters and their receptors. The interaction arises from the electrostatic attraction between a cation and the quadrupole of an aromatic. Much work has focused on studying this interaction by incorporating non-canonical amino acids *in vivo* to modulate the electrostatic potential of aromatic side chains. Recent work has also shown that cation-π interactions are useful in organic synthesis, transition state stabilization, and self-adhesion. This chapter describes computational work with the aim of further describing this interaction. First, several computational methods were evaluated for their ability to reproduce experimental cation-π binding energies. We find the DFT method M06 with the 6-31G(d,p) basis set performs best of several methods tested. The binding of benzene to a number of different cations (sodium, potassium, ammonium, tetramethylammonium, and guanidinium) was studied. Substituted derivatives of benzene and indole were studied as well, and binding energies obtained were used to reaffirm the validity of the "fluorination strategy" to study cation-π interactions *in vivo*. In addition, anomalous behavior of large diffuse cations' binding to substituted aromatics was discussed.

2.2: Introduction

The importance of studying noncovalent interactions in biological systems cannot be overstated. Nevertheless, it remains a challenging task to confirm the importance of a noncovalent interaction in a biological system, and even more challenging to quantify one. For some time, our group has been studying a particular type of interaction – the cation-π interaction – in a range of proteins and other supramolecular systems.[1-5] Our group's work in studying proteins – notably ligand-gated ion channels (LGICs) – has concentrated on the use of noncannonical amino acid mutagenesis. Commonly, cationic ligands will bind to groups of aromatic amino acids in these ion channels. In order to determine whether a putative cation-π interaction is functionally significant, these amino acids are replaced with fluorinated variants.[6-7] A loss of function that can be related to the decrease in cation-π binding is a good indication of the presence of an interaction.[8] Recent crystal structures of LGICs confirm these observations.[9-11]

Understanding the effect of aromatic fluorination on the cation-π interaction necessarily requires an understanding of the interaction itself. The cation-π interaction is a primarily electrostatic attraction between the negative face of an aromatic molecule and a cation.[4, 6] This negative electrostatic potential is generated by the benzene or indole quadrupole. The π component of the cation-π interaction therefore need not be aromatic: olefins can perform cation-π interactions as well. This effect is highly relevant in proteins, particularly in LGICs that bind positively charged neurotransmitters.

However, it has also been extensively studied outside of these contexts, and exploited for use in non-biological systems. For example, there are many examples of organic and inorganic catalysis that use the cation-π interaction to direct stereochemistry.[12]

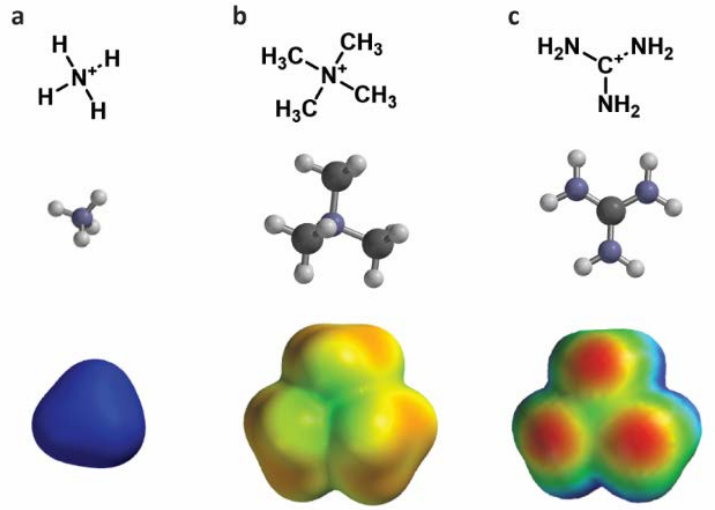


Figure 1: Cations examined in this study. (a) the ammonium ion, (b) the tetramethylammonium ion, and (c) the guanidinium ion. Pictured are molecular structures and potential energy surfaces (DFT//M06/6-31G(d,p), 95.6 to +167.3 kcal/mol)

In particular, both cation-π and anion-π (wherein an electron-deficient aromatic group can bind to an anion) noncovalent interactions have been designed in to enantioselective catalysts for a number of different reaction classes.[13-14] In addition to catalysis, the cation-π interaction is a well-studied organizational interaction in supramolecular chemistry, a subject well-studied in our lab through the synthesis and characterization of cyclophanes.[3] Recent work has also demonstrated the importance of these interaction in underwater adhesion and self-assembly.[15]

Our work on these receptors has relied on the comparison of *in vivo* collected data to *ab initio* calculated cation-π binding energies. A linear trend between the activation of a receptor by a cationic ligand and the binding energy of a sodium ion to the equivalent substituted aromatic ring (indoles to mimic the side chain of Trp or benzenes to mimic Phe/Tyr) was considered compelling evidence for a functionally relevant cation-π interaction. This "fluorination strategy" is surprisingly general. Linear plots have been seen in over 30 cases, spanning a range of proteins and ligand types. Drug-like molecules with widely differing structures have been studied, including quaternary ammonium ions (acetylcholine) and protonated amines, including primary (glycine, GABA, serotonin),

secondary (epibatidine, cytidine, varenicline) and tertiary (nicotine). In addition, more complex cations such as granisetron, ondansetron,[16] and the guanidinium toxin tetrodotoxin have shown linear fluorination plots.[17]

In contrast, a study of a guanidinium compound, meta-chlorophenyl biguanide (mCPBG) binding to the 5-HT_3 (serotonin) receptor showed behavior that was difficult to interpret.[18] Although we normally associate a lack of correlation in a fluorination plot with no functional cation-π interaction, the unusual charge distribution of guanidinium merits further study. In all cases we compared experimental data to the binding of Na^+ to the appropriate aromatics. While it may be reasonable to assume that a primary ammonium ion (RNH_3^+) is well modeled by Na^+, more complex ions such as a quaternary ammonium or a guanidinium show much different charge distributions (Figure 1) and so may display different binding behaviors.

To address this issue we have computationally evaluated fluorination effects on cation-π interactions involving the more complex cations ammonium (NH_4^+), tetramethylammonium (NMe_4^+), and guanidinium (Figure 1). Substituent effects on cation-π interactions and related noncovalent interactions involving benzene have been the subject of several recent investigations, including some with very high levels of theory.[19-21] These studies have revealed some unanticipated effects in such noncovalent interactions. The more modest goals of the present work involve the trends in cation-π binding energies in response to substitutions for various aromatics with the studied cations. When constrained to a cation-π binding geometry, these larger cations mimic the trends seen with Na^+ as probe ion.

2.3: Preliminary Computational Studies

2.3.1: Choice of computational method

We considered several different levels of theory to find which offers the best combination of accuracy and computational efficiency. Since one goal of the present work is to enable evaluation of larger systems related to protein crystal structures, we did not consider overly expensive computational methods such as CCSD, nor did we consider overly large basis sets of the sorts that have often been used to evaluate various methods. In order to evaluate these methods, we considered four prototype cation-π interactions for which experimental gas phase data are available (Table 1).[22-25] Our group had previously evaluated a number of aromatic-Na^+ binding energies for comparison to *in vivo* data.[26] We had found that a relatively simple level of theory – HF/6-31G(d,p) – performed remarkably well. As shown in Table 1, it still performs well in evaluating the relatively small Na^+ and K^+ cations. However, it does perform poorly in matching the gas-phase data for larger cations, even NH_4^+, which is experimentally similar to K^+ in the gas phase.

Surprisingly, the more computational complex and variationally correct MP2 method did not perform significantly better than HF. This could be due to an overestimation of the effect of electron correlation or orbital mixing in the interaction, whereas all available theoretical and experimental data predict that the cation-π interaction is primarily electrostatic in nature. Two density functionals were used to predict the gas phase binding energies. As previously observed, the commonly used functional B3LYP performed poorly. We had previously evaluated this method and found the results to be disappointing without the addition of complex dispersion correcting potentials.[27-32] The addition of Grimme dispersion potentials to B3LYP was considered for this study, but it did not show

Table 1 Evaluation of computational methods by comparing Binding Energies to Benzene (kcal/mol)[a]

Ion	**HF 6-31G(d,p)**	**HF 6-311+G(d,p)**	**MP2 6-31G(d,p)**	**MP2 6-311+G(d,p)**	**Expt.**
Na+	27.1	23.2	29.9	24.9	28
K+	19.3	15.6	23.8	19.5	19.2
NH4+	15.3	14.5	19	19.2	19.3
NMe4+	6.6	5.6	11.3	11.7	9.4
Error[b]	1.9	4.3	2.2	1.5	–
Ion	**B3LYP 6-31G(d,p)**	**B3LYP 6-311+G(d,p)**	**M06 6-316(d,p)**	**M06 6-311+G(d,p)**	**Expt.**
Na+	28.4	23.8	26.8	22.2	
K+	20.8	16.1	20.3	16.3	28
NH4+	17.5	15.8	19.5	18.2	19.2
NMe4+	7.7	6	10.8	10	19.3
Error[b]	1.4	3.6	0.98	2.6	9.4

[a] Computed values are ΔE; experimental values are ΔH. All data in the gas phase. [b] Mean absolute error of the four values.

improved results relative to B3LYP with no improvement to computational time. The recently established Minnesota functional M06[33] showed the closest matching with experimental results. Interestingly, the closely related M06-2x functional (containing twice the nonlocal exchange than M06), which had previously been demonstrated to be the optimal functional for noncovalent interactions, did not perform as well in predicting these gas phase binding energies. It overestimated the binding energies of Na^+ and K^+ to benzene by 2 and 4 kcal/mol respectively, so was not considered further.

The use of a larger 6-311+G(d,p) basis set gave poorer results for every method but MP2. This is likely due to a number of factors, including a cancelation of effects to some extent, but the results are consistent across the various cations studied. My own speculation is that a larger basis set, much like MP2, will overestimate dispersion and orbital mixing effects between the cation and aromatic at the expense of electrostatic effects. Given the results of Table 1, we have settled on M06/6-31G(d,p) as the optimal level of theory. It gives quite good results for the systems considered. We note that the simpler HF/6-

31G(d,p) method used previously by our group replicates all of the trends presented in this chapter. The largest differences are with the guanidinium and tetramethylammonium ions Appendix 2 includes these data and a brief analysis of HF in much of the same context as work in this chapter and Chapter 3.

2.3.2: Cation-π interactions with benzene

The cation-π binding energies and optimized geometries were calculated at M06/6-31G(d,p) for a number of different ions (Table 2). For the simple metal ions Na^+ and K^+ as well as NH_4^+, the cation binds to benzene in one conformation: directly centered in the middle of the benzene ring (*en face*). To evaluate the tetramethylammonium-benzene interaction, the complex was minimized in three separate cation-π geometries: with one, two, and three methyl groups *en face* (Figure 2). Geometries with more methyl groups *en face* had higher binding energies, consistent with previously reported work.[34] It should be noted that it is possible for NH_4^+ to bind to benzene in multiple conformations: i.e: with a different number of protons *en face*. However, because of the small differences in binding energies between these minor conformational differences, they were not considered. This could be due to the level of theory utilized. Other work has investigated these differences at various levels of theory.[35]

Table 2 Cation-π Binding Energies (kcal/mol)[a]

	Benzene	**F-Benzene**	**F_2-Benzene**		**F_3-Benzene**	
			1,3	**1,4**	**1,2,3**	**1,3,5**
Sodium	26.8	22.7	19.2	18.9	16.1	15.6
Potassium	20.3	17.7	14.6	14.4	12.0	11.7
Ammonium	19.5	16.2	12.8	12.9	10.1	10.1
Tetramethylammonium (1)	6.5	5.3[b]	4.3[b]	-	2.9[b]	3.0
Tetramethylammonium (2)	8.7	7.2[b]	5.5[b]	5.7	4.0[b]	-

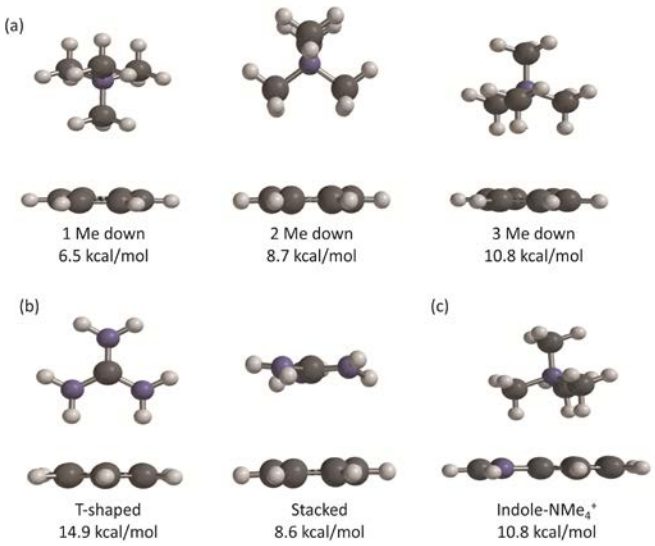


Figure 2: Examples of large cations binding to benzene. (a) Tetramethylammonium ions binding with one, two and three methyl groups towards the face of benzene. (b) Guanidinium ions binding in T-shaped and stacked conformations. (c) Tetramethylammonium binding to indole. Cation-π binding energies

Cation-π energies of guanidinium ions to benzene were calculated in two separate conformations: T-shaped and stacked (see Figure 2). As has been seen in other studies, T-shaped interactions have a much stronger binding energy.[36] Interestingly, according to M06/6-31G(d,p) the guanidinium ion adopts a propeller shape in the stacked conformation, with the amine groups adopting a 25 degree torsion angle from the plane of the cation. Interestingly, this is a similar structure to hexamethylguanidinium, which exists in two enantiomers and can be resolved by cation-π binding to cyclophanes in solution.[3] This increase in the torsional angle is likely to increase the amount of cationic surface area presented to the benzene surface, as the positive electrostatic potential is concentrated on the guanidinium protons (Figure 1).

2.3.3: The diborane-benzene interaction

An alternate cation-π – like interaction studied was the interaction between benzene and diborane, B_2H_6, in the gas phase. It is well known that diborane is formed via

a stable three-center, two-electron (3c2e) bond. This bond leaves the bridge protons somewhat electron deficient, so partially positively charged. One would assume that this would form a stable interaction with benzene. In fact, it does, though comparatively less than a cation would. This supports the initial assumption that the cation-π interaction is primarily governed by electrostatic forces (Figure 3).

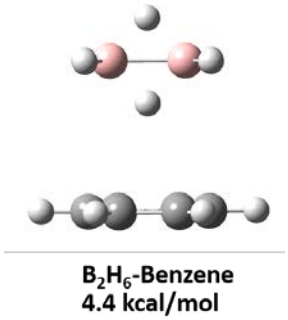


Figure 3: Diborane-benzene interaction

2.4: Substituent effects of the cation-π interaction

2.4.1: Initial fluorination studies:

In our experimental studies of a possible cation-π interaction to a Phe or a Tyr,[37] we have used 4-F-Phe, 3,5-F_2-Phe, and 3,4,5-F_3-Phe as our cation-π interaction probes, while for Trp we use the 5-, 5,7-, 5,6,7-, and 4,5,6,7-fluorotryptophans. As such, we have focused on the analogous benzenes and indoles for our computational studies (Tables 2 and 3). For the most part these studies were straightforward: the ion remained roughly centered over the ring and the cation-π binding energy progressively decreased as fluorines were introduced. The exceptions occurred with the larger ions tetramethylammonium and guanidinium binding to the fluorobenzenes and highly fluorinated indoles. In these cases full geometry optimization led to edge-on binding, with the cation attracted to the fluorine(s).

In the past, we have observed with complex cations (such as guanidinium) data that has been difficult to interpret. A study of such complex cations could be instructive in understanding *in vivo* results. We believe the level of theory employed here is adequate for computation of cation-π interactions. The edge-on interactions observed for diffuse cations

such as tetramethylammonium and guanidinium display a complex response to substitution that suggests possible challenges in evaluating potential cation-π interactions for ions of this type *in vivo*. These secondary interactions, while observed when using HF or M06, may not be best modeled by these levels of theory. Thus, emphasis should be placed on the trend in the data, rather than achieving the best quantitative evaluation of a particular binding interaction. Table 4 shows the binding energies of various cations to fluorobenzene at equilibrium geometries, indicating which cations are significantly offset from the center.

Table 4: Geometry-minimized *En face* Binding Energies (kcal/mol)[a]

	Benzene	F-Benzene
Sodium	26.8	22.7
Potassium	20.3	17.7
Ammonium	19.5	16.2
Tetramethylammonium (1)	6.5	6.3[b]
Tetramethylammonium (2)	8.7	9.0[b]
Tetramethylammonium (3)	10.8	10.7[b]
Guanidinium (T-Shaped)	14.9	12.3
Guanidinium (Stacked)	8.6	10.1[b]

[a] DFT//M06/6-31G** calculations; full geometry optimization. [b] Structure is substantially offset from the center of the ring.

Tetramethylammonium ions deviated significantly from the center of fluorobenzene, regardless of one, two, or three methyl groups *en face*. When more than one fluorine was introduced, tetramethylammonium ions did not exhibit clear local minima in a typical cation-π binding geometry. Rather, the tetramethylammonium cations favored the side of the benzene that was fluorinated (Figure 4a-c). The cation-fluorine binding appeared to be competitive with cation-π binding to the unfluorinated ring (Table 5; Figure

Table 5: *En side* Binding Energies (kcal/mol)[a]

	F-Benzene	**1,3-F_2-Benzene**	**1,2,3-F_3-Benzene**
Sodium	22.1	19.7	27.5
Potassium	17.1	14.9	21.6
Ammonium	15.9	13.3	18.5
Tetramethylammonium (1)	5.9	4.8	-
Tetramethylammonium (2)	7.9	6.5	8.4
Tetramethylammonium (3)	10.6	9.0	-
Guanidinium	14.2	12.3	14.8

[a] DFT//M06/6-31G** calculations; full geometry optimization in all cases

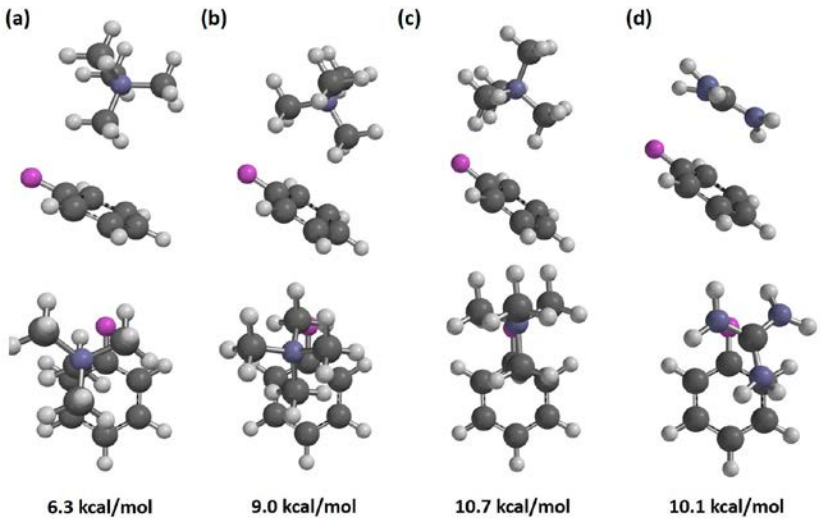


Figure 4: Geometry-minimized structures of tetramethylammonium binding with (a) one, (b) two, and (c) three methyl groups down as well as (d) guanidinium binding in a stacked fashion fluorobenzene alternate

5a-c). As expected, tetramethylammonium-fluorine binding scaled with the number of methyl groups facing the fluorine.

The stacked guanidinium conformation deviated significantly from the center of fluorobenzene, favoring the side of the benzene containing a fluorine, although the stacked arrangement was maintained (Figure 4d). However, the guanidinium-fluorine interaction with fluorobenzene has a higher binding energy than the stacked guanidinium-benzene interaction and is competitive with the T-shaped interaction. (Table 5; Figure 5d). Similar results have been seen in the binding of guanidinium to phenol. In the stacked geometry the calculated binding energy to fluorobenzene was 10.1 kcal/mol, more favorable than the stacked interaction with the parent benzene.

The T-shaped guanidinium-benzene interaction was also probed with fluorination. In fluorobenzene as well as 1,3-difluorobenzene complexes, the position of the guanidinium relative to the center of the aromatic molecule did not change. However, with 1,2,3-trifluorobenzene, the T-shaped guanidinium complex deviated significantly from the

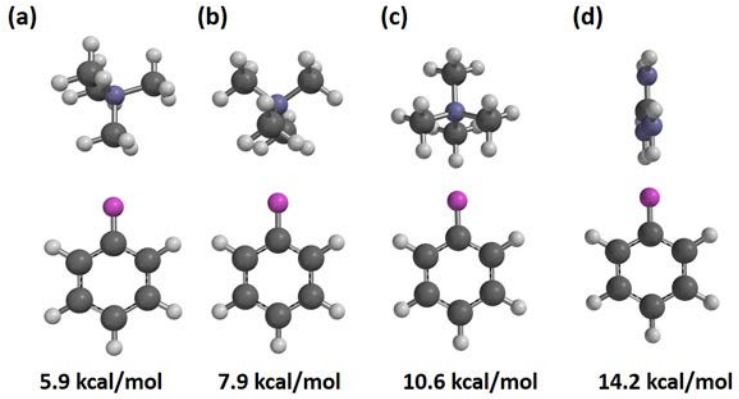


Figure 5: Binding energies of tetramethylammonium ions to the side of a fluorobenzene molecule. Binding energies computed at DFT//M06/6-31G** and are all geometry minimized structures. NMe_4^+ ions were computed with one, two and three methyl groups facing the fluorine.

unfluorinated complex, with the guanidinium ion favoring the fluorinated side of the benzene. As such, an interaction energy for this complex was obtained by the single-point method, where the binding energy was determined by adding the fluorine atoms on to the benzene at the appropriate bond lengths and taking single point energies of the complex as well as the two substituents with no geometry minimization. As with stacked interactions, symmetrically-substituted fluorinated benzenes can bind to guanidinium in a T-shaped conformation without significant deviation of the cation from the face of the aromatic molecule.

While these edge-on structures are interesting with regard to possible gas phase studies, the focus of the present work is the binding of cationic ligands to proteins, where structural constraints are likely to discourage edge-on geometries. We developed two strategies to determine the effect of fluorination on such constrained systems. First, we simply added fluorines to the ion-benzene complex (which is in an *en face* cation-π geometry) and performed single point energy (SPE) calculations. This methodology involved adding fluorines at the exact bond lengths and angles as for equilibrium geometry calculations for the fluorinated aromatics. Alternatively, we used the symmetrical 1,4- and

1,3,5-fluorobenzenes, for which the ions stayed centered over the ring. The two strategies produced similar results, with the expected trends on progressive fluorination.

2.4.2: Substituent effects calculations

Once a general method for calculating substituent effects for fluorination was established, binding energies for the Na^+, NH_4^+, and tetramethylammonium cations to various substituted benzene and indole rings were computed. In addition to providing a useful database of cation-π binding energies for use in ours and other groups' experimental studies, this work demonstrates that the method by which edge cases are calculated is general to a large variety of substituents. Most substituents' binding energies could be obtained without the use of the single point method. However, as noted in the table, several aromatics necessitated the methodology, where the cations deviated from binding the aromatic alone when geometry minimized. The single point method entails the addition of substituents at relevant geometries and bond lengths to an aromatic in a cation-π binding geometry followed by SPE calculations.

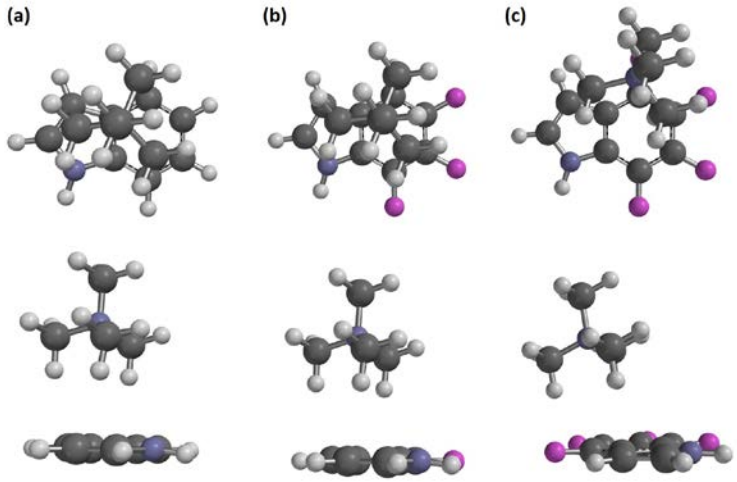


Figure 6: Geometry-minimized structures of tetramethylammonium binding with (a) indole, (b) 5,6,7-trifluoroindole, and (c) 4,5,6,7-tetrafluoroindole with alternate views of all complexes below. Binding energies computed at DFT//M06/6-31G** and all structures are geometry minimized.

With the larger and more electron-rich indole ring, the use of SPE calculations was necessary less often for diffuse cations. Only highly electron deficient rings such as 4,5,6,7-tetrafluoroindole, 5-bromoindole, and 5-cyanoindole necessitated the SPE method for this reason. A fully geometry minimized tetramethylammonium – 4,5,6,7-tetrafluoroindole gave an energy of 10.0 kcal/mol, roughly the same as for trifluoroindole (Figure 6, Table 6). The fluorine-cation binding energy of fluoroindole with tetramethylammonium is 13.8 kcal/mol. Basic nitrogen atoms also complicated the analysis, in particular in analyzing the binding of NH_4^+ cations. 7-aza-indole, N-Methylindole, and 5-aminoindole all showed an equilibrium geometry with NH_4^+ cations where the cationic nitrogen (i.e: the extra proton) was on the substituent, rather than the ligand. SPE calculations were utilized here to

Table 6: Substituent effects of cation-π interactions with various aromatic groups, representing noncannonical amino acids that can be substituted into *X. Laevis* oocytes.[a]

	Tryptophan Derivatives				**Tyrosine/Phenylalanine Derivatives**		
	Sodium	**Ammonium**	**NMe_4^+**		**Sodium**	**Ammonium**	**NMe_4^+**
Trp	32.7	25.5	15.3	**Phe**	26.8	19.5	10.8
5-F_1Trp	28.8	22.3	13.5	**4-F_1Phe**	22.9	16.2	9.0[b]
5,7-F_2Trp	25.7	19.3	11.9	**3,5-F_2Phe**	19.2	12.8	7.1[b]
5,6,7-F_3Trp	22.5	17.1	10.1	**3,4,5-F_3Phe**	16.1	10.1	5.3[b]
4,5,6,7-F_4Trp	18.9	14.4	9.1[b]	**4-BrPhe**	22.7	16.2	10.4[b]
4-F_1Trp	29.2	22.4	13.8[b]	**4-CNPhe**	16.5	10.4	5.9[b]
6-F_1Trp	28.7	21.6	13.3	**4-MePhe**	28.3	20.8	11.9
5-BrTrp	28.7	22.8	13.4[b]	**4-OMePhe**	29	21.8	12.0[b]
5-CNTrp	22.9	16.5[b]	9.8[b]				
5-MeTrp	34	26.8	16.4				
NpAla	28.7	20.8	13.5				
7-AzaTrp	27.5[b]	20.8[b]	12.9[b]				

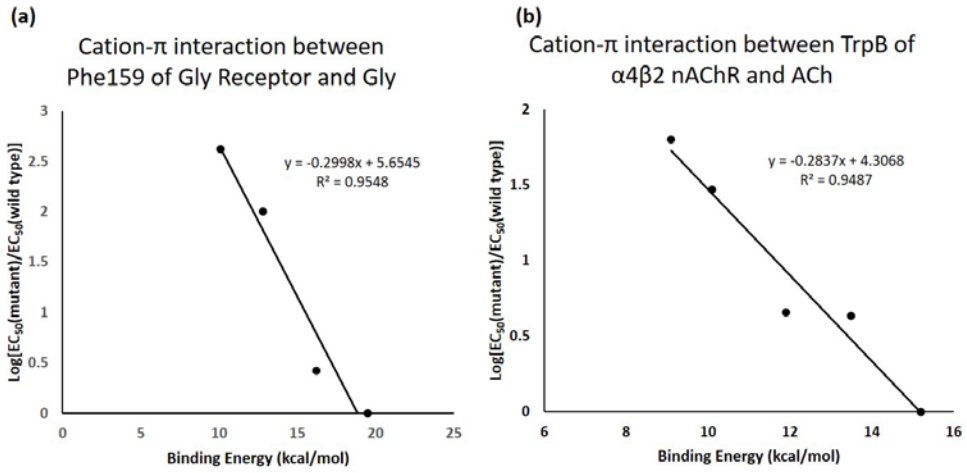


Figure 7: Cation-π binding plots for (a) Glycine at the glycine receptor using computed cation-π binding energies for ammonium ions to fluorinated benzenes using *in vivo* data collected for activation of Glycine receptor mutants

demonstrate consistency with previous results. Benzene derivatives necessitated SPE calculations with nearly every substituent when analyzing binding with the diffuse tetramethylammonium ion. The only exception for this was with the binding to toluene, a relatively electronically inactive substituent.

2.4.3: Comparison of substituent effect calculations with in vivo data

As noted above, we previously compared experimental results for a fluorination study with calculated binding energies for Na^+ to the aromatic of interest. With these new

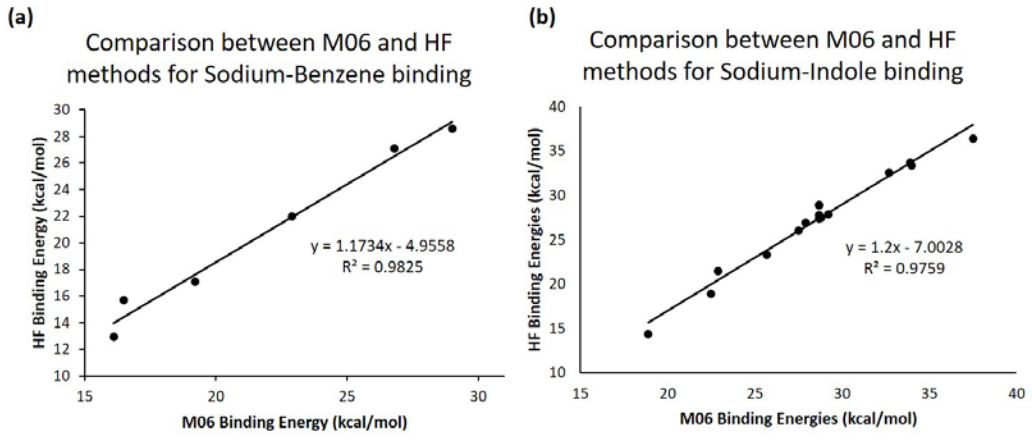


Figure 8: Comparison of binding energies of sodium ions with (a) substituted benzenes and (b) substituted

substituent effects in hand – using both an updated more accurate computational method and more accurate ligands – we can now perform more realistic comparisons. For primary amines, the previously described cation-π interaction between glycine and fluorinated phenylalanine residues in the glycine receptor[38] was re-evaluated using the results for the binding of an ammonium ion to fluorinated benzenes. As shown in Figure 7, a linear trend between agonist potency and cation-π binding energy is still obtained.

We have several demonstrations of strong cation-π interactions between ACh and nicotinic ACh receptors.[39] The relevant calculation then is tetramethylammonium binding to fluorinated indoles. Unlike with tetramethylammonium-benzene, we only considered tetramethylammonium binding to the indole ring with three methyl groups *en face* due to the challenges with a lack of symmetry. In each case, the calculated cation-π binding energies line up well with experimental results, confirming that the results of such an analysis of receptor binding are independent of the nature of the probe cation (Figure 7).

In addition to these two pictorial examples, we simulated nearly every fluorination plot that our group has studied. The receptors included in the study include several variants of the mouse muscle and nicotinic acetylcholine receptor, the 5-HT_{3A} (serotonin) receptor, as well as the $GABA_C$ and Glycine receptors. We collected the slope and R^2 regression for the plot and compared the M06-derived energies to the HF-derived ones. In these studies DFT//M06/6-31G(d,p) was used for M06 calculations, and HF/6-31G(d,p) for HF calculations. For simplicity, Na^+ was still used as the probe ion, to aid in comparison. The results show that for all Tryptophan residues studied, the change in R^2 value was 0.020±0.023, a very small deviation. The difference in slope was relatively constant, with the ratio between the M06-derived and HF-derived correlations being 1.31±0.05. This is

expected, as M06 tends to result in stronger binding energies for Na^+ ions when compared to HF. With benzene-derived aromatics, similar results were found. The differences in R^2 values was 0.048±0.047. The difference in slope between the two was similarly constant, with the result being 1.10±0.060. Regressions of the HF and M06-derived energies show that the binding energies relate to each other in a consistent manner (Figure 8). This follows with the previous analysis: M06 binding energies can be used in place of HF energies with no loss of accuracy in predicting the presence of cation-π interactions.

Table 7: Fluorination plots evaluated using both M06 and HF cation-π substituent effects. The Residue refers to the amino acid probed for a cation-π interaction. The R^2 values refer to the correlations when using either HF or M06 methods with a 6-31G(d,p) basis set with a sodium ion probe.

Receptor	Residue	Drug	R^2 (HF)	R^2 (M06)
mm nAChr	TrpB	ACh	0.959	0.946
	TrpB (L9'S)	ACh	0.938	0.884
nAChR ($(a4)_2(b2)_3$))	TrpB	ACh	0.923	0.950
	TrpB	Nicotine	0.959	0.962
	TrpB	Cytisine	0.863	0.859
	TrpB	Varenicline	0.988	0.988
nAChR ($(a4)_3(b2)_2$))	TrpB	ACh	0.929	0.925
	TrpB	Nicotine	0.991	0.990
	TrpB	Cytisine	0.971	0.968
	TrpB	Varenicline	0.949	0.946
nAChR (a6b2)	TrpB	ACh	0.961	0.927
	TrpB	Nicotine	0.647	0.708
	TrpB	TC299423	0.722	0.743
nAChR ($(a4)_2(b4)_3$)	TrpB	ACh	0.813	0.889
	TrpB	Nicotine	0.937	0.952
$5\text{-}HT_{3A}$	TrpB	Seratonin	0.994	0.992
	TrpB	5-HTQ	0.881	0.873
$GABA_C$	TyrC2	GABA	0.939	0.928
nAChR (a7)	TyrA	ACh	0.919	0.963
	TyrA	Epibatidine	0.880	0.996
TyrC2	TyrC2	Epibatidine	0.970	0.948

In addition to an HF/M06 comparison, substituent effect plots were also compared between Na^+ and tetramethylammonium ions. In this case, both used the M06/6-31G(d,p)

method. As before, for Tryptophan-based aromatics studied, the difference in R^2 was very small: 0.025±0.032. However, because the cation used has a significantly different binding energies, the derived slope is different as well. However, as expected this difference is constant at 2.01±0.13. The phenylalanine- and tyrosine-based receptors show similar ability to replicate the correlations. The change in R^2 was 0.0069±0.0022 and the change in slope was, as expected, similar to the difference with the calculations involving indole, at 2.08±0.0092. These results demonstrate that the M06-derived substituent effects are effective for utilization in free energy relationships. The regressions closely match correlations obtained with previously-used HF binding energies, and the slopes depend on both the difference of method chosen and cation.

2.5: Discussion

The interaction of more diffuse, complex cations to substituted aromatics can be complicated. Because the substituents on substituted benzenes have some electrostatic potential, cationic groups can experience multiple significant attractive and repulsive forces simultaneously. Especially in the case of the relatively smaller attractive force of a cation-π interaction on diffuse cations, cation-substituent interactions can become favorable quickly, especially as the electrostatic potential from the benzene quadrupole decreases. The key question in all of this is whether the introduction of electron-withdrawing substituents in the presence of a diffuse cationic ligand *in vivo* can have an effect on the ability to measure cation-π interactions.

The fact that our group has been able to measure many cation-π interactions with diffuse cations such as tertiary and quaternary amine ligands such as nicotine and acetylcholine suggests that the addition of fluorines onto benzene or tryptophan *in vivo*

would not affect anything more than the cation-π interaction being probed. This could be because the ligand itself is held cooperatively by 4-5 aromatic residues in an aromatic box motif (more on this in Chapter 3). This motif would largely restrict the freedom of movement of the ligand. However, a study of the 5-HT3 receptor ligand *meta*-chlorophenylbiguanide was not able to discern the presence of a strong cation-π interaction. This is an example of an especially diffuse cation. More extensive time-resolved QM-MM calculations would likely be necessary to determine if this computationally observed phenomenon is viable *in vivo*.

That being said, our results allow more reasonable modeling of the experimental studies. We find, perhaps not surprisingly, that the linear trends between receptor activation and cation-π binding energy arise regardless of the precise nature of the cation used in the calculations. These more realistic calculations do, however, allow some interesting quantitative comparisons to be made. We have several studies of ACh binding to a series of fluorinated Trp derivatives. The largest effect we have seen is a 540-fold drop in affinity on going from Trp to F_4-Trp, corresponding to 3.7 kcal/mol.[7] Our calculations for tetramethylammonium show a corresponding drop of 6.1 kcal/mol. For a simple ammonium ion, our biggest effect has been seen for GABA binding to the $GABA_A$ receptor, with a 16,500-fold drop in affinity on going from Phe to 3,4,5-F_3-Phe, corresponding to 5.75 kcal/mol.[7] Our calculations on ammonium show a corresponding drop of 8.4 kcal/mol. We consider these results to be quite encouraging. As noted above, one might expect some attenuation of a cation-π interaction on going from the gas phase to a receptor binding site. The fact that the trend between the two studies is in the right direction, with the protein binding being roughly 60% of the gas phase binding in both

cases, suggests that the present computational model may be able to provide semi-quantitative analyses of protein cation-π binding sites.

2.6: Materials and Methods

All calculations were performed using Spartan 14[40] unless otherwise stated.

Calculating Cation-π Energies:

Cation-π interactions to benzene and derivatives were evaluated with full geometry optimization at M06/6-31G(d,p)[33] with energies calculated using the equation below:

$$BE = (E_{+} + E_{\pi}) - E_{TOT}$$

where E_{TOT} is the total complex energy, E_{+} is the energy of the cation alone, and E_{π} is the energy of the aromatic system, whether it is a single aromatic or an aromatic box bereft of a cation. In this form, a positive BE signifies a favorable interaction.

Substituent Effects:

In most cases, full geometry optimization was performed for complexes of cations to fluorinated aromatics. In some cases, however, this resulted in geometries that would not be considered cation-π interactions. In such cases, fluorines were appended to the benzene ring with C-F bond distances set to values determined by optimization of the isolated aromatic system. This allowed determination of a cation-π binding energy with the cation at the position appropriate to the benzene-ion complex. We will refer to this as the single-point binding energy. The structures were then allowed to relax to the lowest-energy geometry in the gas phase. The binding energy itself was calculated using equation 1.

Several cation-π structures represented relatively shallow local minima, and were challenging to isolate. For the three separate orientations of the tetramethylammonium ion (one, two and three methyl groups facing the benzene), starting structures were the

molecular mechanics-minimized structure. This enforced a symmetry on the structure that was maintained while optimizing at a higher level of theory. To achieve a stacked conformation for guanidinium-benzene, the distance from the center of the guanidinium molecule to three meta carbons of benzene was constrained to 3.4 Å and the (constrained) equilibrium geometry was calculated at DFT//M06/6-31G(d,p). The final stacked geometry was obtained using this restricted geometry as a starting point for full geometry optimization.

2.7: References

1. Dougherty, D. A., Cation-π Interactions in Chemistry and Biology. A New View of Benzene, Phe, Tyr, and Trp. *Science* **1996,** *271*, 163-168.
2. Dougherty, D. A., The cation-pi interaction. *Accounts of chemical research* **2013,** *46* (4), 885-93.
3. Kearney, P. C.; Mizoue, L. S.; Kumpf, R. A.; Forman, J. E.; Mccurdy, A.; Dougherty, D. A., Molecular Recognition in Aqueous-Media - New Binding-Studies Provide Further Insights into the Cation-Pi Interaction and Related Phenomena. *J Am Chem Soc* **1993,** *115* (22), 9907-9919.
4. Ngola, S. M.; Dougherty, D. A., Concerning the effects of aromatic ring fluorination on the cation-pi interaction and other molecular recognition phenomena in aqueous media. *J Org Chem* **1998,** *63* (14), 4566-4567.
5. Ngola, S. M.; Kearney, P. C.; Mecozzi, S.; Russell, K.; Dougherty, D. A., A selective receptor for arginine derivatives in aqueous media. Energetic consequences of salt bridges that are highly exposed to water. *J Am Chem Soc* **1999,** *121* (6), 1192-1201.
6. Ma, J. C.; Dougherty, D. A., The Cation-π Interaction. *Chem. Rev.* **1997,** *97* (5), 1303-1324.
7. Van Arnam, E. B.; Dougherty, D. A., Functional Probes of Drug–Receptor Interactions Implicated by Structural Studies: Cys-Loop Receptors Provide a Fertile Testing Ground. *Journal of Medicinal Chemistry* **2014,** *57* (15), 6289-6300.
8. Dougherty, D. A.; Van Arnam, E. B., In Vivo Incorporation of Non-canonical Amino Acids by Using the Chemical Aminoacylation Strategy: A Broadly Applicable Mechanistic Tool. *ChemBioChem* **2014,** *15* (12), 1710-1720.

9. Olsen, J. A.; Balle, T.; Gajhede, M.; Ahring, P. K.; Kastrup, J. S., Molecular recognition of the neurotransmitter acetylcholine by an acetylcholine binding protein reveals determinants of binding to nicotinic acetylcholine receptors. *PloS one* **2014,** *9* (3), e91232.
10. Min, J.; Zhang, Y.; Xu, R. M., Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & development* **2003,** *17* (15), 1823-8.
11. Morales-Perez, C. L.; Noviello, C. M.; Hibbs, R. E., X-ray structure of the human alpha4beta2 nicotinic receptor. *Nature* **2016,** *538* (7625), 411-415.
12. Kennedy, C. R.; Lin, S.; Jacobsen, E. N., The Cation–π Interaction in Small-Molecule Catalysis. *Angewandte Chemie International Edition* **2016,** *55* (41), 12596-12624.
13. Liu, L.; Cotelle, Y.; Klehr, J.; Sakai, N.; Ward, T. R.; Matile, S., Anion-[small pi] catalysis: bicyclic products with four contiguous stereogenic centers from otherwise elusive diastereospecific domino reactions on [small pi]-acidic surfaces. *Chemical Science* **2017,** *8* (5), 3770-3774.
14. Wang, C.; Miros, F. N.; Mareda, J.; Sakai, N.; Matile, S., Asymmetric Anion–π Catalysis on Perylenediimides. *Angewandte Chemie International Edition* **2016,** *55* (46), 14422-14426.
15. Gebbie, M. A.; Wei, W.; Schrader, A. M.; Cristiani, T. R.; Dobbs, H. A.; Idso, M.; Chmelka, B. F.; Waite, J. H.; Israelachvili, J. N., Tuning underwater adhesion with cation–π interactions. *Nat Chem* **2017,** *9* (5), 473-479.
16. Duffy, N. H.; Lester, H. A.; Dougherty, D. A., Ondansetron and granisetron binding orientation in the 5-HT(3) receptor determined by unnatural amino acid mutagenesis. *ACS chemical biology* **2012,** *7* (10), 1738-45.
17. Santarelli, V. P.; Eastwood, A. L.; Dougherty, D. A.; Horn, R.; Ahern, C. A., A cation-pi interaction discriminates among sodium channels that are either sensitive or resistant to tetrodotoxin block. *Journal of Biological Chemistry* **2007,** *282* (11), 8044-8051.
18. MIles, T. F. Binding Site Structure and Stoichiometry in Serotonin Type 3 Receptors. California Institute of Technology, 2015.
19. Parrish, R. M.; Sherrill, C. D., Quantum-Mechanical Evaluation of π–π versus Substituent−π Interactions in π Stacking: Direct Evidence for the Wheeler–Houk Picture. *J Am Chem Soc* **2014,** *136* (50), 17386-17389.
20. Wheeler, S. E.; Bloom, J. W. G., Toward a More Complete Understanding of Noncovalent Interactions Involving Aromatic Rings. *The Journal of Physical Chemistry A* **2014,** *118* (32), 6133-6147.

21. Wheeler, S. E.; Houk, K. N., Substituent Effects in Cation/π Interactions and Electrostatic Potentials above the Centers of Substituted Benzenes Are Due Primarily to Through-Space Effects of the Substituents. *J Am Chem Soc* **2009,** *131* (9), 3126-3127.
22. Guo, B. C.; Purnell, J. W.; Castleman, A. W., THE CLUSTERING REACTIONS OF BENZENE WITH SODIUM AND LEAD IONS. *Chemical Physics Letters* **1990,** *168* (2), 155-160.
23. Meot-Ner (Mautner), M.; Deakyne, C. A., Unconventional Ionic Hydrogen Bonds. 1. $CH^{d+\cdots}X$. Complexes of Quaternary Ions with n- and π-Donors. *J. Am. Chem. Soc.* **1985,** *107*, 469-474.
24. Meot-Ner, M.; Deakyne, C. A., Unconventional ionic hydrogen bonds. 2. NH+.cntdot..cntdot..cntdot..pi.. Complexes of onium ions with olefins and benzene derivatives. *J Am Chem Soc* **1985,** *107* (2), 474-479.
25. Sunner, J.; Nishizawa, K.; Kebarle, P., Ion-solvent molecule interactions in the gas phase. The potassium ion and benzene. *The Journal of Physical Chemistry* **1981,** *85* (13), 1814-1820.
26. Mecozzi, S.; West Jr., A. P.; Dougherty, D. A., Cation-π Interactions in Aromatics of Biological and Medicinal Interest: Electrostatic Potential Surfaces as a Useful Qualitative Guide. *Proc. Natl. Acad. Sci. USA* **1996,** *93*, 10566-10571.
27. Grimme, S., Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry* **2004,** *25* (12), 1463-1473.
28. Johnson, E. R.; DiLabio, G. A., Structure and binding energies in van der Waals dimers: Comparison between density functional theory and correlated ab initio methods. *Chemical Physics Letters* **2006,** *419* (4–6), 333-339.
29. Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A., Application of 25 density functionals to dispersion-bound homomolecular dimers. *Chemical Physics Letters* **2004,** *394* (4–6), 334-338.
30. Kristyán, S.; Pulay, P., Can (semi)local density functional theory account for the London dispersion forces? *Chemical Physics Letters* **1994,** *229* (3), 175-180.
31. Torres, E.; DiLabio, G. A., A (Nearly) Universally Applicable Method for Modeling Noncovalent Interactions Using B3LYP. *The Journal of Physical Chemistry Letters* **2012,** *3* (13), 1738-1744.
32. Tsuzuki, S.; Lüthi, H. P., Interaction energies of van der Waals and hydrogen bonded systems calculated using density functional theory: Assessing the PW91 model. *The Journal of Chemical Physics* **2001,** *114* (9), 3949-3957.
33. Zhao, Y.; Truhlar, D., The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two

new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Account* **2008,** *120* (1-3), 215-241.

34. Pullman, A.; Berthier, G.; Savinelli, R., Theoretical study of binding of tetramethylammonium ion with aromatics. *Journal of Computational Chemistry* **1997,** *18* (16), 2012-2022.
35. Duffy, E. M.; Kowalczyk, P. J.; Jorgensen, W. L., Do denaturants interact with aromatic hydrocarbons in water? *J Am Chem Soc* **1993,** *115* (20), 9271-9275.
36. Gallivan, J. P.; Dougherty, D. A., Cation-π interactions in structural biology. *Proceedings of the National Academy of Sciences* **1999,** *96* (17), 9459-9464.
37. Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A., From ab initio quantum mechanics to molecular neurobiology: a cation-pi binding site in the nicotinic receptor. *Proc Natl Acad Sci U S A* **1998,** *95* (21), 12088-93.
38. Pless, S. A.; Hanek, A. P.; Price, K. L.; Lynch, J. W.; Lester, H. A.; Dougherty, D. A.; Lummis, S. C. R., A Cation-pi Interaction at a Phenylalanine Residue in the Glycine Receptor Binding Site Is Conserved for Different Agonists. *Molecular Pharmacology* **2011,** *79* (4), 742-748.
39. Xiu, X.; Puskar, N. L.; Shanata, J. A.; Lester, H. A.; Dougherty, D. A., Nicotine binding to brain receptors requires a strong cation-pi interaction. *Nature* **2009,** *458* (7237), 534-7.
40. Wavefunction Inc, I. C., Spartan '14.

Chapter 3: Computational studies of the aromatic box motif and cooperativity in cation-π interactions

3.1: Introduction

The cation-π interaction is a common noncovalent interaction observed in proteins. Physiologically relevant examples of this interaction have been observed in a common protein motif known as an aromatic box, where 3-5 aromatic residues bind to a single cationic ligand cooperatively. Our group has studied this motif in ligand gated ion channels (LGICs) that bind to cationic neurotransmitters such as acetylcholine, glycine, and GABA; or cationic drugs such as nicotine and varenicline. This chapter details computational work to further describe this binding motif. First, the distance dependence of the cation-π interaction is established, to determine if non-optimal interaction distances can be viable. Then, *ab initio* boxes were computed and analyzed to determine the cation-π interaction's cooperativity. Finally, a number of crystal structures containing aromatic box motifs, mostly ligand gated ion channels, were analyzed and their individual cation-π interactions quantified. This data can be used to both validate previously obtained *in vivo* data and establish subjects for further study.

3.2: Introduction:

Cation-π interactions are very common in proteins. A study from our lab has demonstrated that a cation-π interaction between amino acid residues occurs in roughly one in every seventy seven residues in the Protein Data Bank (PDB).[1] One of these examples is shown in Figure 1a.[2] Single cation-π interactions have also been observed for drug-protein binding. One example is the interaction between CDK8, a transcription factor, and cortinistatin A. Here, Trp105 and the tertiary amine in cortinistatin A make a relatively strong (9.8 kcal/mol) cation-π interaction (Figure 1b).[3] The energies of these cation-π interactions (calculated by single point energies using the same computational methodology discussed in Chapter 2) suggest that energetically-significant cation-π interactions can be an organizing force in biology.

For a number of years, we have been studying this interaction in a range of proteins, most typically neuroreceptors and ion channels.[4-7] Only in recent years has structural information become available for such systems.[5, 8-10] Remarkably, the available structures point to a general structural motif for cation-π interactions, the so-called aromatic box. This is formed by 3 to 5 aromatic side chains that can encapsulate the positive charge of a cationic molecule. This molecular cage has been seen in many receptors and in other

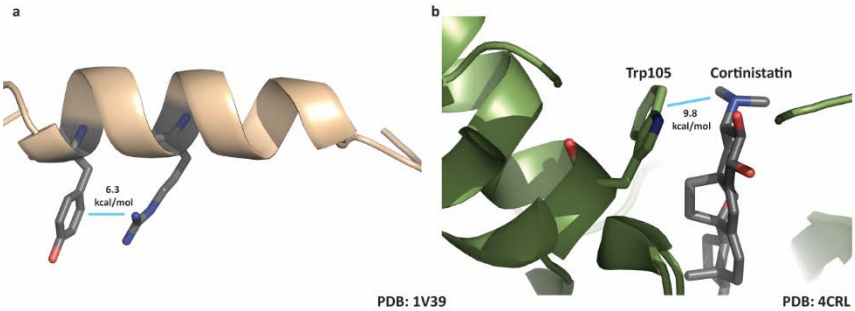


Figure 1: Cation-π interactions in proteins. (a) The interaction between an arginine and tyrosine in an alpha helix in the Vaccinia Virus Protein VP39. (b) The interaction between a tryptophan in the CDK8 transcription factor and cortinistatin A

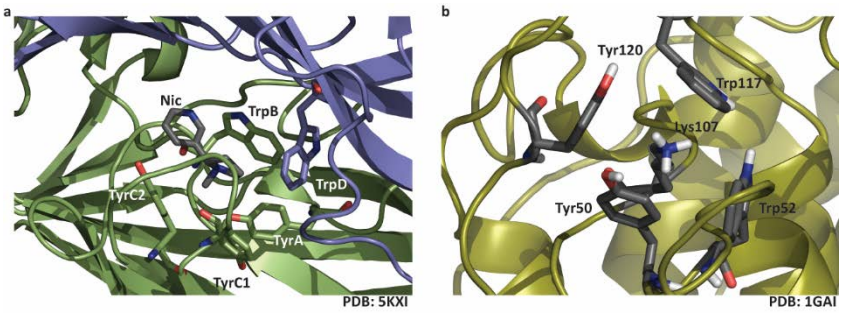


Figure 2: Examples of aromatic boxes in biology. (a) The α4β2 nicotinicic acetylcholine receptor (nAChR) bound to Nicotine by five aromatic amino acids (labeled)[18]. (b) Glucoamylase-471 structure featuring an organizing intramolecular aromatic box.[19]

binding sites for cationic guests.[11-13] It is not known what the advantage of this motif is, nor is it clear to what extent 3-5 aromatics can simultaneously contribute to the binding of a cation.

Most aromatic boxes our lab has studied have similar structures: aromatic amino acids from loops in at the interface of two subunits in a ligand gated ion channel (LGIC) contribute to bind a cation.[14-17] The residues are named for the loops they are located on (Figure 2a, PDB: 5KXI[18]). These aromatic residues and the loops they are located on are conserved among a large number of these neuroreceptors. In addition to LGICs, the aromatic box motif is observed in a number of different proteins. In our group's study of the PDB, several stood out.[1] For example, in the Glucoamylase-471 protein (PDB: 1GAI), four aromatic amino acids from two loops cooperatively bind a lysine residue to tie together a significant portion of the tertiary structure (Figure 2b).[19] The aromatic box motif is also well-established as a binding site for trimethyllysine. This is a common posttranslational modification for histones. Many histone binding proteins will contain a cluster of aromatic residues that can tightly bind to the quaternary ammonium of trimethyllysine.[20-24]

A notable difference between the cation-π interactions studied computationally in the previous chapter and those of this chapter is one of geometry. All of the unsubstituted cation-benzene and cation-indole calculations were geometry optimized in the gas phase,

so they reflect the lowest energy structure with only the cation and aromatic. Most aromatic boxes do not exhibit such ideal binding geometries. Figure 3 shows two aromatic boxes: the acetylcholine binding protein (AChBP, PDB: 3WIP[25]) and a Histone Binding Protein (PDB:1PFB[26]) with the distances between the cation and protein residues indicated. This chapter establishes the bounds by which a cation-π interaction can be considered relevant by considering the distance dependence of the interaction, then further applies this to both *ab initio* generation of aromatic box analogues and the study of crystal structures of aromatic boxes.

One further consideration for this chapter is that our lab's observation of *in vivo* aromatic boxes normally only reveals one functionally significant cation-π interaction in an aromatic box.[7, 27] For example, the only functionally relevant cation-π interaction in most nicotinic acetylcholine receptors (nAChRs) is an aromatic residue on the B loop.[28] For the most common examples, this residue is a tryptophan. However, there are five residues in this aromatic box: two from loop C (TyrC1 and TyrC2), one from loop A (TyrA), and a residue from loop D, on the complimentary subunit (TrpD). Simple observation of the aromatic box motif suggests that multiple aromatic residues must be necessary for cooperative binding of a cationic ligand or amino acid. This chapter will examine

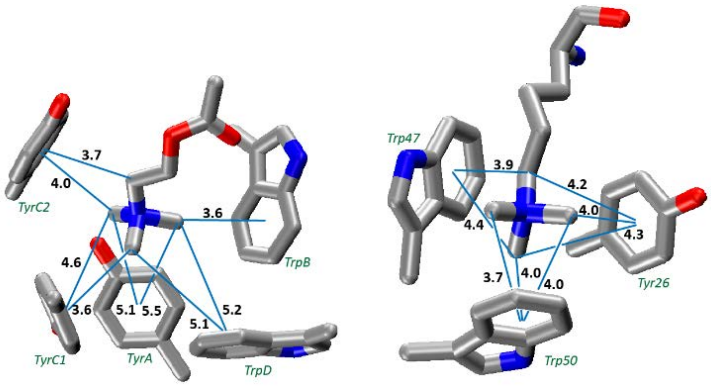


Figure 3: The aromatic box motif binding sites of (a) ACh binding to AChBP (PDB: 3WIP)[25]. Residues are labeled according to their conventional designation in the nAChR. (b) Trimethyllysine binding to the polycomb chromodomain from Drosophila histone H3 protein (PDB: 1PFB).[26] Distances (Å) are from a carbon to the ring centroid; the calculated optimal distance for such an interaction is 3.6Å

crystal structures of these aromatic boxes and compare to *in vivo* data in order to attempt to resolve this apparent contradiction.

3.3: Establishing cooperativity of the aromatic box

3.3.1: Distance dependence calculations

In biological systems cations might bind to aromatic rings at non-optimal geometries (the optimum being essentially van der Waals contact), due to other constraints imposed by the protein and/or ligand. This might be especially true for the aromatic box motif, as discussed below. As such, it is important to probe non-optimal binding geometries. To evaluate this situation, we have determined the distance dependence of a cation-π interaction to benzene for Na^+, K^+, NH_4^+, and tetramethylammonium (1-methyl *en face*) to the center of the benzene ring. As shown in Figure 4, similar trends are seen, and the cation-π interaction distance dependence is not overly steep. For all three cations, displacement by a full Å from the optimal distance results in roughly a 40% drop in binding energy. When fit to a $1/r^n$ function, n values of 1.9, 2.0, 2.8, and 2.8 are obtained for Na^+, K^+, NH_4^+, and tetramethylammonium, respectively. These results indicate that even a non-optimal cation-π interaction can contribute significantly to ligand binding or protein stability. Further work on angle dependence has been carried out by other groups, with similar conclusions.

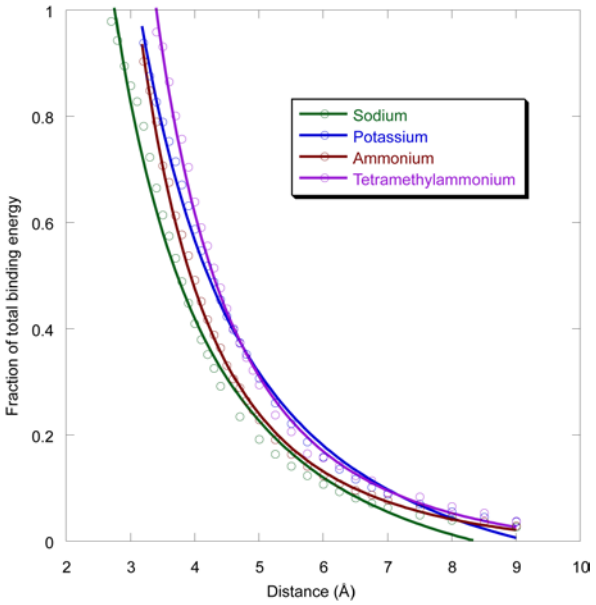


Figure 4: Distance dependence of cation-π binding for sodium, potassium, ammonium, and NMe4+ (one methyl down) ions. Binding energy is relative to the optimal binding, and distance is from the center of the benzene to Na+ or K+; N atom of ammonium; or methyl C of tetramethylammonium. Complexes with very short distances deviate from this electrostatic plot and were not included in the fitting.

Table 1 Binding energies (kcal/mol) to *ab initio* generated aromatic boxes

	Benzene Rings	Binding Energy	Theoretical[a]	Sum[b]	Discrepancy[c]
NH_4^+	3	48.7	58.5	53.4	9.8
	4	59.6	78.0	71.1	18.4
	5	62.3	97.5	-	35.2
NMe_4^+	3	30.6	32.4	31.7	1.8
	4	39.2	43.2	42.3	4.0
	5	44.0	54.0	-	10.0

[a] The binding energy of a cation-benzene complex multiplied by the number of benzenes. [b] Four times the binding energy of a single cation-box aromatic calculated. [c] The difference between the theoretical binding energy and the calculated binding energy. Binding energies computed at M06/6-31G(d,p)

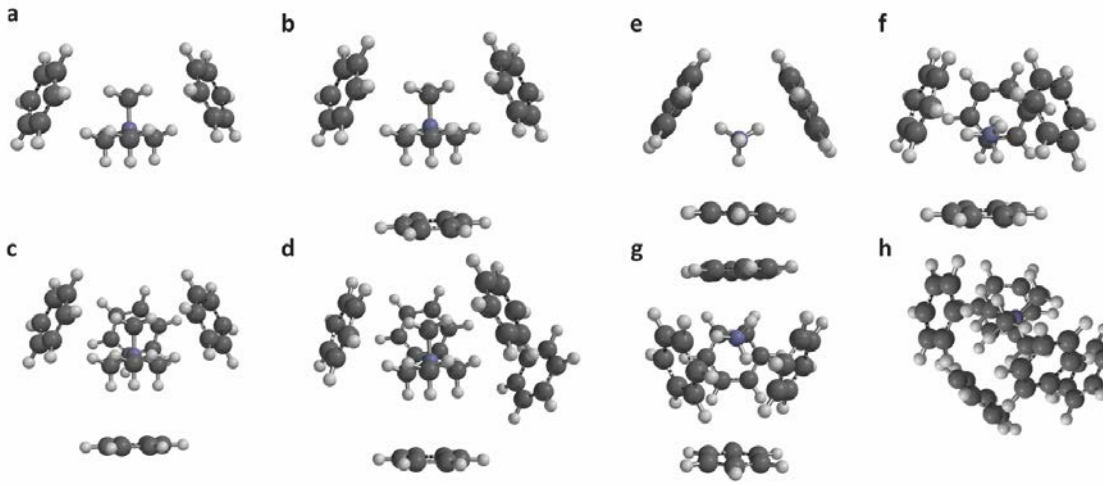


Figure 5: Geometry-optimized complexes of (a-d) tetramethylammonium or (e-h) ammonium cations to (a) two, (b,e), three, (c,f) four, or (d,g,h) five benzenes. Binding energies computed at M06/6-31G(d,p).

3.3.2: Ab initio aromatic box calculations

We first probed the aromatic box by studying the simultaneous binding of 4 benzene rings to a tetramethylammonium. As seen in Table 1 and Figure 5a-c, a tetramethylammonium can easily bind to multiple benzenes simultaneously. The total binding energy is 39.2 kcal/mol. Each pairwise interaction has three methyl groups "down", and so four optimal interactions of this sort would produce a binding energy of 43.2 kcal/mol (4 x 10.8; listed as "Theoretical" in Table 1). The individual binding energies are nearly optimal, as evidenced by the "Sum" column in Table 1. Similar results are seen for 2, 3, or 5 benzenes, although the 5th benzene binds less tightly than the others. This is

because the fifth benzene breaks the symmetry of the system. That and the respective sizes of the benzene and NMe_4^+ cation restrict the effective *ab initio* aromatic box to four benzenes.

Receptors that bind primary ammonium ions (RNH_3^+) such as GABA or serotonin also have a cluster of aromatic amino acids at the agonist binding site, so we considered the binding of ammonium to an aromatic box. Again, four benzenes can fit comfortably around an ammonium ion with no obvious steric conflicts (Figure 5b). Experimental studies of the potassium cation, which is similar in size to ammonium, have shown that multiple benzenes can complex the ion.[29-31] The total binding energy of ammonium to four benzenes is 59.6 kcal/mol, but now the individual interactions are not optimal; the sum value (Four times the binding energy of a single cation-box aromatic) differs considerably from the true binding energy. The even larger difference between the theoretical value and the binding energy could reflect a small negative cooperativity due to conflicts between induced dipoles generated by the binding. Such induced dipoles are expected to be smaller when tetramethylammonium is the ion, consistent with the smaller deviation seen in Table 1.

Calculation of these three and four-aromatic boxes with indole or phenols was attempted, to better model an *ab initio* aromatic box relative to tyrosine or tryptophan. However, due to the comparatively lower symmetry of these molecules, calculations of these types were generally more expensive and less successful, and were not pursued further.

3.3.3: Calculation of the effects of edge-on benzene-benzene binding on the binding energy of a benzene-sodium interaction

Calculations such as the *ab initio* aromatic box as described above use as an 'aromatic reference' a point calculation of the benzenes forming the box without the cation. This has the advantage of canceling out any benzene-benzene edge-edge interactions (which will likely be repulsive) to calculate only the cation-π interactions. This methodology has the disadvantage of not considering any benzene-benzene interaction in potential modulation of the cation-π interaction. This section describes calculations towards describing the effect of benzene-benzene interactions on a cation-π interaction. A later section will describe the effect of hydrogen bonding on the cation in the cation-π pair in a validated interaction in biological receptors.

Recent investigations of the muscle-type nicotinic acetylcholine receptor have revealed a secondary Tyrosine residue positioned behind Tryptophan B, the primary cation-π forming residue in its aromatic box.[32] Some speculation has focused on the potential that the primary interaction is being modulated by a slipped-stacked aromatic

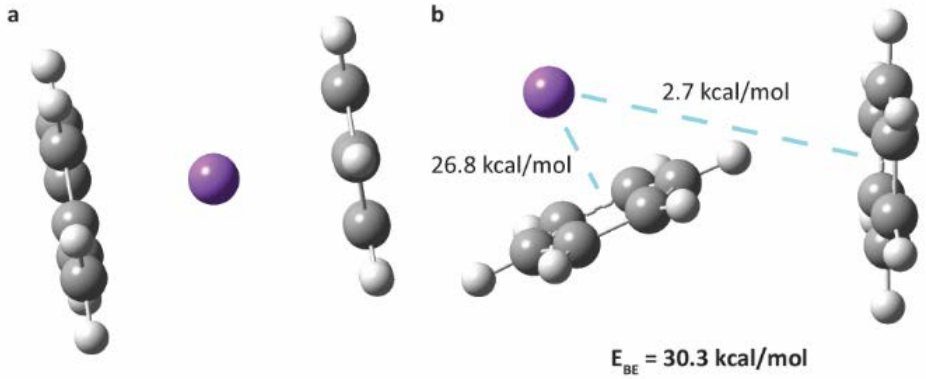


Figure 6: Sodium cation complexes with two benzenes (a) Sandwich complex, the global minimum, has a total of 49.2 kcal/mol binding energy for the Na^+ ion. (b) the combination of a T-shaped benzene interaction and a $Na^{\pm}$ benzene cation-π interaction. The cation-π interaction is not appreciably affected by the T-shaped interaction, and the total interaction energy of the sodium cation is the sum of the two.

interaction. To test this hypothesis, a proof-of-concept complex was attempted. An ideal complex would have been a slipped-stacked benzene interaction with one benzene making a cation-π interaction with Na^+. However, due to the high favorability of the sandwich complex (Figure 6a), the most useful geometry-minimized structure achieved was a T-shaped interaction between two benzenes, with one benzene forming a cation-π interaction (Figure 6b).

This complex suggests that there is no inductive effect on the cation-π interaction. The two individual sodium-benzene interactions together add to 29.5 kcal/mol, which is not significantly different from the calculated binding energy of sodium to the benzene complex of 30.3 kcal/mol. Additional benzene-benzene interactions do not have an effect on cation-π binding – this is likely because while benzene has a strong permanent quadrupole, it is not especially polarizable. It would be reasonable to expect that while the Tyr-Trp interaction discussed earlier may not affect cation-π binding, it may have a structural/organizational effect on the aromatic box itself, which could account for stronger cation-π binding interactions.

3.4: Determination of binding energies of crystal structure aromatic boxes

3.4.1: Study of a histone binding protein binding to trimethyllysine

Crystal structures have provided a valuable confirmation to the original hypothesis that cationic ligands were bound by a number of aromatic protein residues. However, as discussed above, each interaction for the aromatic box is not the optimal one. A computational methodology for systematically determining both the overall aromatic box binding energy and individual cation-π interactions was developed. The binding of a trimethyl lysine to the polycomb chromodomain from Drosophila histone H3 protein

Table 2: Binding energies (kcal/mol) of tetramethylammonium to the histone binding protein aromatic box

Complex	Binding energy
Aromatic Box	32.3
Predicted[b]	42.7
Trp50	13.9
Trp47	12.0
Tyr26	9.0
Sum	34.9

[a] M06/6-31G(d,p) calculations; single-point energies.[b] Binding energy if all interactions were optimal and independent.

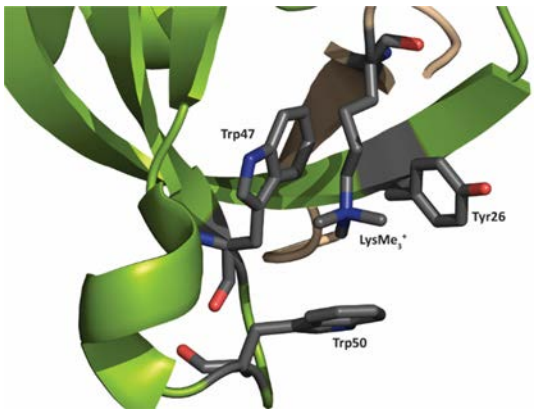


Figure 7: Crystal structure of a histone binding protein (green) bound to a histone peptide (beige) containing a trimethyllysine. Cation and aromatic box residues are grey

(PDB: 1PFB) was analyzed.[26] A total binding energy of 32.3 kcal/mol for the tetramethylammonium was found. Trp47 and Trp50 contribute the bulk of the binding energy, but Tyr26 is still significant (Table 2, Figure 7).

Based on the distance dependence of tetramethylammonium binding from Figure 4, it is evident that *all eight* potential methyl-aromatic interactions of Figure 3 contribute to the total binding energy. For the protein structure the calculated total binding energy is slightly less than the sum of the individual interactions, as was seen in the other calculated structures. Also for this system as with all previous systems, the total binding energy is much less than what would be expected if all cation-π interactions were optimal; a deviation of 10 kcal/mol is seen for the histone binding protein. As is evident from Table 3, this is a much smaller deviation than has been seen for other receptors, and is likely due to the nearly optimal geometries for the three aromatic amino acids relative to the trimethyllysine cation.

3.4.2: AChBP aromatic box binding calculations

Ligand gated ion channels, in particular the set of nicotinic acetylcholine receptors (nAChRs) are an important class of proteins that bind to cationic ligands via the aromatic box motif. There are a range of neurotransmitter-binding proteins that mimic the ligand-binding domain of nAChRs, called acetylcholine binding proteins (AChBPs).[25] These structures are much easier to crystallize, and as a result, there is a wealth of crystal structure data with a variety of ligands bound. A set of acetylcholine binding proteins (AChBPs) was studied by the same method as above. First, the acetylcholine (ACh) of an AChBP crystal structure (PDB: 3WIP) was replaced with tetramethylammonium and the total binding energy and the interactions to each individual ring were evaluated (Figure 8a, Table 3). We find a total binding energy of 28.9 kcal/mol, with four of the five aromatics contributing

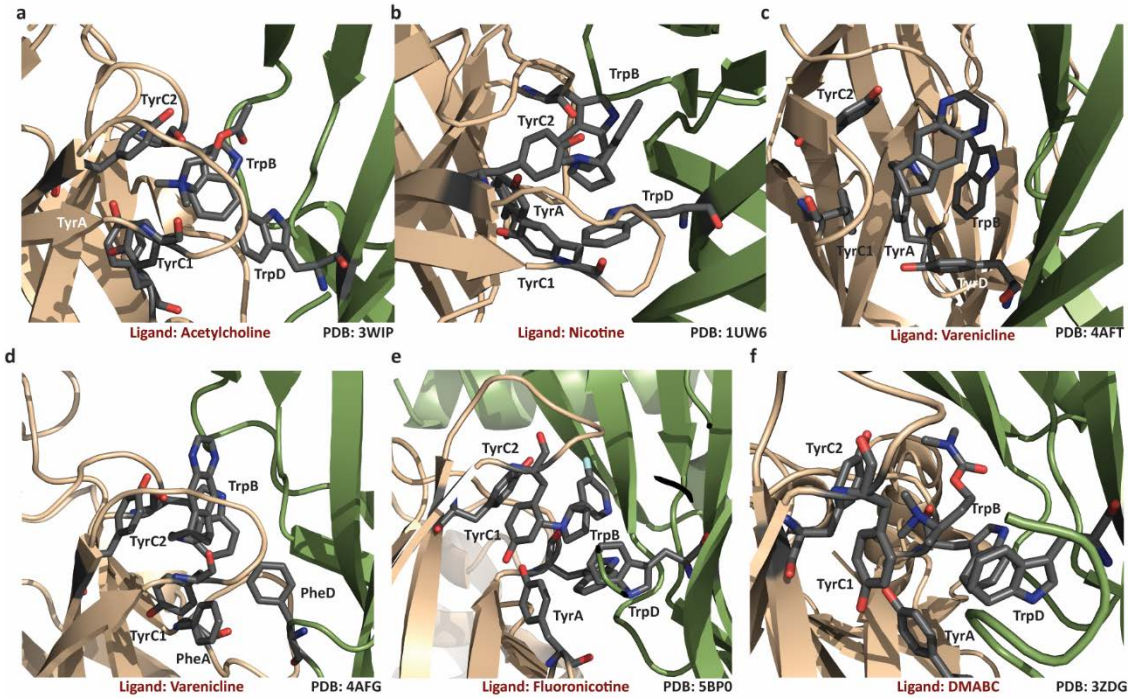


Figure 8: Acetylcholine binding protein crystal structures studied in this work. All crystal structures are of the species *Lymnaea stagnalis* with the exception of the varenicline bound structure in (d), which is from the species *Capitella teleta*. Ligands bound include (a) acetylcholine, (b) nicotine, (c,d) varenicline, (e) fluoronicotine, and (f) DMABC, a carbanylcholine derivative.

significantly. The largest contribution comes from TrpB, the residue that has been shown experimentally to make a functionally important cation-π interaction in most nAChRs. As with the previous *ab initio* structures, the calculated binding energy for the whole aromatic box is slightly less than that calculated from the individual cation-π interactions. Also, for all systems studied, the total binding energy is much less than that would have been expected if all cation-π interactions were optimal. In particular, the ACh-bound AChBP structure has a theoretical binding energy of 67.7 kcal/mol, much less than the calculated binding of 28.9 kcal/mol. This is likely due more to the deviations from an optimal interactions than any induced dipole effects.

A crystal structure functions well as a snapshot of ligand binding. However, this is a dynamic process, and multiple data points are useful for determination of optimal binding partners, especially in the context of a cooperative system. As such, a number of other AChBP structures were also evaluated, binding to different ligands. Crystal structures studied included binding of acetylcholine (PDB: 3WIP) nicotine (Figure 8b PDB: 1UW6), varenicline (Figure 8c,d PDB: 4AFG, 4AFT), fluoronicotine (Figure 8e PDB: 5BP0), and 3-(dimethylamino)butyl dimethylcarbamate (DMABC), a carbamylcholine derivative

Table 3: Binding energies (kcal/mol) of simplified ligands to the aromatic boxes of AChBPs.[a]

Ligand	Acetylcholine	Nicotine	Varenicline	Varenicline[d]	Fluoronicotine	DMABC
Aromatic Box	32.6	30.4	37.3	22.6	33.2	29.5
TrpB	8.3	11.2	10.2	10.6	11.1	8.9
TyrC1	7.5	3.9	6.4	4.2	5.2	4.2
TyrC2	8.0	6.3	6.8	7	7.4	10.1
TrpD	4.8	3.8	1.2[e]	0.4[f]	3.9	3.9
TyrA[b]	6.5	7.8	16.7	1.6[f]	8.2	4.5
Sum[c]	35.1	33.0	41.3	23.8	27.6	31.6

[a] M06/6-31G(d,p) calculations; single-point energies [b] TyrA makes a dipole-cation interaction with the ligand: the phenol hydroxyl group is pointing at the cation. [c] The sum of the individual cation-π interactions [d] A varenicline-bound AChBP from a different species than those used in this study. [e] This residue is Tyr in this structure [f] This residue is Phe in this structure.

(Figure 8f, PDB: 3ZDG).[25, 33-36] There are several other AChBP structures and derivatives currently available on the Protein Data Bank, but this study focused on AChBP structures that resembled the $(\alpha_4)_3(\beta_2)_2$ receptor configuration, as it is a well-studied receptor in our group.

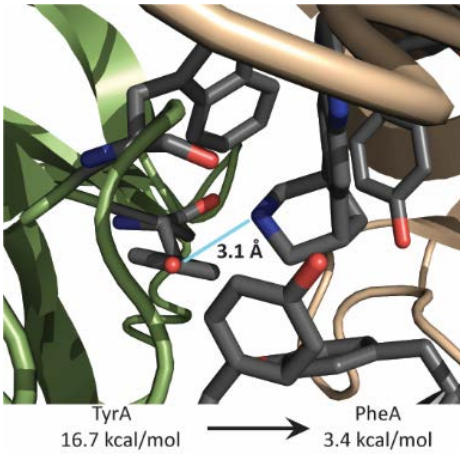


Figure 9: Strong interaction between TyrA and the varenicline cation.

The collected results (Table 3) show that there are several trends conserved between ligands. First, TrpB forms a very strong cation-π interaction, and is often among the strongest. TrpD is the opposite. In one structure (4AFG) of Varenicline binding to a different subtype of AChBP, TrpD does not participate in the aromatic box, and a Phe residue from a different part of the β subunit becomes a member of the box. Note that in each structure in Figure 8, TyrA does not appear to form a cation-π interaction with the cation, and instead is facing the cation, with the hydroxyl group pointed towards the positive charge, in a likely cation-dipole interaction. To test that this interaction is present, each Tyrosine in the 4AFT structure was 'mutated' to Phenylalanine by changing the phenol ring to benzene in the calculation. The only residue in the aromatic box which exhibits a significant difference is TyrA (Figure 9). This observation is mimicked in each of the AChBP structures studied, with the exception of 4AFG, in which the aromatic residue on loop A is already phenylalanine (Table 5).

3.4.3: Computational study of ELIC

While AChBP is a useful model for the nAChR, it is nevertheless still a binding protein, so does not contain any information about gating. The first full-length LGIC crystallized was ELIC, a prokaryotic receptor that shares structural similarities with neuronal LGICs. The full-length receptor has been crystallized in several forms with various ligands bound. Two

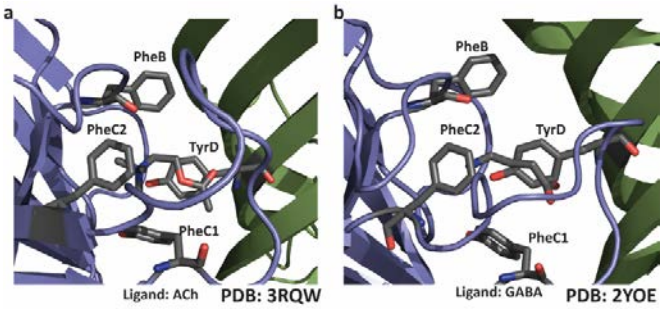


Figure 10: Full length ELIC crystalized with (a) acetylcholine and (b) GABA as ligands. (+) subunit in blue and (-) in green

Table 4: Binding energies (kcal/mol)[a] of simplified ligands to aromatic boxes of various full-length crystallized LGICs

Receptor	**ELIC**		**ELIC**		**a4b2 nAChR**	
Ligand	**GABA**		**Acetylcholine**		**Nicotine**	
	Aromatic	**BE (kcal/mol)**	**Aromatic**	**BE (kcal/mol)**	**Aromatic**	**BE (kcal/mol)**
Aromatic Box		22.8		20.2		32.7
Residue B	Phe	3.5	Phe	2.0	Trp	12.5
Residue C1	Tyr	6.6	Tyr	7.3	Tyr	3.4
Residue C2	Phe	7.3	Phe	9.1	Tyr	6.8
Residue D	Tyr	6.9	Tyr	2.8	Trp	5.4
Residue A	-	-	-		Tyr	7.3
***Sum*[b]**		*24.3*		*21.2*		*35.4*
Receptor	**GlyR**		**GlyR**		**$GABA_AR$**	
Ligand	**Glycine**		**Strychnine**		**Benzamidine**	
	Aromatic	**BE (kcal/mol)**	**Aromatic**	**BE (kcal/mol)**	**Aromatic**	**BE (kcal/mol)**
Aromatic Box		28.5		13.7		19.6
Residue B	Phe	8.2	Phe	2.7	Tyr	2.6
Residue C1	Tyr	8.2	Tyr	2.3	Phe	4.6
Residue C2	Phe	8.5	Phe	5.7	Tyr	5.1
Residue D	Phe	6.0	Phe	3.4	Tyr	4.8
Residue A	-	-	-	-	Tyr	2.1
***Sum*[b]**		*30.9*		*14.1*		*19.1*

[a] M06/6-31G(d,p) calculations; single-point energies [b] The sum of the individual cation-π interactions

structures were analyzed for this work: ELIC bound to acetylcholine (ACh, PDB: 3RQW) and γ-aminobutyric acid (GABA, PDB: 2YOE).[37-38] As before, the binding energies of both the full aromatic box and each individual cation-π interaction were calculated. These results are shown in Figure 9 and Table 4. The residues are named for their homologues to eukaryotic LGICs. For example, ELIC has a phenylalanine residue at the same relative position as TrpB.

The results suggest that the aromatic box binds to different cations in similar manners. As with AChBP and the *ab initio*-generated aromatic boxes, ligand binding appears to be cooperative. Previous fluorination studies of ELIC have shown that PheC2 and PheB respond to progressive fluorination in fluorination plots when the agonist is GABA.[38] This tracks partially with the results from our computational analysis of the crystal structure, which suggests that GABA binds most strongly with PheC2, and equally well to TyrD and TyrC1. Unlike the *in vivo* results, the crystal structure does not suggest a strong cation-π interaction between PheB and GABA. Computational analysis of the same receptor bound to ACh showed similar qualitative results.

3.4.4: Investigation of a full-length α4β2 nAChR

Recent work has yielded crystal structures of full-length LGICs with cationic ligands bound. The aromatic box appears to be conserved. Binding energies for several

Table 5: Binding energies (kcal/mol)[a] of various cations to TyrA and a modified residue without a hydroxyl group, PheA.

Receptor	**nAChR**	**AChBP**	**AChBP**	**AChBP**	**AChBP**	**AChBP**
Cation	**Nicotine**	**Acetylcholine**	**Nicotine**	**Fluoronicotine**	**Varenicline**	**Carbamylcholine**
TyrA	7.3	6.5	7.8	8.2	16.7	4.5
PheA	-0.1	2.9	2.2	2.0	3.4	1.8
Difference	7.4	3.7	5.6	6.2	13.2	2.6

[a] M06/6-31G(d,p) calculations; single-point energies

receptors' aromatic box motifs were calculated in a similar manner as above. These receptors included the $(\alpha_4)_3(\beta_2)_2$ nAChR bound to nicotine, the homopentameric 5HT-3_A receptor, the homopentameric $GABA_A$ receptor bound to benzamidine, and the human glycine receptor bound to both glycine and strychnine. I will first focus on the recently published α4β2 nAChR, which we have the most *in vivo* data to compare to (Figure 11a, Table 4).

In the α4β2 AChR structure (PDB: 5KXI), TrpB was measured as the strongest cation-π interaction, matching the computations on similar AChBP structures (see above) and previous *in vivo* studies of α4β2.[7, 18, 28] Recent studies have suggested that secondary amine ligands (such as Varenicline/Chantix and nornicotine, a demethylated nicotine derivative) can make a second cation-π interaction two TyrC2, the second strongest cation-π interaction studied.[39] As in the AChBP structures, TyrA makes a dipole-cation interaction with the ligand rather than a cation-π interaction (Table 5). Changing the phenol groups to benzene groups for each of the other two Tyr residues in the box results in much smaller losses of binding. This effect is likely due to the lower binding energy of a benzene relative to a phenol.

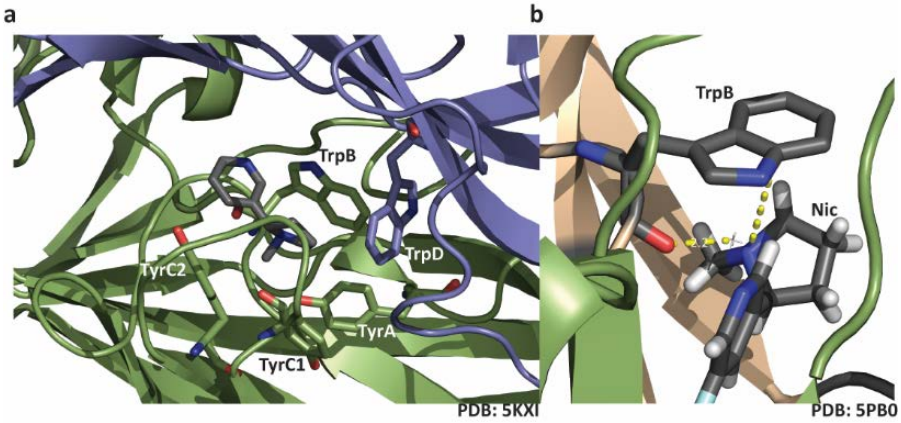


Figure 11: (a) Crystal structure of the α4β2 nAChR with aromatic box residues labeled. α subunit is in green, β subunit in blue. (b) Hydrogen bond between the N-H of the fluoronicotine cation and the carbonyl group of TrpB yellow dotted lines denote 2.2 Å (left) and 3.4 Å (left). As before, in AChBP, α subunit in beige and β in green.

Table 6: Binding energies (kcal/mol)[a] of various cations to the aromatic residue on the B loop of various LGICs or LGIC derivatives. Direct cation-π calculations were done as previously.

Interaction	nAChR Nicotine	AChBP Nicotine	AChBP Fluoronicotine	AChBP Varenicline	GlyR Glycine[c]
TrpB+Hydrogen Bond	26.0	30.4	32.1	31.6	22.6
Hydrogen Bond	14.4	19.3	22.5	23.0	15.2
TrpB, Direct	12.5	11.2	11.1	10.2	8.2
TrpB, Indirect[b]	11.7	11.1	9.5	8.6	7.3
Discrepancy	0.8	0.1	1.5	1.6	0.8

[a] M06/6-31G(d,p) calculations; single-point energies [b] Found by subtracting the calculated hydrogen bonding energy from the combined hydrogen bonding and cation-π binding to TrpB (top row). [c] The residue on loop B for this receptor is Phe

One important pharmacological difference between ACh and nicotine (as well as other nAChR agonists) is the hydrogen bond between the carbonyl of TrpB and the N-H of the pyrrolinium ring. In order to better understand this interaction, the cation-π interaction between nicotine's cationic moiety and both TrpB and the whole aromatic box was conducted in the presence and absence of the carbonyl making the hydrogen bond (Figure 11b). The results suggest that the presence of a hydrogen bond to the cationic nitrogen does not appreciably modulate the strength of the cation-π interaction. This computation was repeated for the AChBP structure with nicotine (PDB: 1UW6), fluoronicotine (PDB: 5BP0), and varenicline (PDB: 4AFT) and similar results were obtained (Table 6).

With the structures of both the α4β2 receptor and AChBP bound to nicotine in hand, the comparative accuracy of the AChBP as a nAChR receptor homologue can be analyzed. One measure of this accuracy is the comparable relative binding energies of the two structures, which has already been established. Further, the program TuftsViewer can be used to compare two crystal structures by their secondary structure homology. Using this program, the relative positions of the aromatic box amino acids can be compared. Figure

12a shows the comparison between alpha subunits of the full receptor and AChBP. The relative positions of the amino acids in the aromatic box do not change, with the exception of TyrA. The same is true for a comparison of two AChBP structures with different ligands; nicotine (PDB: 1UW6) and varenicline (PDB: 4AFT) in Figure 12b.

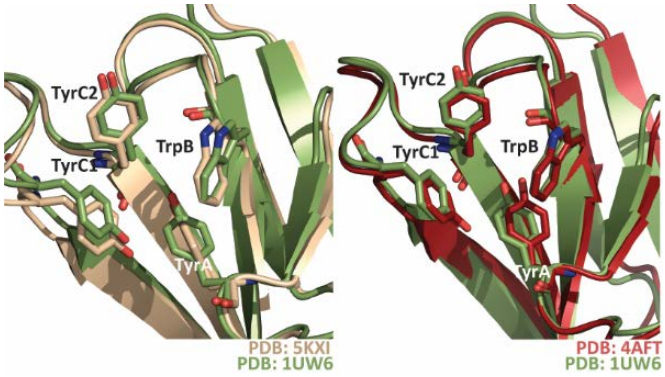


Figure 12: Structural homology comparisons of (a) the full length nAChR (beige) and AChBP (green) structures bound to nicotine and (b) two AChBP structures binding to varenicline (red) and nicotine (green).

3.4.5: Computational investigation of full-length LGICs

Other full-length neurotransmitter receptors have been recently crystallized, and subjected to similar computational analysis. The glycine receptor, an inhibitory LGIC, has been crystallized with both Strychnine (PDB: 3JAD) and glycine (PDB: 5TIO) in the neurotransmitter binding site.[40-41] Computational analysis predicts that the strongest cation-π interaction is PheC2, though TyrC1 and PheB both have similar binding energies (Figure 13, Table 3). While there is a Phe residue in the place of TyrA, it does not participate in

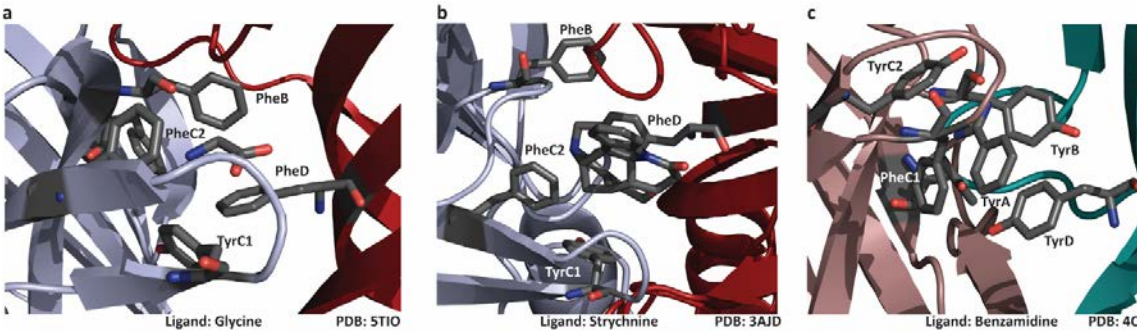


Figure 13: Crystal structures of the aromatic box domains of the Glycine receptor (GlyR) bound to (a) glycine and (b) strychnine. Blue is the (+) subunit and red is the (-). (c) The homomeric $GABA_AR$ bound to benzamidine. Salmon is the (+) subunit and teal is (-).

the aromatic box so is not included in the analysis. Computational analysis of the strychnine structure predicts that PheB will have the strongest interaction. Our lab's own studies on the glycine receptor have suggested that PheB will make the functionally significant cation-π interaction.[42]

The binding energies for the strychnine crystal structure (PDB: 3JAD) were generally weaker than with glycine (PDB: 5TIO) for two possible reasons. First, the probe ion for the glycine calculations is methylammonium, where Strychnine's is trimethylammonium. Our results from the previous chapter suggest that additional methyl groups make the positive electrostatic potential more diffuse, and will decrease the binding energy. In addition, Strychnine is a much larger ligand, so it is possible that steric hindrance may move the aromatic groups farther away from the ligand.

To test the second hypothesis, the two structures can be analyzed using the TuftsViewer program; the relative positions of the aromatic groups were then inspected. The (+) subunits of the glycine receptors were aligned using this program, and the relative positions of the aromatic box residues PheB, TyrC1, and PheC2 analyzed. The relative positions of the Strychnine-bound receptor's aromatic box residues are shifted from their positions in the glycine-bound structure. As shown in Figure 14, the displacement distances can be as much as 2 Å. This, along with the distance dependence data shown in section *3.3.1* suggests the size of the ion can have an effect of the ability of the aromatic box to bind to it.

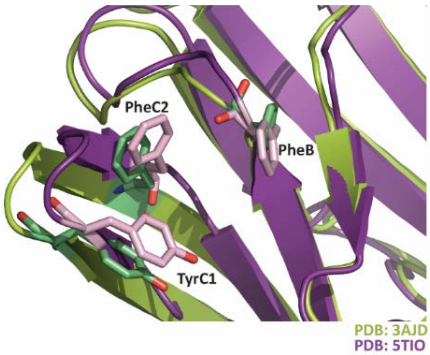


Figure 14: Structural homology comparisons of the full length GlyR structures bound to glycine (purple) and strychnine (green).

The GABA(A)R-β3 homopentamer has also been crystallized with the previously unknown agonist benzamidine in the binding site (PDB: 4COF).[43] This aromatic box was also evaluated for comparison to the others. As before, there appears to be cooperativity in binding, though the strongest cation-π interaction was to TyrC2. Oddly, the measured *in vivo* cation-π interaction for the $\alpha_1\beta_2$ $GABA_AR$ receptor showed that Tyr A made a strong cation-π interaction.[44] The computational data suggest that, similarly to the nAChR analogues studied, TyrA makes a dipole-cation interaction between the hydroxyl group of TyrA and the cationic ligand. However, both the subtype of receptor and agonist studied do not match those studied *in vivo*. This could in part explain the differences in the data, but the discrepancy merits further study.

3.5: Discussion

The aromatic box motif is common in protein cation-π binding sites. This suggests that multiple aromatics can simultaneously have favorable interactions with a cation, although it seems likely that not all interactions would be optimal in a protein structure. This prompted us to revisit the issue of the distance dependence of the cation-π interaction. As in an earlier, preliminary study, we find that the distance dependence is not steep (Figure 3).[4] For example, it could be argued that the cation-π interaction should be treated as an ion-quadrupole interaction. Ideally, an ion-quadrupole interaction should show a $1/r^3$ distance dependence for the binding energy. However, the cation-π interaction is better modeled by a $1/r^2$ distance dependence for the prototype Na+ and K+ ions. For larger ions a roughly $1/r^3$ distance dependence is seen, likely reflecting the increased importance of polarizability for the larger ions. In any case, displacement of the ion by a full angstrom

from van der Waals contact (the optimal binding arrangement) diminishes the interaction by less than half.

Two approaches have been taken to evaluate the aromatic box motif. First, we evaluated the binding of multiple benzenes to both tetramethylammonium and ammonium ions under optimal conditions. We find that four aromatics can easily bind simultaneously to either ion. For tetramethylammonium, the total binding energy is close to what would be expected if all four individual benzene tetramethylammonium interactions were optimal and independent. For ammonium there is a significant deviation from expectation based on four optimal ammonium-benzene interactions. Mostly, this is because the individual interactions are not optimal; clearly the ammonium ion is a bit too small to perfectly bind four benzenes simultaneously. There is a second, non-additivity factor that we attribute to a clash between induced dipoles in the four benzenes. Nevertheless, the global conclusion is that multiple aromatics can readily contribute to the binding of a cation through cation-π interactions.

We next considered two exemplary aromatic boxes from various protein structures. These structures focused on LGICs because of their known importance in binding cationic ligands though the aromatic box. In addition, many of the receptors with published full-length crystal structures had been studied *in vivo* by our lab, so the aromatic box amino acids that were deemed ‘functionally important’ had already been established *in vivo*. Therefore, the functionally important amino acids can be compared to the amino acids calculated to make the strongest cation-π interactions in our computation.

In each case we truncated the ligand a simpler homologue. We anticipated that all contacts between the cation and an aromatic of the box would not be optimal, and indeed

that is the case. As shown in Tables 2 and 3, the “predicted” binding energy – that which would arise if each cation-π interaction was optimal – is substantially larger than the actual binding energy. Recalling that an optimal cation-π interaction for tetramethylammonium to Tyr or Trp would be 12.5 or 15.1 kcal/mol, it is clear that each interaction in every receptor studied evaluated individually is weaker than optimal. To use two protein as an example: the histone binding protein bound to Me_3Lys (PDB: 1PFB) has three interactions approaching the ab initio-generated geometries and interaction energies, whereas some components of the AChBP box bound to ACh (PDB: 3WIP) make only relatively minor contributions to the overall binding energy. Nevertheless, it is clear that a ligand can benefit simultaneously from multiple cation-π interactions to components of an aromatic box. Gratifyingly, the largest interaction in the AChBP crystal structure is to TrpB, which is the residue most commonly involved in a functionally significant cation-π interaction in studies of nAChRs.

It is also interesting that, despite the presence of 4 cation-π interactions and one cation-dipole interaction, the AChBP shows a smaller binding energy than the histone binding protein with 3 cation-π interactions, because the cation-π interactions in the latter are individually much stronger. Experimentally, AChBP binds ACh with a ~4mM affinity, while affinities on the order of 0.3 mM have been determined for the binding of KMe_3^+ to a histone binding protein.

We also considered the extent to which the total binding energy deviates from the sum of the individual interactions. In each case, the deviation is not large, suggesting a relative lack of cooperativity (positive or negative) in the biological aromatic boxes. We conclude that while the positioning of the aromatics in the biological structures is not

optimal, a cation can still make energetically significant cation-π interactions to multiple amino acid side chains. The shallow distance dependence of the cation-π interaction presaged this observation. While not as large as is theoretically possible, total binding energies for the two aromatic boxes are quite large. Like most noncovalent interactions, the cation-π interaction is attenuated on moving from the gas phase to a condensed medium, although the attenuation is not as large as is typically associated with a hydrogen bond or an ion pair, since the desolvation penalty is less severe.[45]

These calculations do not include any adjustments that the protein structure must make to adapt to the ligand; here we are considering perfectly preorganized binding sites. This brings up an important distinction between studies such as this one and studies conducted *in vivo* in intact protein structures over a time period. These calculations capture energetics at a fixed time: somewhat of a thermal minimum for binding the ligand. However, binding a ligand and undergoing a conformational change is a dynamic process. This could in part explain the dichotomy between this work and the *in vivo* work of the Dougherty group. The 'functionally significant' cation-π interactions that can be measured by the addition of fluorines on unnatural amino acids capture the amino acids essential in the kinetic process of binding the ligand and/or gating the receptor, while the aromatic box itself is essential for achieving a fully bound state. In the cases where multiple fluorination plots are observed, multiple amino acids are important for initial ligand binding.

3.6: Materials and methods

All calculations were performed using Spartan 14 unless otherwise stated.[46]

Calculating Cation-π Energies:

Cation-π interactions to benzene and derivatives were evaluated with full geometry optimization at M06/6-31G(d,p) with energies calculated using equation 1:

$$BE = (E_{+} + E_{\pi}) - E_{TOT} \quad \textbf{(1)}$$

where E_{TOT} is the total complex energy, E_{+} is the energy of the cation alone, and E_{π} is the energy of the aromatic system, whether it is a single aromatic or an aromatic box bereft of a cation. In this form, a positive BE signifies a favorable interaction.

Distance Dependence:

Distance dependence data were obtained by constraining the distance of the ion in question to benzene and performing a constrained geometry-optimized DFT//M06/6-31G(d,p) calculation. This distance constraint was increased iteratively from the starting (optimal) geometry, and the binding energy calculated using equation 1. These data were fit using the KaleidaGraph™ software (Synergy Software).

Aromatic Box Calculations:

Ab initio aromatic boxes with a complexed ion were generated using Spartan 14. Geometry-minimized (M06/6-31G(d,p)) structures were obtained for ammonium bound to 3 or 4 benzene molecules and for tetramethylammonium binding to 3, 4, or 5 benzene molecules. The binding energies were obtained using equation 1, where E_{π} is the energy of the aromatic box without the tetramethylammonium ion.

To calculate the structure of sodium binding to two benzenes, the structures were first optimized using Gaussian 09: first a geometry minimization (HF/6-31G(d,p) followed by a secondary minimization with a higher level of theory (DFT//M06/6-31G(d,p)). Binding energies to each benzene were then calculated by deleting each benzene and

computing binding energies via single point energies (M06/6-31G(d,p)) between each benzene and the sodium ion.

Existing aromatic box motifs were imported from the Protein Data Bank using Gaussview 5, and protons were added to the structures.[47] Cation-π binding energies were calculated using Gaussian09.[48] Aromatic boxes were generated by deleting backbone and β atoms of the amino acids as well as the noncationic portion of each cationic ligand or amino acid. In most cases this generated an ammonium cation of the formula $NMe_rH_{(4-r)}^+$. The sole exception to this is the calculation arising from the benzamidine-bound $GABA_AR$ structure, where the cation used was benzamidine. The cation-π binding energy was computed by taking the single-point energy (M06/6-31G(d,p)) with and without the cation. Individual cation-π interactions between the truncated cation and aromatic residues were found by deleting all but the cation and aromatic group of interest and computing a single point energy.

Hydrogen-bond calculations

As with the previous aromatic box calculations, the cooperativity between the hydrogen bond of a protonated amine and a carbonyl on an aromatic amino acid was studied by importing the relevant amino acids from the PDB crystal structure in to Gaussview 09. Unlike with the previous experiment, the amino acid was not pruned to the aromatic. Instead, protons were added as appropriate and the carbonyl was turned into a *trans* amide by adding a tertiary amine group to the aldehyde generated by Gaussview 5 and adjusting the dihedral angle appropriately. As before, the ligand was pruned to a cation of the formula $NMe_rH_{(4-r)}^+$. The binding energies were computed by taking the single-point energies (M06/6-31G(d,p)) with and without the aromatic group. Each of those single point

energies was then computed with and without the cation. The computed hydrogen bond energy (the binding energy of the cation to the carbonyl) was subtracted from the binding energy of both to determine if there was any effect on the cation-π interaction energy from the hydrogen bond.

Structural comparison

To compare the relative positions of aromatic box amino acids, the PDB structures in input into the *TuftsView* program, which aligns the structure of single-subunit proteins by their secondary structure.[49-50] Only the alpha subunits of each protein were analyzed in this manner, as they had the majority of the aromatic box amino acids. The structures were then imported into Pymol for rendering and analysis. Distances between similar atoms in each crystal structures were measured using Pymol's software.

3.7: References

1. Gallivan, J. P.; Dougherty, D. A., Cation-π interactions in structural biology. *Proceedings of the National Academy of Sciences* **1999,** *96* (17), 9459-9464.
2. Hodel, A. E.; Gershon, P. D.; Shi, X.; Wang, S. M.; Quiocho, F. A., Specific protein recognition of an mRNA cap through its alkylated base. *Nature structural biology* **1997,** *4* (5), 350-4.
3. Pelish, H. E.; Liau, B. B.; Nitulescu, II; Tangpeerachaikul, A.; Poss, Z. C.; Da Silva, D. H.; Caruso, B. T.; Arefolov, A.; Fadeyi, O.; Christie, A. L.; Du, K.; Banka, D.; Schneider, E. V.; Jestel, A.; Zou, G.; Si, C.; Ebmeier, C. C.; Bronson, R. T.; Krivtsov, A. V.; Myers, A. G.; Kohl, N. E.; Kung, A. L.; Armstrong, S. A.; Lemieux, M. E.; Taatjes, D. J.; Shair, M. D., Mediator kinase inhibition further activates super-enhancer-associated genes in AML. *Nature* **2015,** *526* (7572), 273-6.
4. Dougherty, D. A., Cation-π Interactions in Chemistry and Biology. A New View of Benzene, Phe, Tyr, and Trp. *Science* **1996,** *271*, 163-168.
5. Dougherty, D. A., The cation-pi interaction. *Accounts of chemical research* **2013,** *46* (4), 885-93.
6. Ma, J. C.; Dougherty, D. A., The Cation-π Interaction. *Chem. Rev.* **1997,** *97* (5), 1303-1324.

7. Van Arnam, E. B.; Dougherty, D. A., Functional Probes of Drug–Receptor Interactions Implicated by Structural Studies: Cys-Loop Receptors Provide a Fertile Testing Ground. *Journal of Medicinal Chemistry* **2014,** *57* (15), 6289-6300.
8. Cheng, J.; Goldstein, R.; Gershenson, A.; Stec, B.; Roberts, M. F., The Cation-π Box Is a Specific Phosphatidylcholine Membrane Targeting Motif. *Journal of Biological Chemistry* **2013,** *288* (21), 14863-14873.
9. Meyer, E. A.; Castellano, R. K.; Diederich, F., Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angewandte Chemie International Edition* **2003,** *42* (11), 1210-1250.
10. Salonen, L. M.; Ellermann, M.; Diederich, F., Aromatic rings in chemical and biological recognition: energetics and structures. *Angew Chem Int Ed Engl* **2011,** *50* (21), 4808-42.
11. Kearney, P. C.; Mizoue, L. S.; Kumpf, R. A.; Forman, J. E.; Mccurdy, A.; Dougherty, D. A., Molecular Recognition in Aqueous-Media - New Binding-Studies Provide Further Insights into the Cation-Pi Interaction and Related Phenomena. *J Am Chem Soc* **1993,** *115* (22), 9907-9919.
12. Ngola, S. M.; Dougherty, D. A., Concerning the effects of aromatic ring fluorination on the cation-pi interaction and other molecular recognition phenomena in aqueous media. *J Org Chem* **1998,** *63* (14), 4566-4567.
13. Ngola, S. M.; Kearney, P. C.; Mecozzi, S.; Russell, K.; Dougherty, D. A., A selective receptor for arginine derivatives in aqueous media. Energetic consequences of salt bridges that are highly exposed to water. *J Am Chem Soc* **1999,** *121* (6), 1192-1201.
14. Dougherty, D. A., Cys-loop neuroreceptors: structure to the rescue? *Chemical reviews* **2008,** *108* (5), 1642-53.
15. Horenstein, N. A.; Leonik, F. M.; Papke, R. L., Multiple pharmacophores for the selective activation of nicotinic alpha7-type acetylcholine receptors. *Molecular pharmacology* **2008,** *74* (6), 1496-511.
16. Jensen, A. A.; Frolund, B.; Liljefors, T.; Krogsgaard-Larsen, P., Neuronal nicotinic acetylcholine receptors: structural revelations, target identifications, and therapeutic inspirations. *J Med Chem* **2005,** *48* (15), 4705-45.
17. Thompson, A. J.; Lester, H. A.; Lummis, S. C., The structural basis of function in Cys-loop receptors. *Quarterly reviews of biophysics* **2010,** *43* (4), 449-99.
18. Morales-Perez, C. L.; Noviello, C. M.; Hibbs, R. E., X-ray structure of the human alpha4beta2 nicotinic receptor. *Nature* **2016,** *538* (7625), 411-415.
19. Aleshin, A. E.; Stoffer, B.; Firsov, L. M.; Svensson, B.; Honzatko, R. B., Crystallographic complexes of glucoamylase with maltooligosaccharide analogs: relationship of stereochemical

distortions at the nonreducing end to the catalytic mechanism. *Biochemistry* **1996,** *35* (25), 8319-28.

20. Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L., Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proceedings of the National Academy of Sciences* **2007,** *104* (27), 11184-11188.
21. Jacobs, S. A.; Khorasanizadeh, S., Structure of HP1 Chromodomain Bound to a Lysine 9-Methylated Histone H3 Tail. *Science* **2002,** *295* (5562), 2080-2083.
22. Li, H.; Ilin, S.; Wang, W.; Duncan, E. M.; Wysocka, J.; Allis, C. D.; Patel, D. J., Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **2006,** *442* (7098), 91-95.
23. Peña, P. V.; Davrazou, F.; Shi, X.; Walter, K. L.; Verkhusha, V. V.; Gozani, O.; Zhao, R.; Kutateladze, T. G., Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **2006,** *442* (7098), 100-103.
24. Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J., How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol* **2007,** *14* (11), 1025-1040.
25. Olsen, J. A.; Balle, T.; Gajhede, M.; Ahring, P. K.; Kastrup, J. S., Molecular recognition of the neurotransmitter acetylcholine by an acetylcholine binding protein reveals determinants of binding to nicotinic acetylcholine receptors. *PloS one* **2014,** *9* (3), e91232.
26. Min, J.; Zhang, Y.; Xu, R. M., Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes & development* **2003,** *17* (15), 1823-8.
27. Dougherty, D. A.; Van Arnam, E. B., In Vivo Incorporation of Non-canonical Amino Acids by Using the Chemical Aminoacylation Strategy: A Broadly Applicable Mechanistic Tool. *ChemBioChem* **2014,** *15* (12), 1710-1720.
28. Xiu, X.; Puskar, N. L.; Shanata, J. A.; Lester, H. A.; Dougherty, D. A., Nicotine binding to brain receptors requires a strong cation-pi interaction. *Nature* **2009,** *458* (7237), 534-7.
29. Cabarcos, O. M.; Weinheimer, C. J.; Lisy, J. M., Competitive solvation of K+ by benzene and water: Cation-π interactions and π-hydrogen bonds. *The Journal of Chemical Physics* **1998,** *108* (13), 5151-5154.
30. Koyanagi, G. K.; Bohme, D. K., Kinetics and thermodynamics for the bonding of benzene to 20 main-group atomic cations: formation of half-sandwiches, full-sandwiches and beyond. *International Journal of Mass Spectrometry* **2003,** *227* (3), 563-575.
31. Sunner, J.; Nishizawa, K.; Kebarle, P., Ion-solvent molecule interactions in the gas phase. The potassium ion and benzene. *The Journal of Physical Chemistry* **1981,** *85* (13), 1814-1820.

32. Bruhova, I.; Auerbach, A., Molecular recognition at cholinergic synapses: acetylcholine versus choline. *J. Physiol.-London* **2017,** *595* (4), 1253-1261.
33. Rucktooa, P.; Haseler, C. A.; van Elk, R.; Smit, A. B.; Gallagher, T.; Sixma, T. K., Structural characterization of binding mode of smoking cessation drugs to nicotinic acetylcholine receptors through study of ligand complexes with acetylcholine-binding protein. *The Journal of biological chemistry* **2012,** *287* (28), 23283-93.
34. Billen, B.; Spurny, R.; Brams, M.; van Elk, R.; Valera-Kummer, S.; Yakel, J. L.; Voets, T.; Bertrand, D.; Smit, A. B.; Ulens, C., Molecular actions of smoking cessation drugs at alpha4beta2 nicotinic receptors defined in crystal structures of a homologous binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (23), 9173-8.
35. Bobango, J., Sankaran, B., Park, J.F., Wu, J., Talley, T.T., Comparisons of Binding Affinities for Neuronal Nicotinic Receptors (NNRs) and AChBPs. *To be published*.
36. Ussing, C. A.; Hansen, C. P.; Petersen, J. G.; Jensen, A. A.; Rohde, L. A.; Ahring, P. K.; Nielsen, E. O.; Kastrup, J. S.; Gajhede, M.; Frolund, B.; Balle, T., Synthesis, pharmacology, and biostructural characterization of novel alpha4beta2 nicotinic acetylcholine receptor agonists. *J Med Chem* **2013,** *56* (3), 940-51.
37. Pan, J.; Chen, Q.; Willenbring, D.; Yoshida, K.; Tillman, T.; Kashlan, O. B.; Cohen, A.; Kong, X. P.; Xu, Y.; Tang, P., Structure of the pentameric ligand-gated ion channel ELIC cocrystallized with its competitive antagonist acetylcholine. *Nature communications* **2012,** *3*, 714.
38. Spurny, R.; Ramerstorfer, J.; Price, K.; Brams, M.; Ernst, M.; Nury, H.; Verheij, M.; Legrand, P.; Bertrand, D.; Bertrand, S.; Dougherty, D. A.; de Esch, I. J.; Corringer, P. J.; Sieghart, W.; Lummis, S. C.; Ulens, C., Pentameric ligand-gated ion channel ELIC is activated by GABA and modulated by benzodiazepines. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (44), E3028-34.
39. Post, M. R.; Tender, G. S.; Lester, H. A.; Dougherty, D. A., Secondary Ammonium Agonists Make Dual Cation-π Interactions in α4β2 Nicotinic Receptors. *eneuro* **2017**.
40. Huang, X., Shaffer, P.L., Ayube, S., Bregman, H., Chen, H., Lehto, S.G., Luther, J.A., Matson, D.J., McDonough, S.I., Michelsen, K., Plant, M.M., Schneider, S., Simard, J.R., Teffera, Y., Yi, S., Zhang, M., DiMauro, E.F., Gingras, J., Crystal Structures of Human GlyRa3 Bound to a Novel Class of Potentiators with Efficacy in a Mouse Model of Neuropathic Pain. *To be published*.

41. Du, J.; Lu, W.; Wu, S.; Cheng, Y.; Gouaux, E., Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature* **2015,** *526* (7572), 224-9.
42. Pless, S. A.; Hanek, A. P.; Price, K. L.; Lynch, J. W.; Lester, H. A.; Dougherty, D. A.; Lummis, S. C. R., A Cation-pi Interaction at a Phenylalanine Residue in the Glycine Receptor Binding Site Is Conserved for Different Agonists. *Molecular pharmacology* **2011,** *79* (4), 742-748.
43. Miller, P. S.; Aricescu, A. R., Crystal structure of a human GABAA receptor. *Nature* **2014,** *512* (7514), 270-5.
44. Padgett, C. L.; Hanek, A. P.; Lester, H. A.; Dougherty, D. A.; Lummis, S. C. R., Unnatural amino acid mutagenesis of the GABA(A) receptor binding site residues reveals a novel cation-pi interaction between GABA and beta(2)Tyr97. *Journal of Neuroscience* **2007,** *27* (4), 886-892.
45. Gallivan, J. P.; Dougherty, D. A., A Computational Study of Cation-π Interactions vs. Salt Bridges in Aqueous Media: Implications for Protein Engineering. *J. Am. Chem. Soc.* **2000,** *122* (5), 870-874.
46. Wavefunction Inc, I. C., Spartan '14.
47. Roy Dennington, T. K., and John Millam, Gaussview, Version 5. Semichem, Inc: Shawnee Mission, KS, 2009.
48. 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.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
49. Menke, M., Visualizing Proteins with Multiple Views. University, T., Ed. 2010.
50. Matt Menke, S. S., Connor Gramazio, Caitlin Crumm, Daniela Extrum-Fernandez and Lenore Cowen, TuftsViewer: An Intuitive Interface for viewing 3D Protein Structures. *Proceedings of ISMB 3DSIG*.

Chapter 4: Progress towards the characterization and incorporation of a fluorinated aliphatic noncannonical amino acid to study hydrogen bonds in LGICs[1]

4.1: Abstract

Probing individual hydrogen bonds in ligand gated ion channels (LGICs) is a useful way to study both ligand binding and receptor gating. Recent crystal structures of the α4β2 nicotinic acetylcholine receptor (nAChR) suggest a hydrogen bond between the indole N-H of Trp149 on the B loop of the α subunit and the carbonyl of Leu119, on the E loop of the β subunit. Recent studies in our lab, namely the substitution of napthylalanine in place of TrpB, show a large gain of function of the receptor, suggesting that this hydrogen bond could be crucial in receptor gating. Because the carbonyl in question is flanked by a proline residue, the normal strategy of α-hydroxy acid substitution cannot be applied. In this work, several potential solutions are proposed and attempted before settling on the substitution of the Leu residue with fluorinated aliphatic amino acids, to include electron withdrawing groups on the side chain. After computational and *in vitro* NMR analysis, it was determined that trifluoroalanine could modulate the hydrogen bonding strength to a visible degree. Substitution at Leu119, however, did not show any change, suggesting that the putative hydrogen bond is not present. Ongoing work is being conducted to reaffirm the use of trifluoroalanine in this context in *X. Laevis* oocytes.

[1] Work presented in this chapter was borne out of a wonderful collaboration with Michael R Post, who synthesized TQOPS'-β-Alanine; performed all of the biological studies with the exception of the α7 nAChR work and the work described in subsection 4.5.2; and contributed to project direction and experimental design. Much of this work is also presented in his thesis.[12]

4.2: Introduction

To study noncovalent interactions in proteins or in protein-ligand interactions, our group has long incorporated noncannonical amino acids (ncAAs) into ligand gated ion channels.[1-2] One example of this methodology is the incorporation of fluorinated aromatic residues to study cation-π interactions (Chapter 3). Our lab has taken a similar approach to studying other noncovalent interactions or other stereoelectronic effects when studying ligand binding and gating of LGICs. Incorporation of an amino acid that specifically changes the property of interest will yield a simplified interpretation of results: any change in ligand binding or receptor gating from the one interaction probed is the only thing that has caused the ligand binding to decrease or increase. This is what enables the use of linear free energy diagrams to link amino acid fluorination to calculated cation-π binding.

Incorporation of amino acids that only subtly change the stereoelectronic properties of amino acid residues is a difficult challenge. The use of orthogonal tRNA synthetase/unnatural amino acid pairs is incompatible, because in many cases the noncannonical amino acid resembles the natural amino acid, so the natural analogue will likely be incorporated as well. Our methodology involves the chemical acylation of tRNA with a synthetic ncAA (Figure 1).[3] This precise control enables a wide variety of amino acids to be incorporated, including hydroxy acids and singly fluorinated analogues of amino acids.[4] However, the overall yield of incorporation is only as high as the amount of chemically acylated tRNA one can prepare. Functional assays that can detect very small quantities of protein are therefore needed.

Fortunately, the *Xenopus laevis* oocyte expression system fits these needs. Acylated tRNA and mRNA for a LGIC containing a stop codon are injected into a mature *X. laevis*

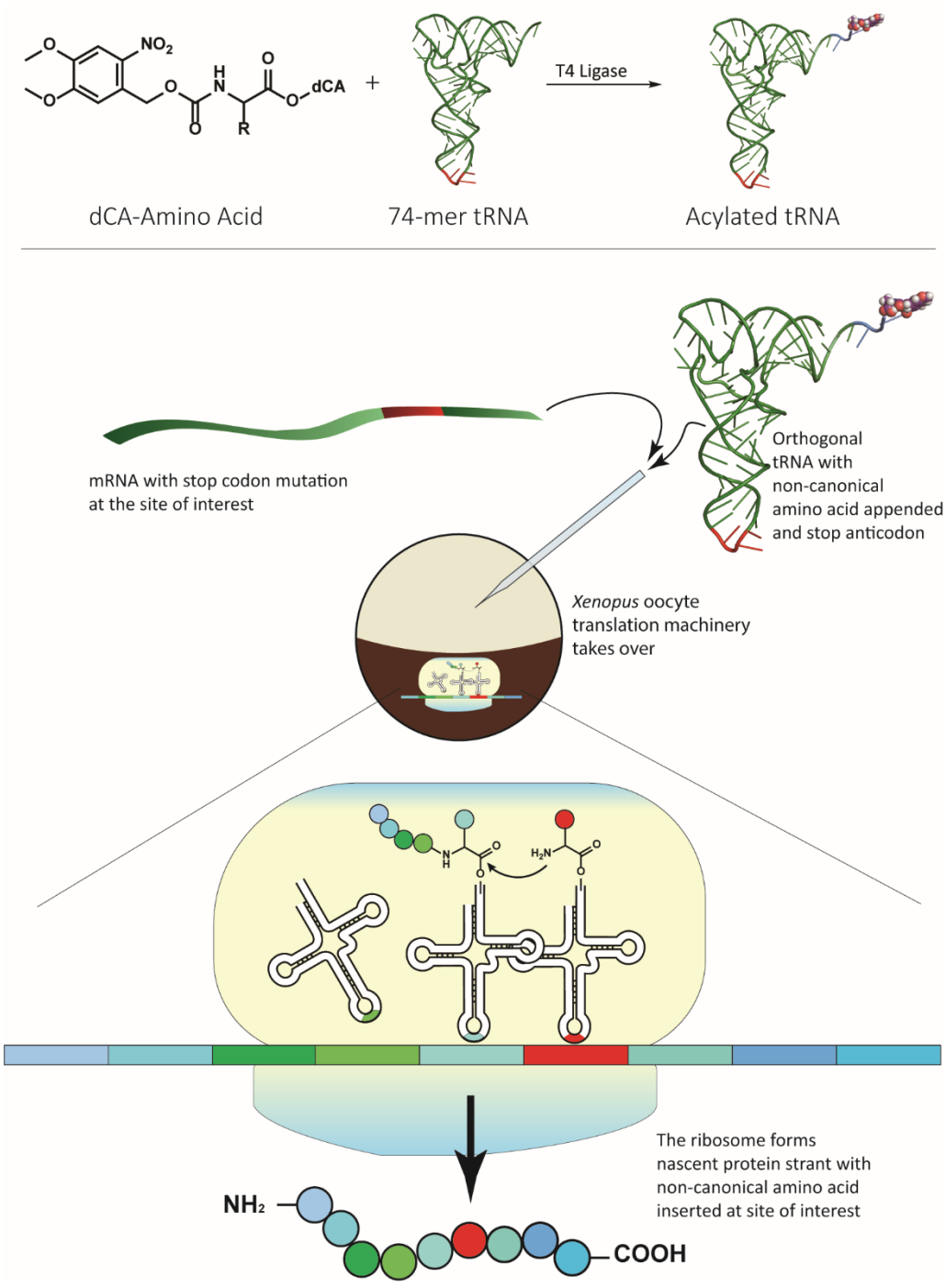


Figure 1: Top: T4 ligase is used to enzymatically append NVOC-protected dCA-amino acid to a 74-mer suppressor tRNA. Bottom, this tRNA-AA is injected alongside stop codon mutated mRNA into oocytes, wherein the ribosome takes over and inserts the amino acid into the nascent strand of the gene.

oocyte and incubated for several days. Then, the cells are monitored using two electrode voltage clamp electrophysiology, and varying amounts of agonist are added.[5] The concentrations at which half the receptors are open (EC_{50}), which is the concentration of

agonist added at half-maximal current, is measured.[6] A change in ligand binding due to incorporation of a ncAA will change the measured EC_{50}. The changes in EC_{50} can then be compared back to empirical or computationally-derived values. An increase of EC_{50} is a loss of function mutation, whereas a decrease is a gain of function mutation.

Probing cation-π interactions has been discussed in previous chapters. We have also probed hydrogen bonding between either ligands or amino acids. The standard methodology for this is incorporation of α-hydroxy acids instead of amino acids.[7] The incorporation of an α-hydroxy acid has a several effects. First, the elimination of an amide negates the ability of that group to be a hydrogen bond donor. Second, the introduction of an ester for an amide will greatly decrease the negative electrostatic potential on the carbonyl. This reduces its ability to be a hydrogen bond acceptor. Several examples of the usefulness of this mutation have emerged. One example is particularly illustrative.

As discussed in the previous chapter, many agonists of nAChRs will make hydrogen bonds between the protonated cationic amine and the carbonyl of the tryptophan on the B loop. Substitution of the next residue on the protein from Trp – a threonine – to its alpha hydroxy acid derivative – threonine α-hydroxy, or Tah – reduces the strength of that hydrogen bond. This results in a concomitant loss of function for the receptor when the agonist is nicotine.[8] Acetylcholine, which does not contain a protonated cationic amine, does not show a loss of function upon Tah incorporation.

One key drawback to this method is that unnatural incorporation must occur at the $i + 1$ residue. If that residue is proline, then the α-hydroxy acid is no longer a sufficient analogue. In that instance, it is necessary to design a new way to study hydrogen bonding while avoiding ablating the unique structural properties of proline. This chapter describes

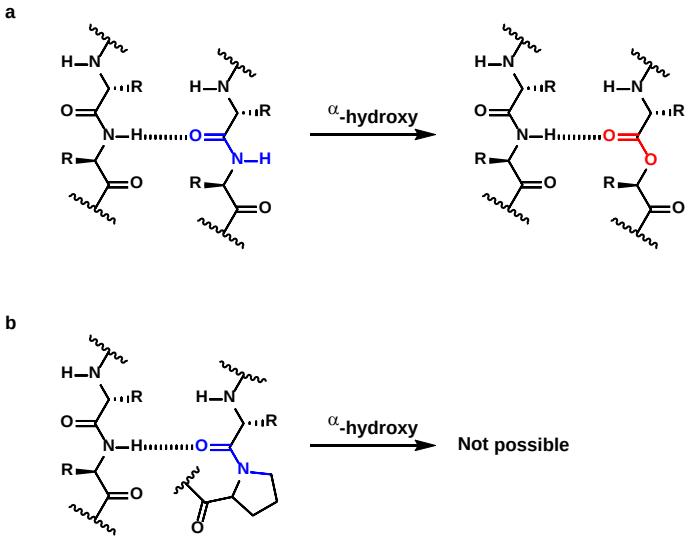


Figure 2: Probing hydrogen bonds involving backbone carbonyls. (a) Incorporation of an α-hydroxy acid will decrease the strength of a hydrogen bond, but (b) when proline is at the *i* + 1 residue, the mutation is not possible.

efforts to design an alternate ncAA methodology to probe hydrogen bonding in LGICs to avoid the complicating structural issues with proline.

A motivation for the design of new ncAAs to solve this problem was due to a recent investigation into an intersubunit hydrogen bond in α4β2 nAChRs. Recent crystal structures have shown a putative hydrogen bond between the indole N-H of TrpB and a carbonyl from a leucine on the adjacent β2 subunit (Figure 3). To probe for this interaction, TrpB was replaced with its naphthyl derivative, 1-naphthylalanine (NpAla). This eliminates the indole N-H but retains the steric and electronic properties of tryptophan. When acetylcholine is used as the agonist, a fivefold gain of function was observed.

This is highly unusual for several reasons. First, when the same mutation is made to muscle-type acetylcholine receptors, a loss of function concomitant with the loss of electron density in NpAla relative to Trp was observed.[9] A similar result was seen with a NpAla substitution into a homologous nAChR, the homopentameric α7 receptor (Table 1).

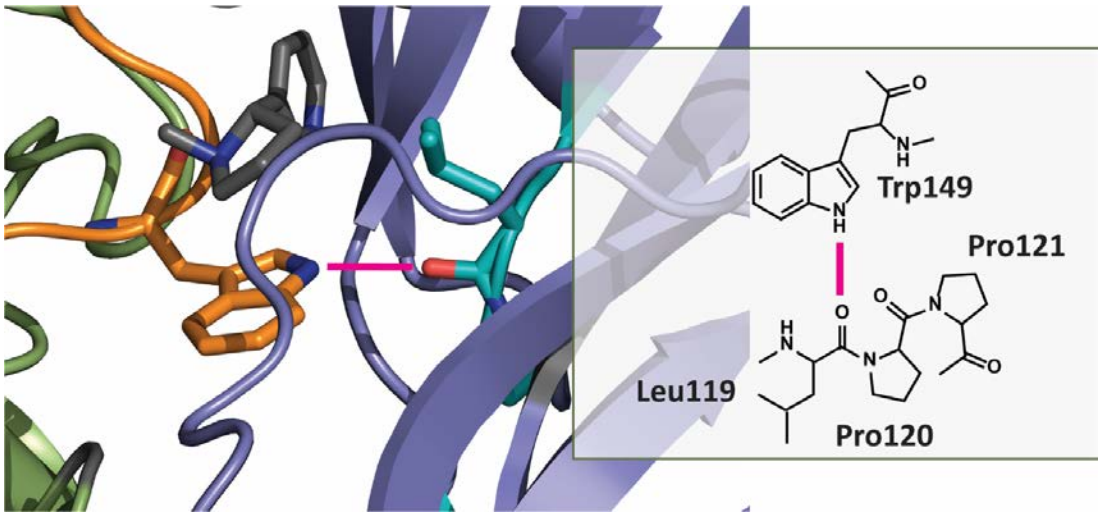


Figure 3: The full length α4β2 nAChR crystal structure (PDB: 5KXI) predicts a hydrogen bond between Trp149 (on Loop B, in orange) and Leu119 (on Loop E, in cyan)

Table 1: Investigating hydrogen bonding from the TrpB N-H to Leu119 in various receptors

$(\alpha4)_2(\beta2)_3$ nAChR	EC_{50} (uM)	n_H	I_{max} (uA)	Fold Shift	N
WT	0.5 ± 0.01	1.19 ± 0.03	4.25 - 30.31	1	10
NpAla	0.11 ± 0.003	1.16 ± 0.03	0.05 - 19.77	0.2	18
$(\alpha7)_5$ nAChR	EC_{50} (uM)	n_H	I_{max} (uA)	Fold Shift	N
WT	150 ± 0.5	1.4 ± 0.03	6.2 - 44.0	1	18
NpAla	920 ± 5	2.6 ± 0.1	1.4 - 18	6.4	14

Because ACh is a quaternary amine and cannot form a hydrogen bond to any part of TrpB, the gain of function is likely not due to ligand binding. Further, gain of function mutations are not commonly encountered when employing ncAA mutagenesis in the binding region of receptors of this type. Therefore, this NpAla discrepancy required further investigation.

Rather than affecting ligand binding, the mutation ablating a hydrogen bond could affect gating. Because EC_{50} is a composite measurement of binding and gating, a change to one or either could affect the measured value. If the NpAla gain-of-function result does not affect ligand binding, it could affect gating alone. One way to depict LGIC function is through an oversimplified three-state model (Figure 4).[10] The first state consists of a closed inactive channel and unbound ligand. The second state is a closed channel bound to agonist, and the third an open channel with agonist bound. The receptor is in a dynamic equilibrium

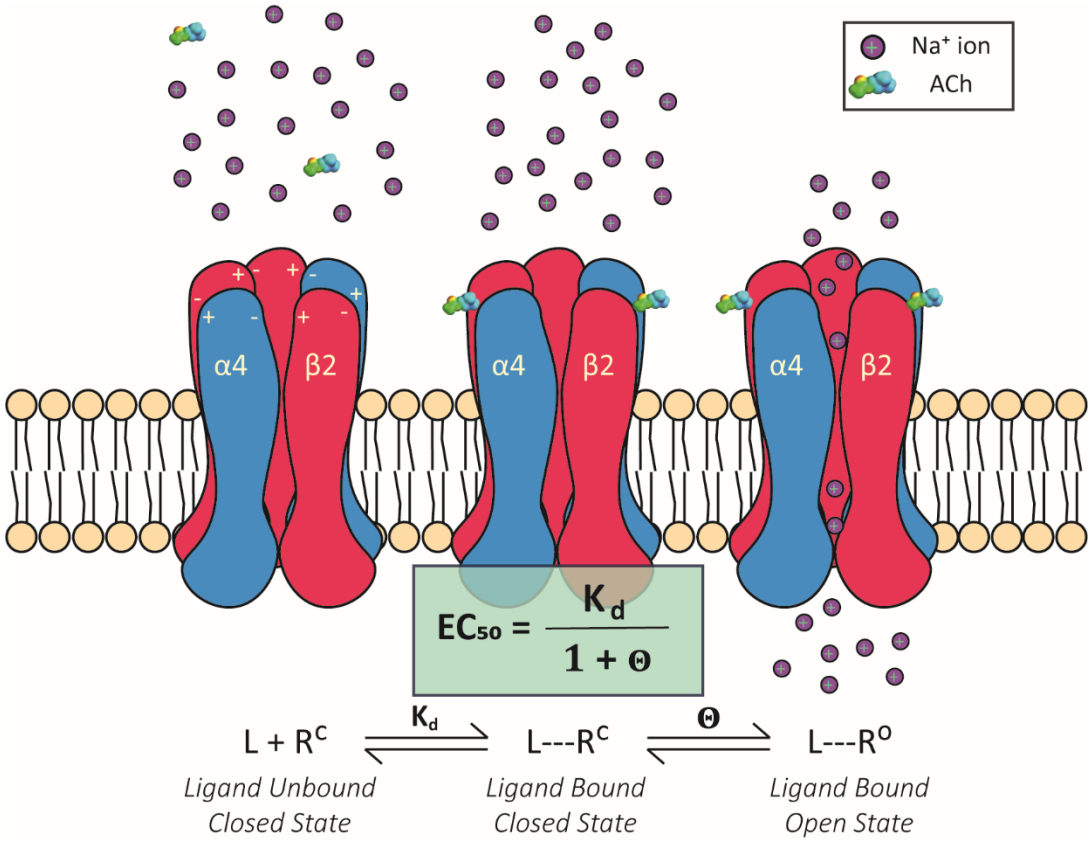


Figure 4: This schematic depicts the simple three state model used to describe ligand-gated ion channel function. The receptor, in this case α4β2 nAChR, is in a dynamic equilibrium among all three. The binding event is characterized by K_d, while the gating event is characterized by Θ. EC_{50} is a composite measurement taking into account both binding (K_d) and gating (Θ).

between the three states with agonist present, with K_d serving as the equilibrium constant describing the binding event. EC_{50} is equal to K_d divided by (1+Θ); therefore, if K_d is unchanged by a mutation (i.e: if the mutation does not affect ligand binding), and a decrease in EC_{50} is observed, it must be due to an increase in Θ. This could be due to either a destabilization of the closed-ligand bound state or a stabilization of the open, ligand bound state.

The potential hydrogen bonding partner for the indole N-H of TrpB appears to be, according to several crystal structures, a leucine on the β subunit, Leu119, on the E-loop of the receptor (Figure 3).[11] This amino acid is adjacent to a proline, part of a highly

conserved Xaa-Pro-Pro motif. Because the proline is present in the *i* + 1 residue of the hydrogen bond-forming carbonyl, traditional replacement with an α-hydroxy amino acid is impossible. This chapter describes several efforts to develop an alternate methodology to probe hydrogen bonding in nAChRs. Ultimately, the best approach appears to be to replace the residue itself with a fluorinated aliphatic amino acid, thus avoiding replacing the proline residue. Using this method, it is evident that the hypothesized hydrogen bonding interaction predicted by structure and hypothesized to explain the gain of function seen with NpAla is not actually functionally relevant, and that this mysterious effect is due to some other rationale.

4.3: Strategies to probe for hydrogen bonding at loop E Leu-Pro-Pro.

The following describe three putative strategies for studying the above described hydrogen bonding interaction.[12] All three follow a similar workflow. The desired amino acid is acylated on to tRNA and incorporated in to either the i or *i* + 1 sites for Leu119 in α4β2 nAChR. To ensure that the desired mutation is due only to changing the hydrogen bond, a double mutant cycle analysis is performed. This compares two separate mutations – in this case NpAla and whichever hydrogen bonding mutation – separately and together. This ensures that if a gain-of-function mutation is observed from one of the putative hydrogen bonding mutations, its relationship to the gain of function from the NpAla mutation can be determined.

Double mutant cycle analysis is a standard methodology to analyze a system in which two mutations are evaluated, individually and in tandem.[13-15] If a mutation in one site has no relation to the mutation in the other site – if the ablation of the tryptophan N-H

is not related to decreasing the hydrogen bonding ability of β2-L119 – simultaneous mutation at both sites will be multiplicative. This can be represented mathematically:

(1) $\Omega = \frac{[EC_{50}(AB) \times EC_{50}(A'B')]}{[EC_{50}(A'B) \times EC_{50}(AB')]}$

where Ω is the coupling constant, and A and B are the endogenous amino acids. A' and B' are the mutated amino acids. A coupling constant of 1 implies that the two mutations are multiplicative, and are not coupled. If the two residues interact, Ω will be either greater or less than the product of their individual effects. This approach has been widely applied, including to LGICs. The coupling constant can be further related to a free energy of interaction by the equation

(2) $\Delta\Delta G = \mathrm{RT}\ln\Omega$

where ΔΔG is the free energy of interaction, calculated directly from the coupling constant. A coupling constant $\Omega > 2$ is considered significant. This corresponds to a $\Delta\Delta G > 0.4$ kcal/mol.

4.3.1: Attempted incorporation of δ-amino acids

The first approach to determine if Loop B and E make an interfacial hydrogen bond was also practically and conceptually the most complicated. Because the Leu-Pro-Pro motif is very well conserved (the vicinal di-proline is found in almost every nAChR subunit from almost every organism in which the gene has been cloned), a strategy was conceived that retained both side chains while eliminating the carbonyl. However, this necessitates the ribosome incorporating an amino acid with a carbon chain long enough to be a dipeptide. In addition, this necessitates the replacement of an sp^2 carbonyl for an sp^3 methylene group, likely changing the conformation properties of the dipeptide analogue. The proposed amino acid is a δ-amino acid (Figure 5).

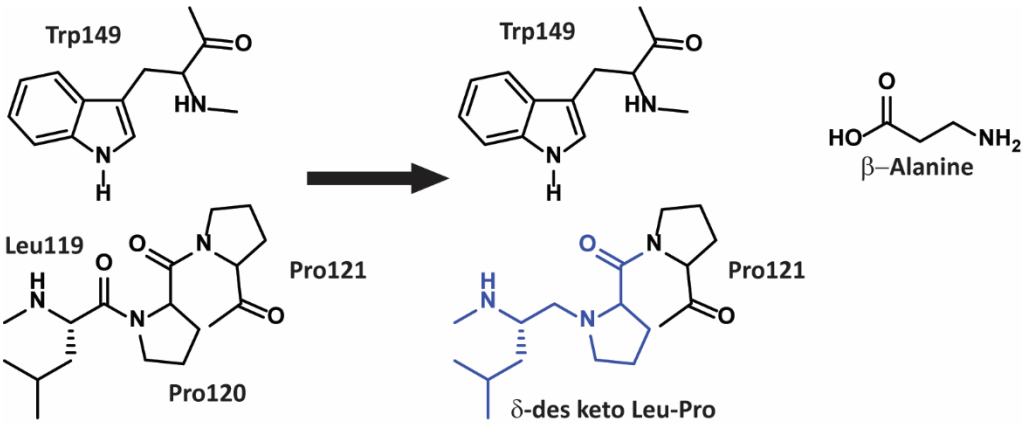
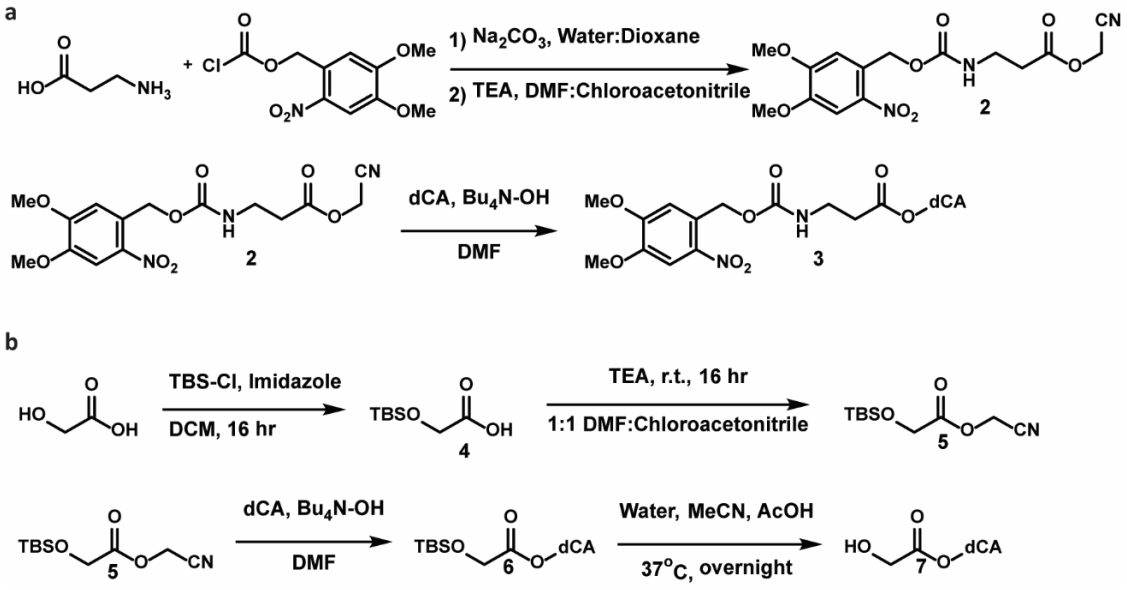


Figure 5: The proposed hydrogen bond that is stabilizing the closed state is between the indole NH of Trp149 (Trp B) and Leu119 (of Loop E). A δ-amino acid containing the Leu and Pro side chains but lacking a carbonyl group was propsoed. As a first step, β-alanine was incorporated (right).

It was unclear whether the oocyte ribosome could incorporate such an amino acid with an extended backbone. There has been some work on incorporation of non-α amino acids (namely β-alanine) using a human ribosome *in vitro* that had extremely poor yields. As such, the incorporation of β-alanine was attempted in muscle nAChR at various glycine sites. The dCA-coupled β-alanine was synthesized (**3**, Scheme 2a) and ligated to THG73 tRNA and tested.[16] These tests either resulted in no functional expression or the observed



Scheme 1: Synthesis of amino acids conjugated to dCA nucleotides, previous to tRNA ligation. (a) β alanine and (b) α-hydroxy glycine, or glycolic acid are described. Note: compound **1** is Nvoc-protected β-alanine.

agonist-induced currents at both β-alanine and 76mer control conditions. In other words, the incorporation of β alanine over reaminoacylated 76mer could not be confirmed. The results suggest that not even a β-amino acid could be incorporated through the ribosome, and further attempts to incorporate longer amino acids were likely to fail. Therefore, this strategy was abandoned.

4.3.2: Incorporation of Gly/Gah at Pro120

Where the strategy of incorporating an extended amino acid was decidedly risky, the subsequent approach was simple. Rather than attempting to leave the conserved proline residue in place, a glycine mutation was attempted. The rationale behind this change was that while proline is a conformationally restricted amino acid, glycine can adopt many conformations due to its lack of a side chain. This would mean that it has the opportunity to mimic a proline conformation, where a substituted amino acid may not. In this case, a glycine/glycine α-hydroxy (Gah) pair would be used for hydrogen bond analysis in the

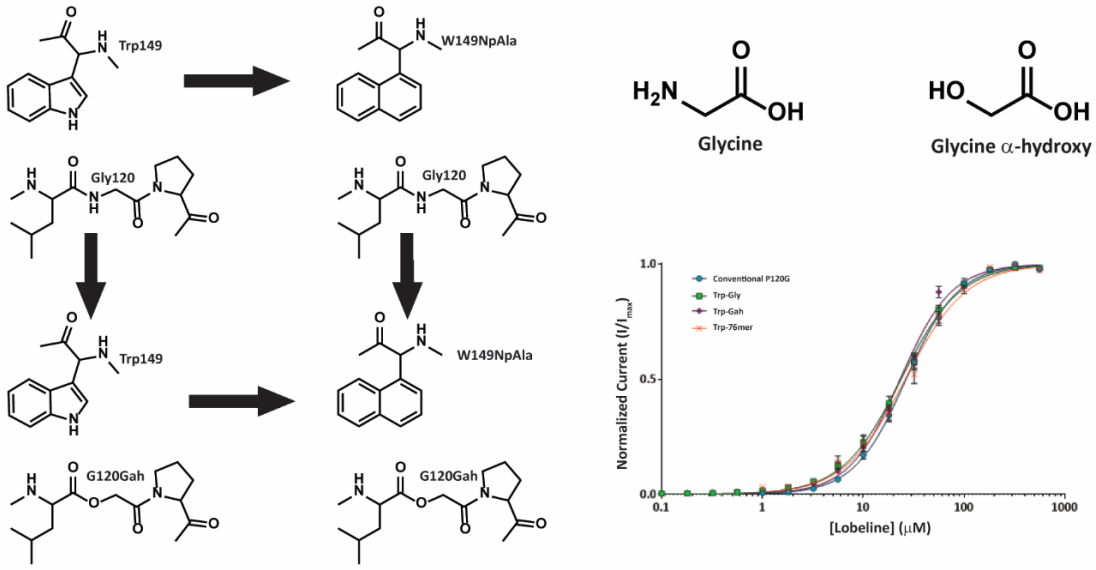


Figure 6: Left: A proposed double mutant cycle using a P120Gly mutant as the "wild-type" and Gah and NpAla as the single-mutants. Top right: the structures of glycine and 'Gah,' or glycine α-hydroxy. Bottom right: dose-response curves of the attempted double mutant cycle, which line up with the 76mer control.

double mutant cycle. In this cycle, as before, the first mutation is incorporation of Gly in the P120 site and NpAla at Trp149. Then, Gah was incorporated with tryptophan unchanged. The third mutation incorporated both Gah and NpAla. If the predicted hydrogen bond is present, the Gly-Gah mutation should result in a decrease in EC_{50}, and should not be coupled to the NpAla mutation by double mutant cycle analysis (Figure 6).

α-hydroxy glycine (Gah) was prepared in a similar manner to other α-hydroxy amino acids (Scheme 2b). First, the hydroxy acid was protected with TBS-Cl and chloroacetonitrile to yield **5**. This acid was ligated to dCA in the normal manner, yielding **6**. This compound could be taken on to ligation, yielding the protected hydroxy acid Gah. Before injection, the amino acid was deprotected with acetic acid under mild conditions(**7**).

Gah was incorporated on an alternate suppressor tRNA, TQOPS, which contains a UGA codon, rather than the UAG codon for the commonly used THG73 codon.[17] This was to enable simultaneous incorporation of two unnatural amino acids: Gly/Gah and NpAla/Trp. This procedure has been repeated previously for two separate subunits of the muscle nAChR. In the double mutant experiment, therefore, α4 Trp149 was mutated to a TAG stop codon and suppressed with THG73 tRNA, and β2 Leu119 was mutated to a TGA stop codon and suppressed with TQOPS' tRNA. The TQOPS' suppressor tRNA has an added benefit of being less likely to get reaminoacylated with endogenous amino acid. This characteristic is important, as residues on Loop E are quite promiscuous in that they can tolerate a wide range of amino acid substitutions, and they are therefore more likely to have an issue with reaminoacylation.

The results (shown in Figure 6, Table 2) show that all four combinations of mutations show high currents and monophasic dose response curves. The Trp-Gly and Trp-

Table 2: Investigation of the interfacial hydrogen bond using a Gly-Gah mutant cycle

α4β2	W149	P120	EC_{50} (μM)	n_H	I_{max} (μA)	Fold Shift	N
WT	Trp	Gly	31.7 ± 0.8	1.49 ± 0.04	2.13 - 18.41	1	13
Mut1	NpAla	Gly	9 ± 0.3	1.34 ± 0.05	0.04 - 0.53	0.3	14
Mut2	Trp	Gah	23.3 ± 0.8	1.69 ± 0.09	0.68 - 9.59	0.7	6
Mut1,2	NpAla	Gah	9 ± 1	1.09 ± 0.13	0.17 - 0.33	0.3	2
						Ω	**1.4**
						ΔΔG (kcal/mol)	**-0.2**

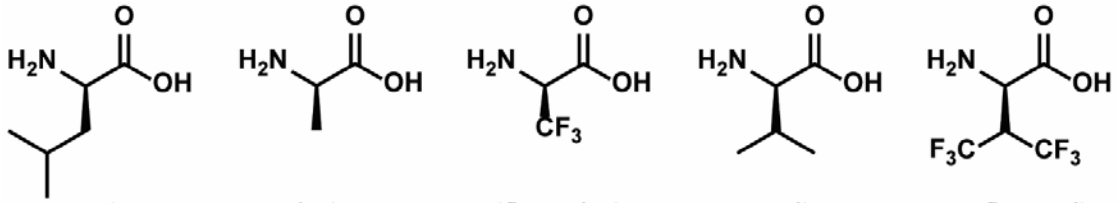
Gah EC_{50} values were not meaningfully different, and the NpAla cases showed the same five-fold decrease in $EC_{50,}$ and no changes were observed between the Gly-NpAla and Gah-NpAla mutation. These results would suggest that no hydrogen bond exists, with a Ω value resulting in a ΔΔG of 0.2 kcal/mol. However, controls with 76mer tRNA, which reaminoacylates with glutamine, showed no meaningful difference from any of the other mutations.[17] Critically, when β4-P120 is mutated to glycine conventionally the dose response curve overlaps very well with the 76mer control. This would imply that while EC_{50} values can be measured for all tested mutations, one cannot conclude that any ncAAs have been incorporated into the β4-P120 site. Because the results themselves cannot be adequately analyzed, the Gly-Gah double mutant strategy was determined to not be viable.

4.3.3: Fluorinated aliphatic amino acids

Clearly, modification of the proline residue is not an acceptable strategy. Modification of the i + 0 residue was instead attempted. Rather than a reduction of the carbonyl's electrostatic potential via ester modification, electron withdrawing groups can be added to the alpha carbon of the amino acid. In order to change only the electronic properties of the amino acid side chain without drastically changing the steric properties, fluorinated aliphatic amino acids can be utilized. While a fluorinated leucine would certainly be an option and would most closely match the endogenous amino acid, other

Table 3: Coventional mutagenesis of Leu119 with aliphatic amino acids

$\alpha4\beta2$	EC_{50} (μM)	n_H	I_{max} (μA)	N
WT	0.50 ± 0.01	1.19 ± 0.03	4.24 - 30.31	10
L119A	1.49 ± 0.08	1.21 ± 0.06	3.05 - 15.98	11
L119V	1.73 ± 0.06	1.12 ± 0.03	5.93 - 26.20	11



Scheme 2: Amino acids of interest for the studying hydrogen bonding by putting electron withdrawing groups on an aliphatic side chain.

perfluorinated aliphatic amino acids – hexafluorovaline (F_6Val) and trifluoroalanine (F_3Ala) – are commercially available so therefore relatively easy to synthesize.

These amino acids have precedent for use in peptides. Perflorous amino acid side chains have been used to simulate the fluorous effect in protein-protein interactions.[18-19] Inclusion of these amino acids at the interface of leucine zipper proteins results in increased packing efficiency and exclusion of water. However, they have not previously been used for their electronic properties.

Before the amino acids themselves were incorporated, preliminary data were collected on the inclusion of alanine and valine in place of β2-L119. This is because these amino acids will be the natural analogues in any eventual double mutant cycle using fluorinated amino acids. The other mutation in the cycle will be, as before, NpAla for α4-Trp149 (TrpB). A NpAla-F_3Ala or NpAla-F_6Val double mutation would serve as a way to measure the additivity and confirm the presence or absence of the putative hydrogen bond (Table 3).

It is essential that incorporation of Ala or Val results in functional receptors for the perfluorous amino acids to be successful. Conventional mutations showed a modest 3-fold

loss of function compared to wild type receptor. This is well within a workable range for a double mutant cycle. Initial nonsense suppression experiments with Ala and Val tRNA matched the conventional mutants, and produced well behaved large current traces with monophasic EC_{50} curves. 76mer tRNA positive controls do not show expression at the α4-Trp149 site, but do show significant current at β2-L119. However, unlike with the Gly/Gah strategy, the dose response curve never turns over even as the concentration of acetylcholine approaches 1000 μM. While the results are promising, more data are required to establish this strategy as a viable one for testing hydrogen bond strength. As such, computational and *in vitro* experiments were conducted to determine the strategy's viability.

4.4: Computational and *in vitro* investigation of aliphatic fluorinated ncAAs

4.4.1: Electrostatic potential calculations for hydrogen bonding

Hydrogen bonding can be modeled as an electrostatic interaction. Much work has already been done to marry computational (semi empirical, DFT, and molecular mechanics) with empirical data, in order to enable conclusions to be drawn directly from *in silico* to *in vivo*. Hydrogen bonding can be directly related to the electrostatic potential at the surface of the molecule.[20] The phenomenon can be described as:

(3) $\log(K) = c_1 \alpha_2^H \beta_2^H + c_2$

where K is the association constant, c_1 and c_2 are solvent-dependent constants, and α_2^H and β_2^H are constants that characterize the properties of a hydrogen bond donor and receptor of interest respectively. This relationship can also be expressed in such a way to determine the change in energy from switching a hydrogen bond from a solvent to a hydrogen bonding partner. Recent work has also enabled the calculation of these values

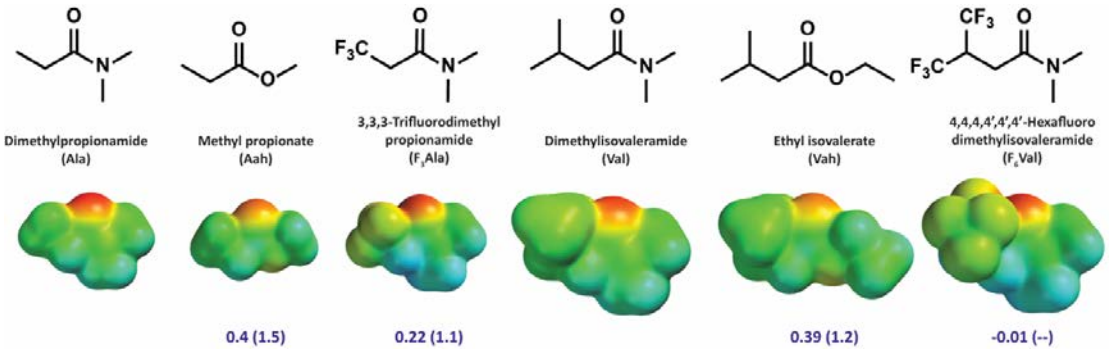


Figure 7: Top: structures of computationally and *in vitro* NMR-studied amino acid derivatives as well as the names they are referred to in this and the next subsections. Bottom: calculated electrostatic potential maps of the above compounds. Equilibrium geometries and electrostatic potentials calculated at DFT//M06/6-31G(d,p). Hydrogen bonding energy difference calculations relative to the base amino acids for computational and NMR (in parentheses) studies are in blue, energies in kcal/mol.

directly from an electrostatic potential surface.[21] The minimum (hydrogen bond acceptor) and maximum (hydrogen bond donor) values of an electrostatic potential surface can be used in the equation:

$$(4)\ \Delta\Delta G_{H-bond} = -(\alpha - \alpha_s)(\beta - \beta_s)$$

In Eq. 4, $\Delta\Delta G_{H\text{-}bond}$ is the change in free energy upon switching a hydrogen bond from solvent to a donor, and α_s and β_s are empirically-derived solvent terms for hydrogen bond donating and accepting respectively. The α and β values are determined from the generated electrostatic potential maps.

The energy of hydrogen bonding was therefore calculated for a number of amino acids derivatives of interest, including several fluorinated versions as well as an ester (ethyl acetate). The substrates chosen matched the compounds used for NMR analysis in the next subsection. The electrostatic potential maps were generated by finding the equilibrium geometry conformation of each amino acid in the gas phase using a DFT//M06/6-31G(d,p) level of theory. This level of theory has previously been used to determine hydrogen bonding to a relatively accurate degree. Electrostatic potential maps were then generated and the appropriate α and β values determined for each substrate.

Table 4: Results from computational electrostatic potential map hydrogen bond predictions. Energies are the energies of switching hydrogen bond from solvent to an Indole N-H

	$E_{min/max}$ (kJ/mol)	α or β	$\Delta\Delta G_{H\text{-}bond}$ (kcal/mol)					
			Water	*Hexanes*	*CCl_4*	*DMSO*	*Acetonitrile*	*DMF*
Indole	236.8	4.5						
EtAc	175.4	3.4	0.5	-3.1	-2.1	2.1	0.9	0.9
Ala	205.7	4	0.2	-3.6	-2.5	1.8	0.5	0.5
Aah	175.3	3.4	0.5	-3.1	-2.1	2.1	0.9	0.9
F_3Ala	191.1	3.7	0.3	-3.3	-2.3	1.9	0.7	0.7
Val	191.5	3.7	0.3	-3.4	-2.3	1.9	0.7	0.7
Vah	164.5	3.2	0.6	-2.8	-1.9	2.1	1.0	1.1
F_6Val	192.4	3.7	0.3	-3.4	-2.3	1.9	0.7	0.7

Because for a double mutant cycle analysis, the relevant number is a fold shift from wild type, the choice of solvent to use in equation (4) does not matter so long as the solvent remains the same for the entire analysis. Therefore, the reported values are the difference between the $\Delta\Delta G_{H\text{-}Bond}$ of the normal amino acid analogue to either the ester or fluorinated version; a $\Delta\Delta\Delta G$ value. Because the NMR studies in the subsequent section were performed in carbon tetrachloride, this solvent was ultimately used to report the final $\Delta\Delta\Delta G$ values. The results are shown in Figure 7 and Table 4.

For the F_3Ala analogue, the calculated shift in $\Delta\Delta G_{H\text{-}bond}$ was determined to be 0.22 kcal/mol, about half as attenuating as alanine α-hydroxy (Aah). There was no meaningful calculated shift in $\Delta\Delta G_{H\text{-}bond}$ of F_6Val compared to Val, with a value of -0.01 kcal/mol. Valine α-hydroxy (Vah), on the other hand, showed a shift in $\Delta\Delta G_{H\text{-}bond}$ of 0.39 kcal/mol. This result was not unexpected, as the inductive effect of groups on the γ carbon relative to the carbonyl decreases rapidly, so any electron withdrawing effects would be unlikely to carry over in an observable manner. While the calculated values seem low, F_3Ala would appear to still be a promising candidate for hydrogen bond testing. Because the shifts for esters was also small, it is likely that this methodology underestimates the actual relative change in binding energies. The observation that F_6Val may not shift the hydrogen-bonding ability greatly is not concerning. It could provide a valuable negative

control; adding addition fluorines without greatly affecting the electronic properties in question.

4.4.2: In vitro *study of fluorinated ncAA derivatives using NMR*

In order to accurately claim that F_3Ala modulates hydrogen bond strength, *bona fide* empirical evidence is necessary. A classic experiment by Gurka and Taft lays out a procedure to measure the ability of hydrogen bond acceptors in carbon tetrachloride.[22] In it, 4-fluorophenol serves as a hydrogen bond donor, and amides or esters such as those described in the previous subsection are the hydrogen bond acceptors. If a hydrogen bond is formed, the electronic environment around the fluorine atom will shift upfield. This shift is measure relative to an internal control of 4-fluoroansiole. These data can be used to calculate a formation constant, K_f, which follows a Van't Hoff relationship, so therefore can be used to calculate a ΔG_f.

A dose-response curve can be obtained using this method (Figure 8a). Increasing amounts of donor was added to a 0.01 M solution of both the phenol and anisole in carbon

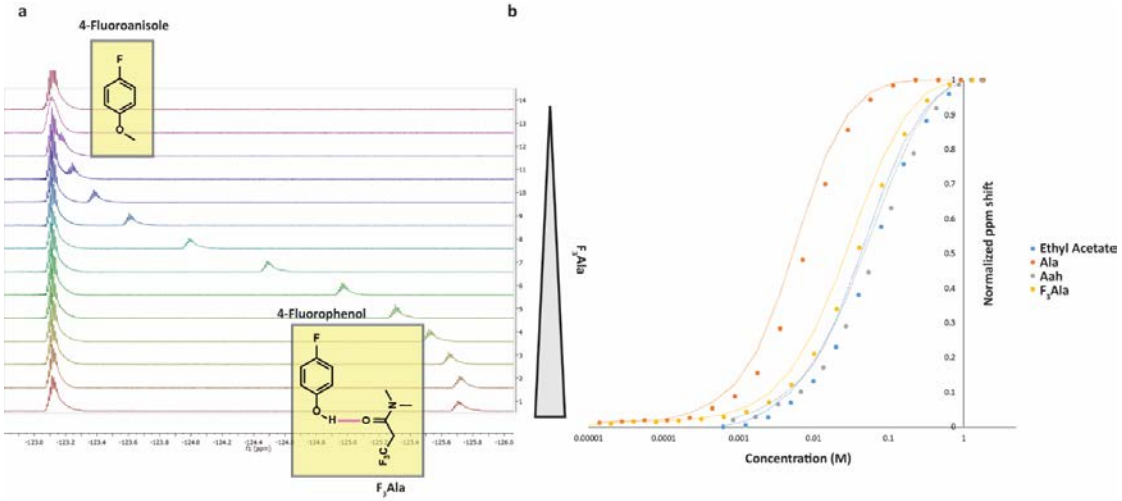


Figure 8: (a) Titration of a hydrogen bond acceptor, F_3Ala, into a solution of 4-fluroanisole and 4-fluorophenol, monitored by ^{19}F NMR. (b) Sigmoidal curves obtained from the data from (a) for several different amino acid derivatives.

tetrachloride. The changes in chemical shift follow a sigmoidal shape (Figure 8b), and the points along the rise of the curve can be used to calculate K_f by the equation:

$$(5)\ K_f = \frac{[p\text{−fluorophenol}]_o\ (\delta/\Delta)}{[p\text{−fluorophenol}]_o(1 - (\delta/\Delta))([\text{acceptor}]_o - [p\text{−fluorophenol}]_o(\delta/\Delta))}$$

where $[X]_o$ is the concentration of each hydrogen bonding partner and (δ/Δ) is the chemical shift of the fluorophenol relative to fluoroanisole at the donor concentration (δ) divided by the maximum observed change in chemical shift (Δ).

The compounds chosen for this experiment are amide or ester derivatives of the amino acids of interest. None of these contain an amino group, so as to not contain any hydrogen bond donors or additional acceptors to confuse the NMR analysis. All compounds matched those of the computational study in the previous subsection with the exception of the F_6Val derivative, which did not show any attenuating ability computationally. For each compound, K_f was determined at several different points and averaged (Table 5).

$\Delta\Delta G$ values can be determined by taking the difference between the normal amino acid analogue (ex: N,N-dimethylpropionamide for Ala or N,N-dimethylisovaleramide for Val) and either its fluorinated or ester derivative. Overall, these experiments show that an Ala-Aah mutation causes a 1.48 kcal/mol loss in binding energy, Ala-F_3Ala a 1.12 kcal/mol loss, and Val-Vah a 1.20 kcal/mol loss. Fortunately, these results match qualitatively to the computational work. In addition, the energies themselves fall into the expected range of results based on previous *in vivo* investigations of

Table 5: ^{19}F NMR-based hydrogen bonding

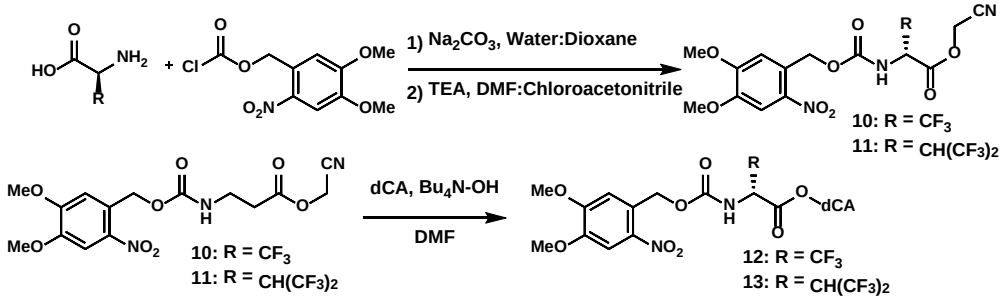
	K_f	$\Delta G_{H\text{-bond}}$	N
EtAc	18.3 ± 0.3	-1.7 ± 0.1	5
Ala	214 ± 6	-3.2 ± 0.7	4
Aah	17.5 ± 0.3	-1.7 ± 0.1	4
F_3Ala	32.6 ± 0.9	-2.1 ± 0.2	4
Val	113 ± 3	-2.8 ± -0.5	5
Vah	13.9 ± 0.9	-1.6 ± -0.2	5

hydrogen bonding in ion channels. The largest fold shift observed for a hydrogen bond is 62 for the $(\alpha4)_2(\beta2)_3$ nAChR – a hydrogen bond from Cytisine to L119. This corresponds to a $\Delta\Delta G$ of 2.45 kcal/mol. While this is still larger than the ^{19}F-NMR derived energy shifts, there are confounding factors in the protein. Further, most changes are more on the order of 1.0-1.4 kcal/mol, much more in the observed range. Based on these data, it is clear the F_3Ala is a viable amino acid to test for the presence of hydrogen bonding. In addition, these studies bolster the evidence for the use of backbone ester mutations.

4.5: Incorporation of F_3Alanine and F_6Valine in α4β2 nAChRs

4.5.1: Synthesis of fluorinated aliphatic amino acids

Scheme 2 shows the synthesis of dCA-coupled F_3Ala and F_6Val. The syntheses themselves are simple and follow from previous work. First, an NVOC protection is performed, followed by cyanomethylation of the acid with chloroacetonitrile to yield **10** and **11**. This doubly protected amino acid is taken on to the dCA-coupled group (**12** and **13**). After HPLC purification, both were ligated on to both 74mer TQOPS' and THG73' tRNA as previously reported. The success of the dCA coupling and ligation were confirmed by MALDI mass spectrometry.



Scheme 3: Synthesis of F_3Ala and F_6Val protected and conjugated to dCA, prior to tRNA ligation.

4.5.2: Progress towards incorporation of F_3Alanine at a validated hydrogen bonding site

Before testing a new ncAA on a putative hydrogen bonding site, it is important to establish that F_3Ala can affect hydrogen bonds in the context of nonsense expression experiments. In order to establish this, a known hydrogen bond will be tested first. The preliminary positive control is a hydrogen bond that has been observed in muscle-type nAChRs and α4β2 that plays a large part in the muscle type receptor's ability to discriminate between nicotine and acetylcholine. The hydrogen bond itself is between Lys158 and Ile201, both on the α subunit (in α4β2 receptors).[23] In muscle type, the residues are glycine and proline respectively. Mutation of the glycine to a lysine in muscle type recovers nicotine sensitivity, and weakening of the hydrogen bond in α4β2 receptors by inclusion of an alpha hydroxy amino acid next to the Ile results in a fivefold loss of function.

In order to test the effectiveness of F_3Ala, Ile201 will be mutated first to Alanine, then to F_3Alanine. Based on the NMR and computational results in the last section, a loss of function that is someone less than the Y202Yah mutation would be expected. Preliminary work has included the incorporation of a stop codon at Ile201 and testing with THG73-F_3Ala as well as THG73-Ala and THG73-Ile. The I201A conventional mutation has also been made and tested along with nonsense suppression of Ala and Ile. The results are shown in Table 6. The conventional mutation of Ile201 to alanine

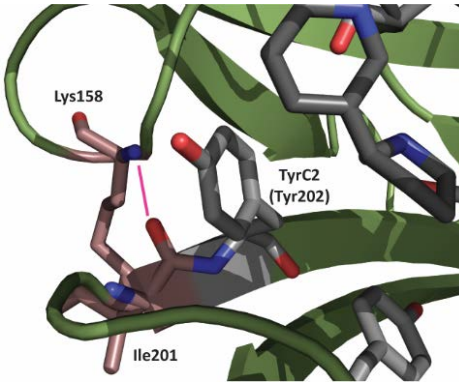


Figure 9: Hydrogen bond between Ile201 and Lys158 in α4β2 nAChR (PDB: 5KXI)

Table 6: Preliminary data to test the hydrogen bond between I201 and K158 in α4β2

	Amino acid	EC_{50} (μM)	n_H	I_{max} (μA)	Fold Shift	N
I201A	-	0.077 ± 0.003	1.37 ± 0.01	1.99 - 9.33	1.0	10
I201TAG	Ile	0.077 ± 0.007	1.16 ± 0.10	0.07 - 0.17	1.0	15
I201TAG	Ala	0.059 ± 0.004	1.21 ± 0.08	0.09 - 0.13	0.76	12

produced a monophasic dose-response curve with large currents. Wild type recovery as well as incorporation of alanine through nonsense suppression suggest that incorporation of F_3Ala is a feasible strategy. Initial experiments to incorporate F_3Ala are currently in progress.

4.5.3: Double mutant cycle at β2-Leu119 and α4-Trp149

With the fluorination strategy proposed supported by *in vitro* and *in silico* data, a double mutant cycle could be attempted to determine the presence of the putative hydrogen bond. Thus, dose response experiments on receptors expressing the following THG73-TQOPS' combinations to make up the mutant cycle: Trp-Ala as the wild-type, NpAla-Ala as the first single-mutant, Trp-F_3Ala as the second single-mutant, and NpAla-F_3Ala as the

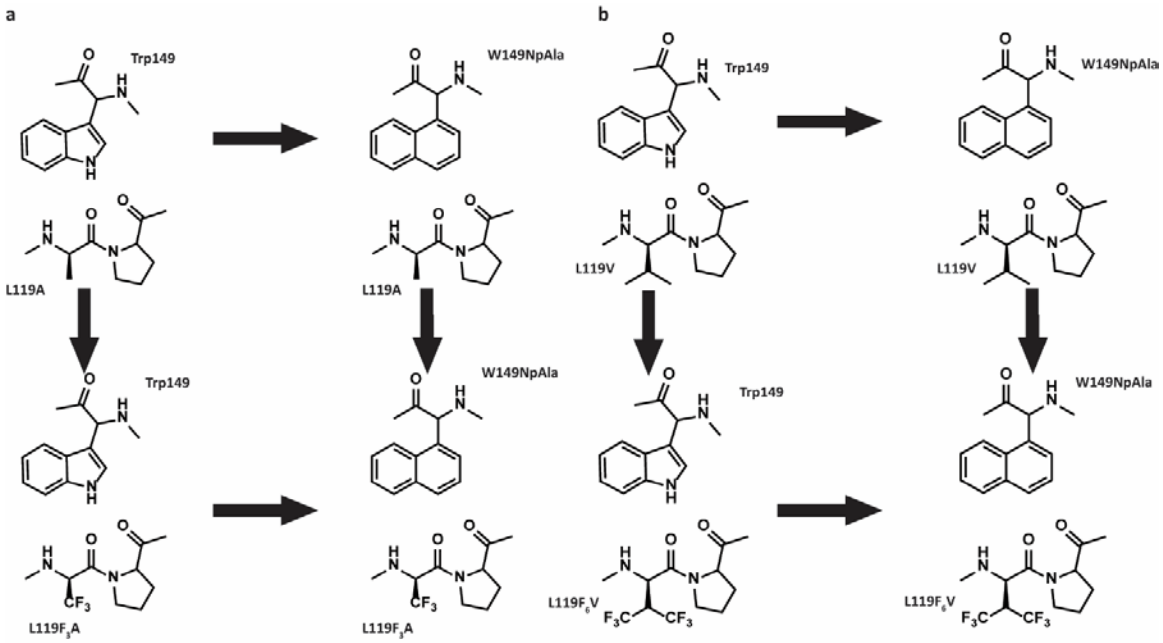


Figure 10: Proposed double mutant cycles to study the interfacial hydrogen bond between TrpB and LeuE utilizing (a) trifluoroalanine and (b) hexafluorovaline.

double mutant. Results from this cycle (Table 7) show that F_3Ala has no effect on EC_{50} and is not meaningfully coupled to the NpAla mutation. The Ω value calculated from this mutant cycle is 1.7, corresponding to a ΔΔG value of around 0.3 kcal/mol.

The F_6Valine double mutant cycle, which based on the computational data is expected to be a negative control, showed similar results. Trp-Val and Trp-F_6Val having similar EC_{50} values about 3- to 2-fold greater than NpAla-Val and NpAla-F_6Val. These values lead to an Ω value of about 1.4 and a ΔΔG of -0.2 kcal/mol, results that are not that different from the alanine cycle.

The results of the alanine cycle alone are not enough to suggest that there is a meaningful hydrogen bond present. Taken with the valine results, which should be a negative control, there is a strong suggestion that the hydrogen bond does not exist. If the valine cycle is considered a baseline, the alanine cycle has a ΔΔG of 0.1 kcal/mol, which is not considered significant. This value is lower than both the computationally predicted and NMR-predicted binding energy. This suggests that the initial hypothesis is incorrect, and the gain-of-function NpAla mutation does not form a hydrogen bond and stabilize the closed state.

Table 7: Mutant cycle results for F_3Ala and F_6Val. Both cases did not conclusively show that a hydrogen bond was present.

α4β2	W149	P120	EC_{50} (μM)	n_H	I_{max} (μA)	Fold Shift	N
WT	Trp	Ala	2.18 ± 0.06	1.15 ± 0.03	0.23 - 1.87	1	10
Mut1	NpAla	Ala	0.388 ± 0.008	1.15 ± 0.02	0.35 - 9.92	0.2	14
Mut2	Trp	F_3Ala	2.0 ± 0.2	0.75 ± 0.03	0.05 - 1.75	0.9	11
Mut1,2	NpAla	F_3Ala	0.62 ± 0.02	0.8 ± 0.02	0.09 - 0.82	0.3	14
						Ω	**1.7**
						ΔΔG (kcal/mol)	**-0.3**

α4β2	W149	P120	EC_{50} (μM)	n_H	I_{max} (μA)	Fold Shift	N
WT	Trp	Val	2.08 ± 0.07	1.2 ± 0.04	0.13 - 5.54	1	15
Mut1	NpAla	Val	0.65 ± 0.01	1.16 ± 0.02	1.25 - 6.36	0.3	13
Mut2	Trp	F_6Val	2.1 ± 0.2	0.78 ± 0.03	0.07 - 29.91	1.0	13
Mut1,2	NpAla	F_6Val	1.22 ± 0.08	0.77 ± 0.02	0.11 - 1.96	0.6	11
						Ω	**1.9**
						ΔΔG (kcal/mol)	**-0.4**

4.6: Discussion

Work to determine the utility of fluorous aliphatic amino acids is still in progress. One outstanding question, however, is the relevance of the hydrogen bond between the indole N-H of TrpB and the carbonyl of Leu119. Our testing here, until the veracity of F_3Ala as a *bona fide* strategy is confirmed, is inconclusive. It does suggest that no such hydrogen bond exists. Another way to test for this is to substitute TrpB for a different amino acid that also does not contain an indole N-H. For this experiment, 3-benzothiophenyl alanine (3-BTA) was chosen. It contains a sulfur atom in place of the indole N-H. While sulfur is significantly larger, it is no longer a hydrogen bond donor.

A 3-BTA substitution was conducted using nonsense suppression at TrpB in the α4 subunit to determine whether the NpAla effect is due to hydrogen bonding (Table 8). The mutation showed a 13-fold loss of function, which directly contradicts the NpAla results. To ensure that these anomalous results were not due to the changes in the electrostatics of the cation-π interaction, the 'expected' binding energies due to the differential cation-π binding energies were calculated (Table 9).

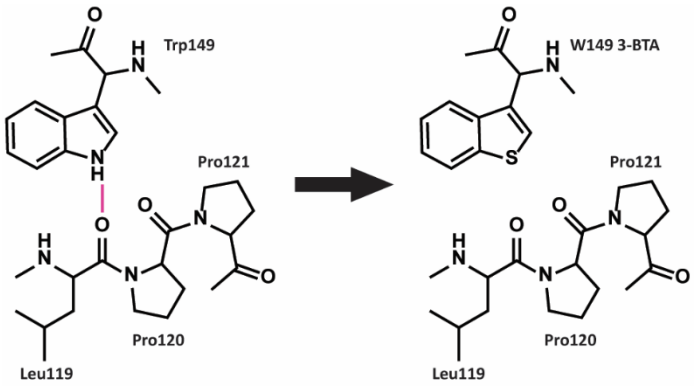


Figure 11: Incorporation of 3-BTA for TrpB to test indole N-H

The predicted 'theoretical' EC_{50} values can be determined from the regression derived from the original $(\alpha4)_2(\beta2)_3$ fluorination plot analysis,

Table 8: Incorporation of 3-BTA at the TrpB site in α4β2 nAChRs

	EC_{50} (μM)	n_H	I_{max} (μA)	Fold Shift	N
Trp	0.6 ± 0.01	1.19 ± 0.03	1.87 - 5.33	1.0	11
3-BTA	8.0 ± 0.3	1.38 ± 0.05	0.99 - 13.37	13	17

Table 9: Comparing experimental EC_{50} values to those predicted from α4β2 TrpB fluorination plot

	Cation-π binding (kcal/mol)*	Predicted EC_{50}	Empirical EC_{50}	New fold-shift
Trp	15.29	0.49	0.54	1.1
NpAla	13.5	1.55	0.12	0.1
3-BTA	13.24	1.83	8	4.4

*Computed binding energies from DFT//M06/6-31G(d,p) calculations of a NMe_4^+ ion to indole, naphtylene, or 3-benzothiophene; predicted EC_{50} based on revised fluorination plots from Chapter 3. Trend line: y = -0.28 + 4.25; Fold-shift = Empirical EC_{50}/Predicted EC_{50}

adjusted for M06 calculations (see Chapter 3). Using this value, a fold shift from the predicted value can be determined. This is essentially a fold shift that is independent from the change in cation-π binding. These values (Table 9) suggest that the NpAla mutation actually causes a 10-fold gain of function, and the 3-BTA a more modest 4-fold loss of function. In any event, it seems unlikely that electrostatics are causing the observed effect or the discrepancy observed.

This mutation remains a mystery, and more work will have to be done to validate its reason. However, as discussed in Chapter 3, data from crystal structures, while incredibly useful in the aggregate, must be treated with caution. In this case, a hydrogen bond that appears in several such crystal structures is likely not present. The most likely reason for the discrepancy is a steric effect: the binding site in α4β2 is sensitive to such steric changes, and it is possible that changes to opposite sides of TrpB could cause differing effects. However, there is not yet enough data to substantiate such a claim.

In the course of these studies, two different ncAAs have ben incorporated in a neuronal nAChR, a technique that can likely be applied to other potential interactions and in different neuronal nAChR subtypes. Further, the strategy of utilizing aliphatic fluorinated amino acids was investigated, and it is likely that the use of fluorination on the β carbon of an amino acid will cause a weakening of hydrogen bonds. This is a useful addition to our lab's toolkit to investigate noncovalent interactions in LGICs.

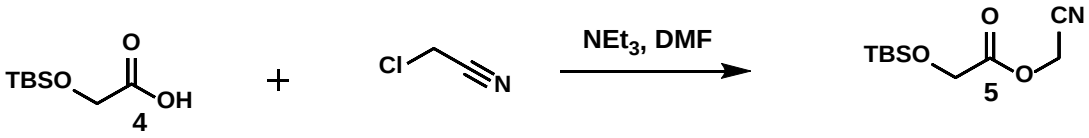
4.7: Methods

Organic Synthesis

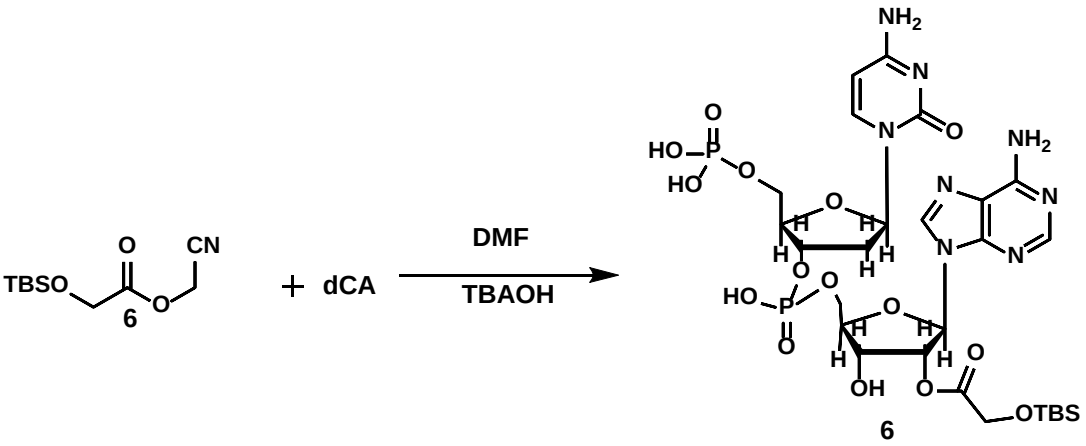
Unless otherwise stated, reactions were carried out under argon. Commercially available reagents were obtained from Sigma Aldrich, AK Scientific, Alfa Aesar, or Acros Organics and used without further purification. Solvents were used as received. Thin-layer chromatography with Sigma Aldrich silica gel coated plates with fluorescent indicator (0.25 mm) was used to monitor reactions. Silica gel chromatography was conducted as described by Still et al.[24], with silica gel purchased from Alfa Aesar (60 Å, 230-400 mesh). NMR spectra were recorded on Varian (300, 400, 500, or 600 MHz) spectrometers. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI) spectra were used to confirm the masses and structures of dCA and tRNA coupled amino acids. Agilent LC-MS was used to obtain mass spectra of precursor compounds. Isolated yields are reported except for compounds purified by HPLC, where crude recovery yields are reported and only a portion of the crude was purified by HPLC.

4

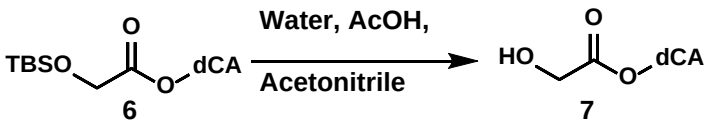
Glycolic acid (50 mg, 0.657 mmol) was mixed with imidazole (44.72 mg, 0.657 mmol) in 6.5 mL dry DCM. To the mixture was added *tert*-butyldimethylsilyl chloride (99 mg, 0.788 mmol). The reactants were stirred at room temperature overnight and it was then quenched by addition of water (150 mL) to form a slurry mixture. This was then extracted with diethyl ether (2 × 40 mL). The organic solutions were combined, dried over anhydrous MgSO4 and concentrated to yield the **4**. ^{1}H NMR (300 MHz, Chloroform-d) δ 4.25 – 4.15 (m, 2H), 0.91 (s, 9H), 0.10 (s, 6H). m/z = 190.3. Yield: 54%.



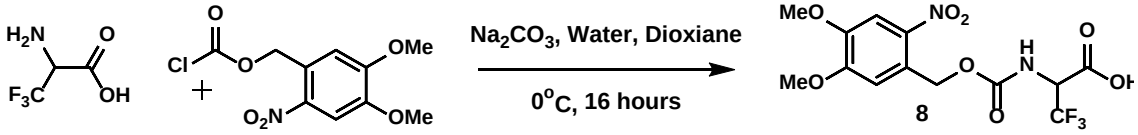
A 5-mL round bottom flask was flame dried with a stir bar, under vacuum, then backfilled with argon. **4** (77 mg, 0.354 mmol) was added to the flask and dissolved in 2 mL DMF. 2 mL chloroacetonitrile was added, followed by 125 µL (0.885 mmol) triethylamine. The reaction was stirred for 24 hours at room temperature, then diluted with ether and washed with water. The organic layer was then concentrated in vacuo and the resulting product was purified by silica gel chromatography. ^{1}H NMR (500 MHz, Chloroform-d) δ 4.83 (s, 2H), 4.39 (s, 2H), 0.98 (s, 9H), 0.17 (s, 6H). m/z = 229.4. Yield: 84%.



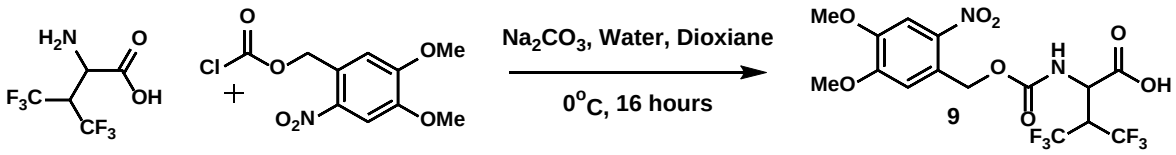
A vial filled with purified dCA and *tert*-butylammonium hydroxide was flushed with argon. Add **6** (1 mg/mL solution in DMF; 0.8 mL), then let stir overnight at room temperature. Purify by HPLC and confirm successful conjugation by MALDI mass spectrometry using a 6-aza-2-thiothymine (ATT)/ammonium citrate matrix. MALDI result: m/z = 808.2.



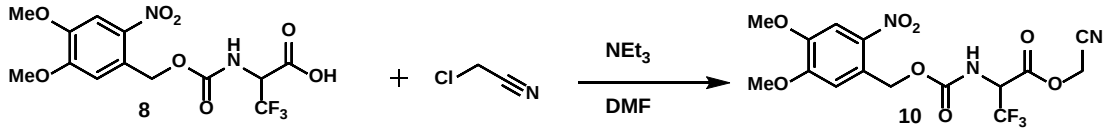
After lyophilization of the dCA coupling, **6** was dissolved in a 1:1:3 mixture of water, acetonitrile:, and acetic acid (total volume: 1.5 mL). Shake overnight at 37°C. Once complete, lyophilize off the solvent, and confirm the reaction is complete via MALDI mass spectrometry using an 6-aza-2-thiothymine (ATT)/ammonium citrate matrix. MALDI result: m/z = 694.1. This product was used in the tRNA ligation reaction.



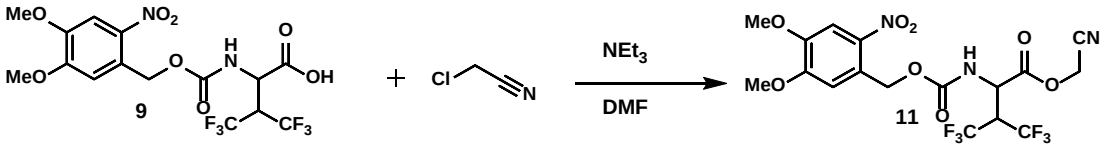
(±)-Trifluoroalanine (25 mg, 0.175 mmol) was dissolved in a 1:1 mixture of water and dioxane (2 mL). Sodium carbonate (93 mg, 0.875 mmol) was added and the mixture cooled to 0°C in an ice bath under stirring. 4,5-Dimethoxy-2-nitrobenzyl chloroformate (58 mg, 0.21 mmol) was added and the mixture stirred for 16 hours overnight, slowly coming to room temperature. Concentrate to remove the dioxane and partition between ethyl acetate/water. Acidify aqueous layer to pH 2-3 before extracting again with ethyl acetate. **8** was obtained as a yellow solid. ^{1}H NMR (400 MHz, Acetonitrile-d3) δ 7.73 (s, 1H), 7.15 (s, 1H), 6.82 (d, J = 9.2 Hz, 1H), 5.50 (s, 2H), 5.08 (p, J = 8.5 Hz, 1H), 3.96 (d, J = 1.5 Hz, 3H), 3.91 (s, 3H). ^{19}F NMR (400 MHz, Acetonitrile-d3) δ -73.00 (d, J = 8.1 Hz). m/z = 382.06. Yield: 96%.



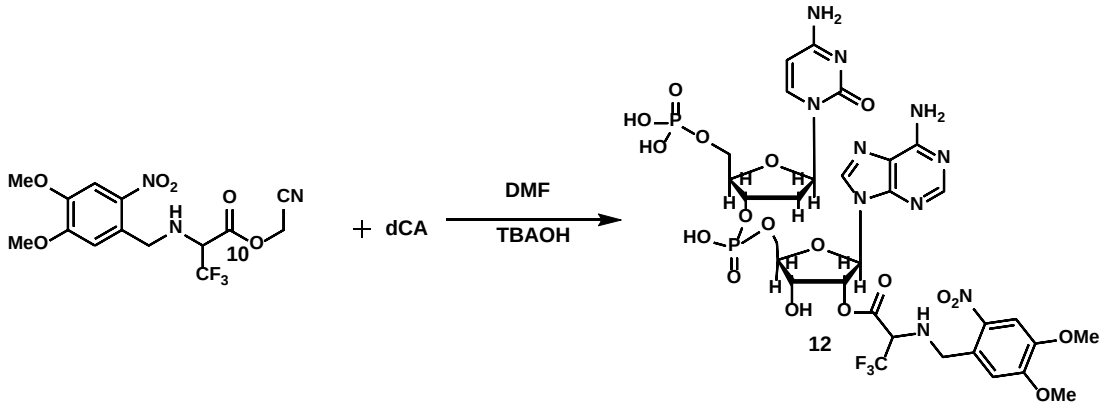
(±)-Hexafluorovaline (50 mg, 0.222 mmol) was dissolved in a 1:1 mixture of water and dioxane (2 mL). Sodium carbonate (118 mg, 1.11 mmol) was added and the mixture cooled to 0°C in an ice bath under stirring. 4,5-Dimethoxy-2-nitrobenzyl chloroformate (73.5 mg, 0.266 mmol) was added and the mixture stirred for 16 hours overnight, slowly coming to room temperature. Concentrate to remove the dioxane and partition between ethyl acetate/water. Acidify aqueous layer to pH 2-3 before extracting again with ethyl acetate. **9** was obtained as a yellow solid. ^{1}H NMR (400 MHz, Acetonitrile-d3) δ 7.69 (s, 1H), 7.11 (d, J = 11.6 Hz, 1H), 6.50 (s, 1H), 5.49 (s, 2H), 5.13 (s, 1H), 4.92 (s, 0H), 4.28 (s, 1H), 3.92 (s, 4H), 3.90 – 3.88 (m, 5H). ^{19}F NMR (376 MHz, Acetonitrile-d3) δ -62.74 , -66.56. Yield: 90%.



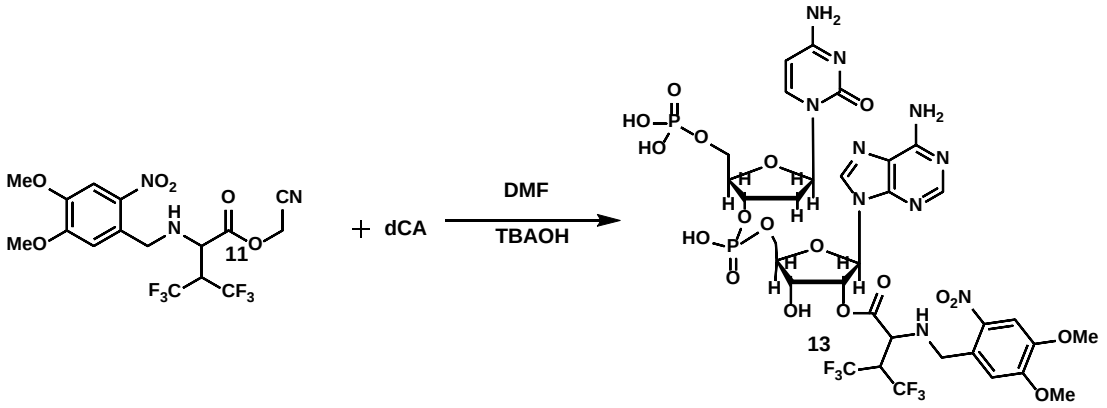
A 5-mL round bottom flask was flame dried with a stir bar, under vacuum, then backfilled with argon. **8** (60 mg, 0.158 mmol) was added to the flask and dissolved in 2 mL DMF. 2 mL chloroacetonitrile was added, followed by 75 μL (0.531 mmol) triethylamine. The reaction was stirred for 24 hours at room temperature, then diluted with ether and washed with water. The organic layer was then concentrated in vacuo and the resulting product **10** was purified by silica gel chromatography. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.10 (s, 1H), 7.51 (s, 1H), 5.87 (d, J = 0.8 Hz, 2H), 5.63 (dq, J = 9.4, 7.9 Hz, 1H), 5.31 (d, J = 0.7 Hz, 2H), 5.29 (s, 1H), 4.32 (s, 4H), 4.28 (s, 3H). ^{19}F NMR (282 MHz, Chloroform-d) δ -67.65 (d, J = 7.8 Hz). Yield: 78%



A 5-mL round bottom flask was flame dried with a stir bar, under vacuum, then backfilled with argon. **9** (60 mg, 0.158 mmol) was added to the flask and dissolved in 2 mL DMF. 2 mL chloroacetonitrile was added, followed by 75 μL (0.531 mmol) triethylamine. The reaction was stirred for 24 hours at room temperature, then diluted with ether and washed with water. The organic layer was then concentrated in vacuo and the resulting product **11** was purified by silica gel chromatography. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.30 (s, 1H), 8.08 (s, 1H), 7.50 (s, 1H), 7.21 (d, J = 10.2 Hz, 1H), 5.89 – 5.84 (m, 2H), 5.74 (d, J = 10.0 Hz, 1H), 5.29 (d, J = 0.5 Hz, 2H), 4.32 (s, 3H), 4.26 (s, 3H). ^{19}F NMR (282 MHz, Chloroform-d) δ -57.57 (p, J = 8.9 Hz), -61.00 (h, J = 9.1 Hz). Yield: 65%.



To a vial filled with purified dCA and *tert*-butylammonium hydroxide, flush with argon. Add **10** (1 mg/mL solution in DMF; 0.8 mL), then let stir overnight at room temperature. Purify by HPLC and confirm successful conjugation by MALDI mass spectrometry using an 6-aza-2-thiothymine (ATT)/ammonium citrate matrix. MALDI result: m/z = 1000.16. This product (**12**) was taken on to tRNA ligation.



To a vial filled with purified dCA and *tert*-butylammonium hydroxide, flush with argon. Add **11** (1 mg/mL solution in DMF; 0.8 mL), then let stir overnight at room temperature. Purify by HPLC and confirm successful conjugation by MALDI mass spectrometry using an 6-aza-2-thiothymine (ATT)/ammonium citrate matrix. MALDI result (**13**): m/z = 1082.16. This product (**13**) was taken on to tRNA ligation.

Molecular Biology

Rat α4L9'A (described as wild-type and/or α4 throughout this chapter) and β2 nAChRs were in the pGEMhe vector, a cDNA plasmid optimized for protein expression in *Xenopus* oocytes. Site-directed mutagenesis was performed by PCR using the Stratagene QuikChange protocol and primers ordered from Integrated DNA Technologies (Coralville, IA). Circular cDNA was linearized with SbfI (New England Biolabs, Ibswich, MA) and then transcribed in vitro using T7 mMessage mMachine kit (Life Technologies, Santa Clara, CA), with a purification step after each process (Qiagen, Valencia, CA). Final concentrations were quantified by UV spectroscopy.

Ion Channel Expression

Xenopus laevis oocytes (stage V to VI) were sourced from both a Caltech facility and Ecocyte Bio Science (Austin, TX). Oocytes were injected with 50 nL solution

containing either 5 or 10 ng mRNA. The α4L9'A:β2 mRNA mass ratio was typically 1:3 for wild-type and conventional mutagenesis experiments. Cells were incubated 24-48 hours at 18°C in ND96 solution (96 mM NaCl, 2mM KCl, 1 mM $MgCl_2$, and 5mM HEPES, pH 7.5) enriched with theophylline, sodium pyruvate, and gentamycin.

Non-canonical Amino Acid Incorporation

The cyanomethylester form of NVOC-protected tryptophan, leucine, alanine, and valine and their corresponding analogues was coupled to dinucleotide dCA and enzymatically ligated to either UAG-suppressor 74-mer THG73 $tRNA_{CUA}$ or UGA-suppressor 74-mer TQOPS'$_{UCA}$ as previously described.[4] The product was verified by MALDI time-of-flight mass spectrometry on a 3-hydroxypicolinic acid matrix. The non-canonical amino acid-coupled tRNA was deprotected by photolysis on a 500 W Hg/Xe arc lamp, filtered with Schott WG-320 and UG-11 filters, immediately prior to coinjection with mRNA containing the UAG or UGA mutation at the site of interest. mRNA and tRNA were typically injected in a 1:1 or 1:2 volume ratio in a total volume of 50 or 75 nL respectively so that 25 ng of mRNA was injected per cell. For double mutant experiments α4L9'AW149TAGβ2L119TGA was injected at a 1:10 mRNA ratio to maintain the $(\alpha4)_2(\beta2)_3$. Stoichiometry was confirmed via the Hill coefficient (cases of a shift in stoichiometry resulted in a large increase in this value). In cases where observed agonist-induced currents were low after 48 hour incubation – likely due to low receptor protein expression – a second injection of mRNA and tRNA was performed after 24 hours. The fidelity of non-canonical amino acid incorporation was confirmed at Trp with a wild-type recovery experiment where tryptophan was loaded onto tRNA. If this experiment yielded similar to EC_{50} to wild-type then the cell incorporated the charged residue and nothing else.

This was accomplished with the Tyr sites by comparing tRNA charged with Phe to a conventional Y-Phe mutation. A read-through/reaminoacylation test served a negative control by injecting unacylated full-length 76-mer tRNA. Results of these controls are discussed in the chapter.

Whole-Cell Electrophysiological Characterization

Acetylcholine chloride and lobeline were purchased from Sigma Aldrich (St Louis, MO). Agonist-induced currents were recorded in TEVC mode using the OpusXpress 6000A (Molecular Devices, Sunnyvale, CA) at a holding potential of -60 mV in a running buffer of Ca^{2+}-free ND96. Agonists were prepared in Ca^{2+}-free ND96 and delivered to cells via a 1 mL application over 15 sec followed by a 2 min wash. Data from dose-response experiments were normalized, averaged, and fit to the Hill equation using Kaleidagraph (Synergy Software, Reading PA), though data are visualized here with Prism (GraphPad Software, La Jolla, CA). Error bars are presented as standard error of the mean, while EC_{50} and Hill coefficient errors are reported by Kaleidagraph and represent the sum of the squared error between the data and the calculated fit.

Computational Evaluation of Hydrogen Bonding

All calculations were performed in Spartan 14 using density functional theory with the M06 functional and the 6-31G(d,p) basis set.[25] Electrostatic potential maps with high resolution were calculated using the Spartan 14 graphical user interface. Electrostatic potential maxima (for indole) and minima (for the acceptors) were determined, and the α, β, and $\Delta\Delta G_{H\text{-}bond}$ values were calculated using methods described in section 6.4.1 with values for α_s and β_s sourced from previously published work.

NMR Determination of Hydrogen Bond Acceptor Strength

All compounds were either purchased from Sigma Aldrich (St Louis, MO) or synthesized using published methods. Solutions of p-fluorophenol and p-fluoroanisole, both at 0.01 M, and hydrogen bond acceptor in varying concentrations were prepared in purified carbon tetrachloride. ^{19}F-NMR measurements were taken on a 400 MHz instrument (Varian Associates, Palo Alto, CA). K_f was calculated as described in section 6.4.2 and ΔG was determined using the van't Hoff equation, $\Delta G = -RT\ln(K_f)$.

4.8: References

1. Van Arnam, E. B.; Dougherty, D. A., Functional Probes of Drug–Receptor Interactions Implicated by Structural Studies: Cys-Loop Receptors Provide a Fertile Testing Ground. *Journal of Medicinal Chemistry* **2014,** *57* (15), 6289-6300.
2. Xiu, X.; Puskar, N. L.; Shanata, J. A.; Lester, H. A.; Dougherty, D. A., Nicotine binding to brain receptors requires a strong cation-pi interaction. *Nature* **2009,** *458* (7237), 534-7.
3. Beene, D. L.; Dougherty, D. A.; Lester, H. A., Unnatural amino acid mutagenesis in mapping ion channel function. *Current Opinion in Neurobiology* **2003,** *13* (3), 264-270.
4. Dougherty, D. A.; Van Arnam, E. B., In Vivo Incorporation of Non-canonical Amino Acids by Using the Chemical Aminoacylation Strategy: A Broadly Applicable Mechanistic Tool. *ChemBioChem* **2014,** *15* (12), 1710-1720.
5. Wagner, C. A.; Friedrich, B.; Setiawan, I.; Lang, F.; Bröer, S., The Use of Xenopus laevis Oocytes for the Functional Characterization of Heterologously Expressed Membrane Proteins. *Cellular Physiology and Biochemistry* **2000,** *10* (1-2), 1-12.
6. Benjamin Peng, H., Appendix A: Solutions and Protocols. *Methods in Cell Biology* **1991,** *36*, 657-662.
7. Gleitsman, K. R.; Kedrowski, S. M. A.; Lester, H. A.; Dougherty, D. A., An Intersubunit Hydrogen Bond in the Nicotinic Acetylcholine Receptor That Contributes to Channel Gating. *Journal of Biological Chemistry* **2008,** *283* (51), 35638-35643.
8. Blum, A. P.; Lester, H. A.; Dougherty, D. A., Nicotinic pharmacophore: The pyridine N of nicotine and carbonyl of acetylcholine hydrogen bond across a subunit interface to a backbone NH. *Proceedings of the National Academy of Sciences* **2010,** *107* (30), 13206-13211.

9. Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A., From ab initio quantum mechanics to molecular neurobiology: A cation–π binding site in the nicotinic receptor. *Proceedings of the National Academy of Sciences* **1998,** *95* (21), 12088-12093.
10. Colquhoun, D., Binding, gating, affinity and efficacy: The interpretation of structure-activity relationships for agonists and of the effects of mutating receptors. *British Journal of Pharmacology* **1998,** *125* (5), 923-947.
11. Morales-Perez, C. L.; Noviello, C. M.; Hibbs, R. E., X-ray structure of the human alpha4beta2 nicotinic receptor. *Nature* **2016,** *538* (7625), 411-415.
12. Post, M. R. Agonist Binding Studies at Two Subtypes of the Nicotinic Acetylcholine Receptor Involved in Parkinson's Disease and Addiction. California Institute of Technology, Pasadena, CA, 2016.
13. Horovitz, A., Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Folding and Design* **1996,** *1* (6), R121-R126.
14. Gleitsman, K. R.; Shanata, J. A. P.; Frazier, S. J.; Lester, H. A.; Dougherty, D. A., Long-Range Coupling in an Allosteric Receptor Revealed by Mutant Cycle Analysis. *Biophysical Journal* **2009,** *96* (8), 3168-3178.
15. Wall-Lacelle, S.; Hossain, M. I.; Sauvé, R.; Blunck, R.; Parent, L., Double Mutant Cycle Analysis Identified a Critical Leucine Residue in the IIS4S5 Linker for the Activation of the CaV2.3 Calcium Channel. *Journal of Biological Chemistry* **2011,** *286* (31), 27197-27205.
16. Saks, M. E.; Sampson, J. R.; Nowak, M. W.; Kearney, P. C.; Du, F.; Abelson, J. N.; Lester, H. A.; Dougherty, D. A., An Engineered Tetrahymena tRNAGln for in Vivo Incorporation of Unnatural Amino Acids into Proteins by Nonsense Suppression. *Journal of Biological Chemistry* **1996,** *271* (38), 23169-23175.
17. Rodriguez, E. A.; Lester, H. A.; Dougherty, D. A., Improved amber and opal suppressor tRNAs for incorporation of unnatural amino acids in vivo. Part 2: Evaluating suppression efficiency. *RNA* **2007,** *13* (10), 1715-1722.
18. Cametti, M.; Crousse, B.; Metrangolo, P.; Milani, R.; Resnati, G., The fluorous effect in biomolecular applications. *Chemical Society Reviews* **2012,** *41* (1), 31-42.
19. Zhang, W., Fluorous Synthesis of Heterocyclic Systems. *Chemical Reviews* **2004,** *104* (5), 2531-2556.
20. Abraham, M. H.; Platts, J. A., Hydrogen Bond Structural Group Constants. *The Journal of Organic Chemistry* **2001,** *66* (10), 3484-3491.
21. Hunter, C. A., Quantifying Intermolecular Interactions: Guidelines for the Molecular Recognition Toolbox. *Angewandte Chemie International Edition* **2004,** *43* (40), 5310-5324.

22. Gurka, D.; Taft, R. W., Studies of hydrogen-bonded complex formation with p-fluorophenol. IV. Fluorine nuclear magnetic resonance method. *Journal of the American Chemical Society* **1969,** *91* (17), 4794-4801.
23. Puskar, N. L.; Lester, H. A.; Dougherty, D. A., Probing the Effects of Residues Located Outside the Agonist Binding Site on Drug-Receptor Selectivity in the Nicotinic Receptor. *ACS Chemical Biology* **2012,** *7* (5), 841-846.
24. Still, W. C.; Kahn, M.; Mitra, A., Rapid chromatographic technique for preparative separations with moderate resolution. *The Journal of Organic Chemistry* **1978,** *43* (14), 2923-2925.
25. Wavefunction Inc, I. C., Spartan '14.

Chapter 5: Synthesis and characterization of a class of cyanine dyes with exceptionally large Stokes shifts

5.1: Abstract

Fluorescence microscopy is an often-used technique that takes advantage of the fluorescence of chromophores to detect biomarkers or measure conditions *in vivo*. Cyanine dyes are useful in that they are bright fluorophores: they have a high absorptivity and quantum yield of fluorescence. However, their usefulness in fluorescence microscopy is somewhat limited because of their small Stokes shift. This chapter describes the synthesis and characterization of a number of cyanine dyes that contain a *meso*-acetylene-aromatic moiety. These dyes, because of orbital mixing between the acetylene-aromatic LUMO and a high lying cyanine orbital, exhibit both normal cyanine fluorescence and cyanine emission when excited with a shorter wavelength. This phenomenon arises from excitation to the S2 state followed by subsequent relaxation the S1 state and normal fluorescence. The brightness of this excitation's emission is comparable to other commercial fluorophores. This mechanism is supported by TDDFT calculations. Crucially, dyes without the acetylene unit were an order of magnitude worse at stimulating short wavelength emission because they stuck in a perpendicular conformation, with rotation barriers roughly an order of magnitude larger than acetylene-containing dyes. In addition, this chapter describes work towards explaining the unorthodox absorbance of an acetylene-phenol Cy5 compound.

5.2: Introduction

Fluorescence microscopy remains one of the most powerful tools to study biological processes *in vivo*. Commonly used fluorescent dyes typically have very high brightness: a large extinction coefficient coupled with efficient fluorescence.[1-3] Polymethine (cyanine), triphenylmethyl cation (fluorescein, rhodamine), pyrene, and coumarin dyes fit these characteristics and derivatives of such dyes are commonly used as fluorescent sensors.[4-8] Cyanine dyes in particular have been used for decades in this application. The absorbance and fluorescence of these dyes can be altered by adjusting the length of the polymethine chain. Synthesis of these dyes is simple and modular, and the end groups and polymethine chain are synthesized separately. This enables attachment of various bioorthogonal reactive groups. Many syntheses of these dyes include modification of the terminal groups or, in the case of trimethine or heptamethine dyes, modification on the central carbon by nucleophilic substitution.[9-10] This work has produced many dyes that are effective as sensors for environmental conditions such as pH and metal concentration. While cyanine dyes are invaluable in modern fluorescence microscopy, they do have some limitations. Their small (20-30 nm) Stokes shift limits their utility, necessitating the use of these dyes in either more complex systems such as FRET or with filter sets that cut out a large portion of both the absorbance and emission spectra.

Organic dyes with large Stokes shifts (in the range of 50-200 nm) have mechanisms that involve charge reorganization in the excited state.[11-17] One example is dual fluorescence: a small Stokes shift fluorescence from a state structurally similar to the ground state (the Franck-Condon state) and a large Stokes shift fluorescence from a reorganized excited state.[5] A prototypical example is the photoacid phenomenon, where

molecules such as 2-naphthol emit from both the protonated state and a transiently stabilized deprotonated state, the latter emission being red-shifted.[18-21] Exceptionally large Stokes shifts can be observed is systems set up for efficient intramolecular energy transfer. For example, a recently reported BOPHY-pyrene dye exhibited exceptionally large Stokes shift via excitation of the pyrene and subsequent energy transfer to and fluorescence from the BOPHY dye.[22] Large Stokes shifts can also be achieved by excitation to the S2 state followed by thermal relaxation to the first excited state, and subsequent emission via Kasha's Rule.[23-26] However, molecules that undergo this mechanism generally do not make for useful fluorescent dyes, as the S0-S2 transition is inefficient.

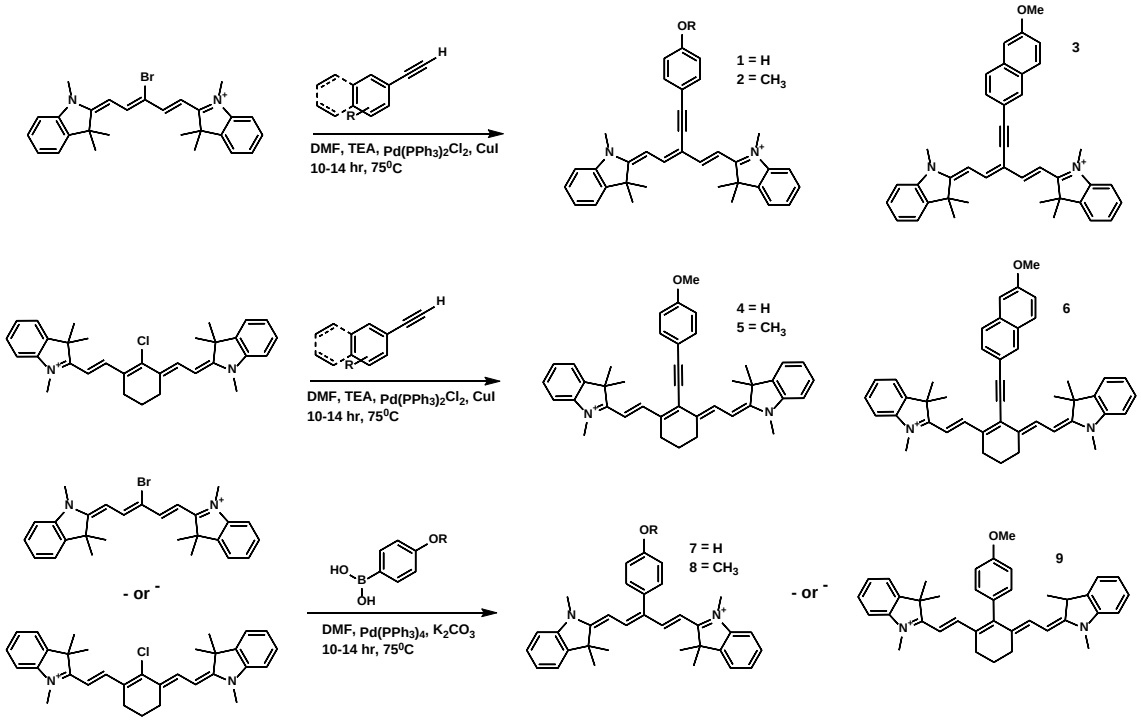
Previous work altered the structure of cyanine dyes to increase their Stokes shift. Cyanine dye hybrids have recently been synthesized to combine elements of a carbocyanine and BODIPPY.[6, 27] While the dye had a moderately large Stokes shift, the synthesis was complex and offered no general solution to creating a class of dyes that, like cyanine dyes, can be easily modified and functionalized to allow for changes in either photophysical properties or bioorthogonal labeling. Here we describe a class of alkyne-linked cyanine-aromatic chromophores. The synthesis of these dyes feature a key Sonagashira reaction from easily-available building blocks. The resulting dyes feature predictable fluorescence at the cyanine absorbance band as well as at a shorter wavelength, with large Stokes shifts of >250 nm, depending on the chromophore structure. The large Stokes shift is mediated via an excitation to the S2 state and subsequent relaxation to the cyanine excited state and normal cyanine fluorescence. The S2 state is made more available by the mixing of the LUMO of the aromatic moiety with a high lying cyanine orbital in a planar conformation, and a lowered rotation barrier due to the acetylene group. Analogous

dyes containing an aromatic group but no acetylene linker exhibited much lower quantum yields for the S2-mediated fluorescence.

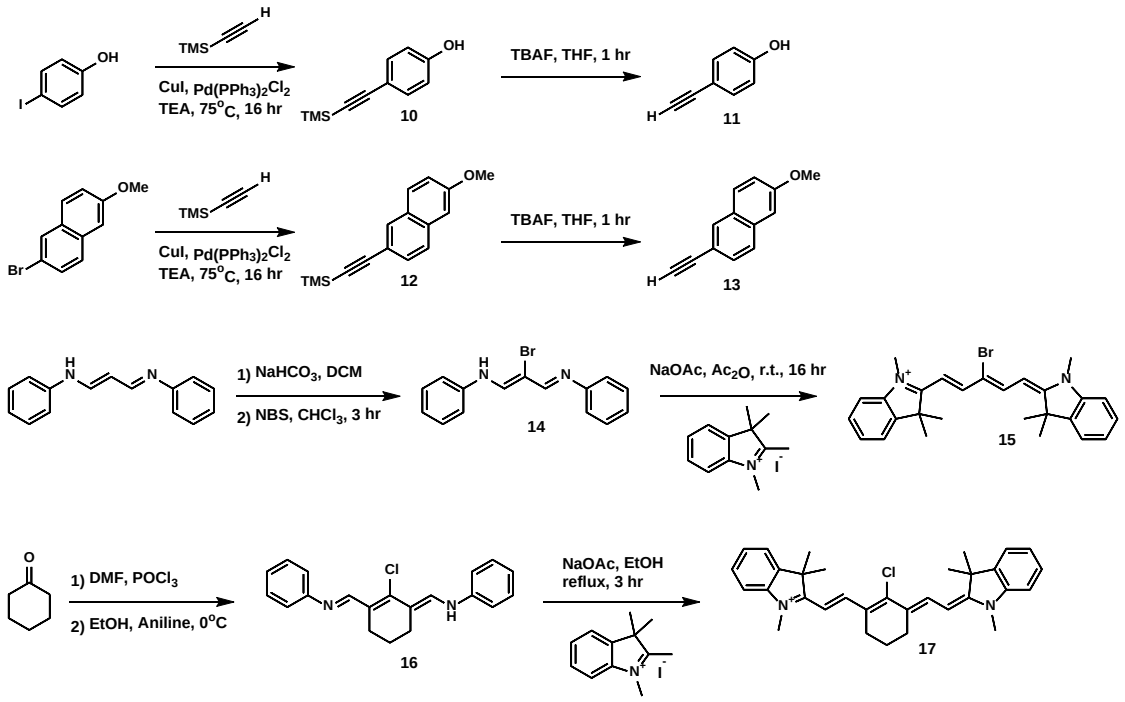
5.3: Synthetic strategy

Scheme 1 shows the syntheses of dyes **1-9**. A Sonagashira reaction was utilized to produce dyes **1-6**, and dyes **7-9** were synthesized through a Suzuki reaction. The use of the Sonagashira and Suzuki reactions to modify cyanine dyes has been detailed previously, and has not been adjusted significantly for this work.[28] Reaction yields are not high, and purification can be challenging. However, due to the brevity and simplicity of the overall process, spectroscopically relevant quantities of all dyes could be produced with relative ease. The inherent high absorptivity of the cyanine dye helps in this as well.

Scheme 2 details the preliminary synthetic steps to produce either partner in the cyanine Sonagashira reaction. Most of these reaction steps have precedent in the literature,



Scheme 1: Key synthetic step of the synthesis of aromatic-acetylene-cyanine dyes **1-6** and phenol-cyanine dyes **7-9** using palladium-catalyzed cross coupling reactions.



Scheme 2: Synthesis of alkyne and halogen-containing synthetic precursers for the cyanine dyes of interest.

and are either very high-yielding or can be scaled up easily to produce a useful amount of product. See the methods section for specific details on these reactions. Compounds **10-13** are aromatic-alkyne groups, and compounds **14-17** are common cyanine chromophore precursors that have extensive precedent in the literature.[29-30]

5.4: Photophysical studies of alkyne-linked cyanine dyes

5.4.1: Steady-state absorbance and long-wavelength fluorescence

All dyes featured normal cyanine absorbance and fluorescence (Figure 1) with small Stokes shifts of 20-30 nm and quantum yields roughly equivalent to previously measured Cy5 and Cy7 fluorescence (Table 1). The two series of dyes, Cy5 and Cy7, refer to the number methine units in their polymethine chains. Cy5 dyes absorb red light (600-640 nm) and appear blue, while Cy7 dyes absorb NIR light (750-800 nm) and appear green. All dyes were soluble in polar organic solvents. Cy5 compounds (**1**-**3**, and **7**-**8**) were

Table 1: Photophysical properties of synthesized dyes. Short wavelength refers to the large Stokes Shift fluorescence observed upon irradiation at 350-450 nm. Long wavelength refers to fluorescence from excitation of the cyanine band. In both cases, emission wavelength is identical. λ_{max}: maximum excitation or emission wavelength. ε: extinction coefficient. φ_{em}: quantum yield of emission. φ_{em} x ε: Fluorescence Brightness

	Long Wavelength Excitation				Short Wavelength Excitation				Emission
Dye	**λ_{max} (nm)**	**ε x10^5 ($M^{-1}cm^{-1}$)**	**φ_{em}**	**φ_{em} x ε**	**λ_{max} (nm)**	**ε x10^5 ($M^{-1}cm^{-1}$)**	**φ_{em}**	**φ_{em} x ε**	**λ_{max} (nm)**
1	583*, 638	0.25	0.024	600	367	0.12	0.0055	66	655
2	648	2	0.16	32000	370	0.14	0.069	966	674
3	648	2	0.18	36000	415	0.11	0.06	660	674
4	770	1	0.16	16000	362	0.05	0.09	450	788
5	795	1	0.081	8100	347	0.092	0.03	276	824
6	795	1	0.091	9100	369, 488	0.092	0.049	451	824
7	638	1	0.081	8100	-	-	-	-	655
8	641	1	0.029	2900	375	0.18	0.0036	65	655
9	751	0.5	0.35	17500	430	0.1	0.0036	36	772

*Nonemissive band

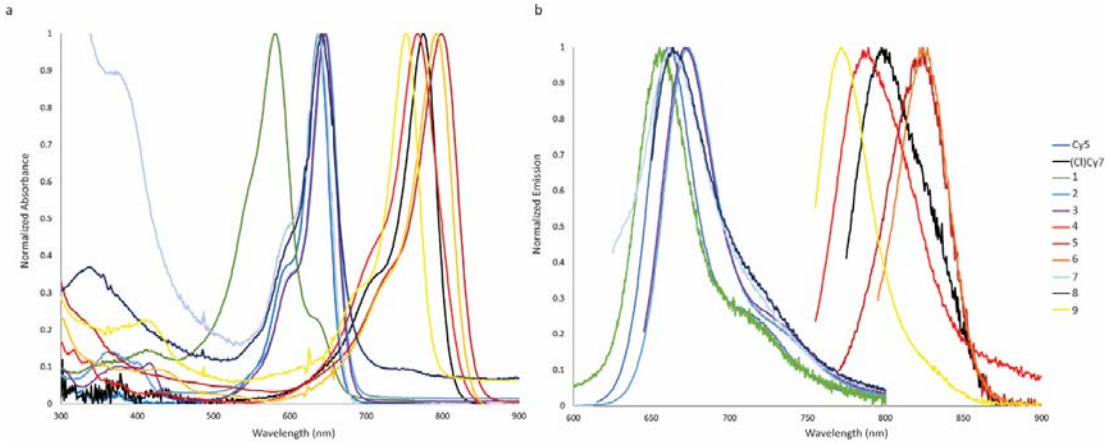


Figure 1: (a) Full absorption spectra and (b) emission spectra of Dyes **1-9** and reference Cy5 and halogenated Cy7 dyes. Emission spectra recorded by excitation of the cyanine absorbance band.

slightly soluble in water, while Cy7 dyes (**4-6** and **9**) aggregated in water into nonemissive H-aggregates (Figure 2).[31-34]

5.4.2: Short-wavelength fluorescence and excited state excimers

In addition to the normal cyanine excitation, emission can be induced by excitation at a shorter wavelength. For the Cy5 compounds **1**, **2**, and **3** this excitation was in the 350-370 nm range. With the Cy7 dyes **4**, **5**, and **6** this excitation is at 350-500 nm. Naphthylene-containing dyes **3** and **6** exhibit slightly extended short-wavelength excitation ranges

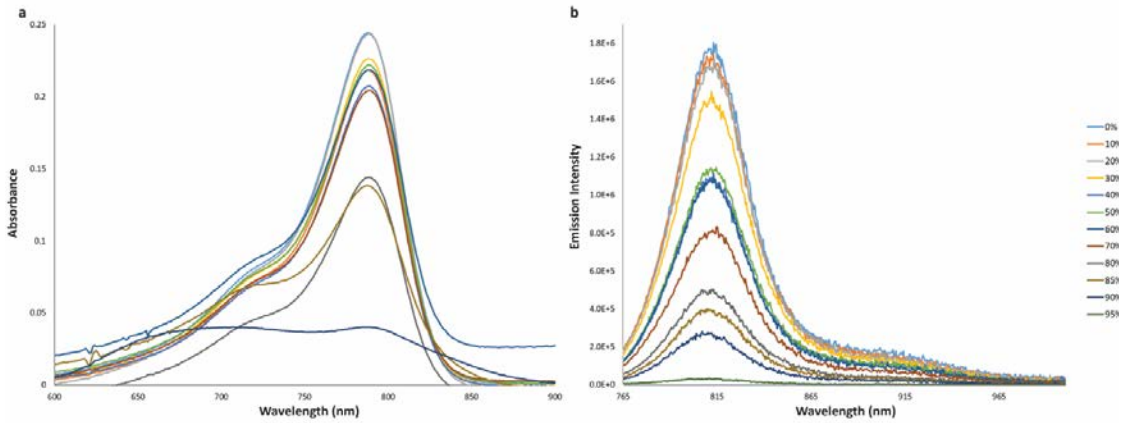


Figure 2: (a) Absorbance and (b) emission of **5** at various concentrations of water in acetonitrile. Percentages refer to the percent volume of water. Blue-shifted absorbance band is an H-aggregate of **5**.

(Figure 3). Short-wavelength fluorescence did not occur with unmodified Cy5 or meso-Cl-Cy7, and notably did not occur with dyes **5** or **6**, homologues to **1** and **3** without the acetylene unit.

Normal and high-energy excitation spectra of the synthesized dyes are shown in Figure 3 on a brightness scale. The fluorescence brightness from short wavelength excitation was higher for the Cy5 dyes **1** and **2**. While the intensity of this fluorescence was never more than 1/5th of the normal cyanine fluorescence, it was still bright enough to be observed at concentrations as low as 0.5 μM. Quantum yields of emission from short wavelength excitation were between one third and two thirds of the long wavelength quantum yields for dyes **1-6**, but several orders of magnitude less for **8** and **9** (Table 1). Short wavelength extinction coefficients are on the same order of magnitude as the native aromatic group.

Most dyes did not exhibit fluorescence in the short-wavelength regime. However, dye **6** did at moderate concentrations emit at 550 nm when excited with 315-400 nm. This emission was determined to originate from an excimer by dilution studies (Figure 4). Upon 1:2 dilution of the dye, the short-wavelength fluorescence decreased by a much wider

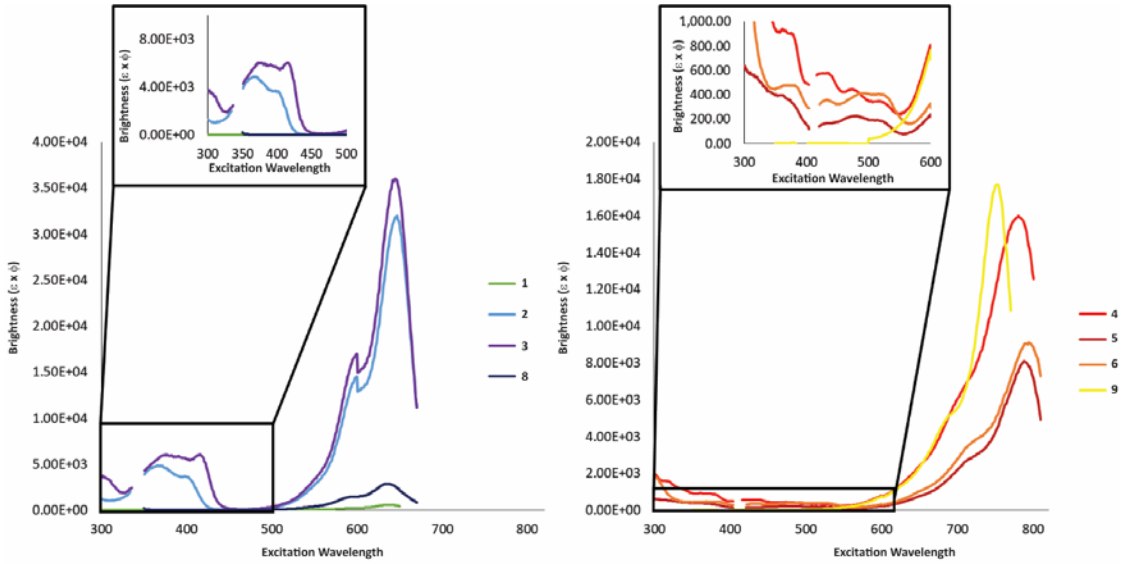


Figure 3: Brightness of dyes (a) **1-3**, and **8**; and (b) **4-6** and **9** in acetonitrile. Quantum yield and extinction coefficient data for dyes were taken from Table 1 and Figure 1. Gaps in spectra in short wavelength regime were to remove backscattering from the observed emission wavelength in the excitation spectra. The discontinuity at 550 nm signifies a switch between the short and long wavelength quantum yield in the brightness measurement of each dye. Inset: brightness of dyes in the short-wavelength range.

margin. We can conclude that the increased nonpolar surface area and preference for planarity (see section **5.5**) increased the preference for an excited state dimer.

*5.4.3: Blue-shifted nonemissive absorbance of dye **1***

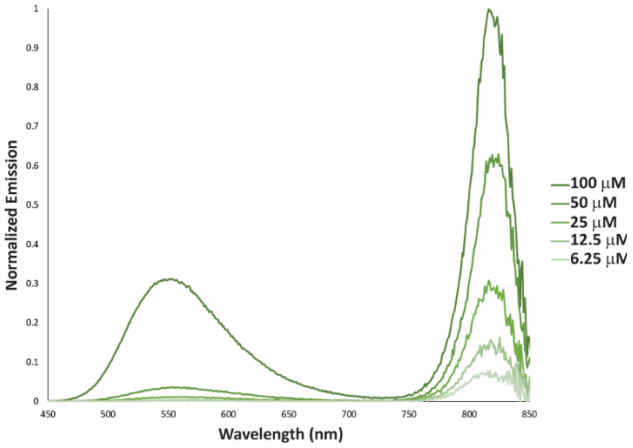


Figure 4: Emission of **6** upon 1:2 dilution in acetonitrile. The cyanine emission band (ca. 800 nm) decreases regularly, while the short-wavelength emission (ca. 500 nm) decreases much faster, suggesting that it arises from an excited state dimer.

Dye **1** featured an anomalous absorbance peak at 585 nm. An excitation analysis of its emission at the normal Cy5 wavelength (~660 nm) shows that a normal cyanine absorbance band (~630 nm) is the one that excites the fluorescence (Figure 5a). This is not a general feature of the

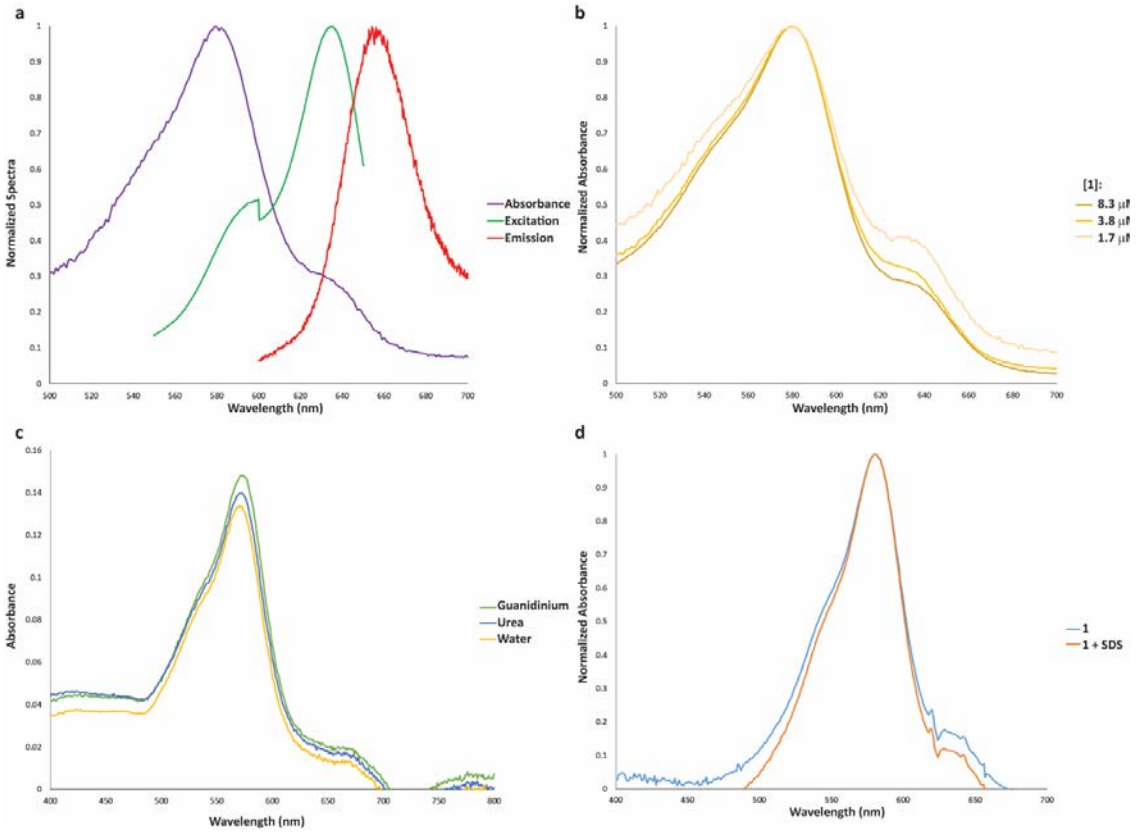


Figure 5: (a) Absorbance, Excitation, fluorescence spectra of **1**. (b) Dilution analysis, (c) addition of denaturants, and (d) addition of SDS to determine the photophysical underpinnings of the blue-shifted nonemissive absorbance band of **1.**

acetylene-phenol moiety, and appears to be unique to the Cy5 variant. Dye **4** has the same acetylene-aromatic group yet does not possess a blue-shifted nonemissive absorbance band. Further complicating matters is that unlike with Cy7 dyes, Cy5 dyes do not have a resonance form that will yield a neutral species upon deprotonation. Therefore, it is unlikely that the blue-shifted absorbance band is an alternate chromophore.

The most likely explanation of this phenomenon is aggregation. Cyanine dyes can aggregate in either of two conformations: H- or J-aggregates. While J-aggregates are characterized by an emissive, slightly red-shifted absorbance band, H-aggregates are blue shifted and nonemissive.[32, 34] Normally, H-aggregation of Cy5 dyes do not occur in organic solvents such as acetonitrile nor at such low concentrations. To test if aggregation is a

possibility, several experiments were conducted. First, **1** was diluted in acetonitrile to undetectable concentrations. Even at very low concentrations (0.02 μM) the proportion of blue-shifted to normal cyanine band did not disappear (Figure 5a).

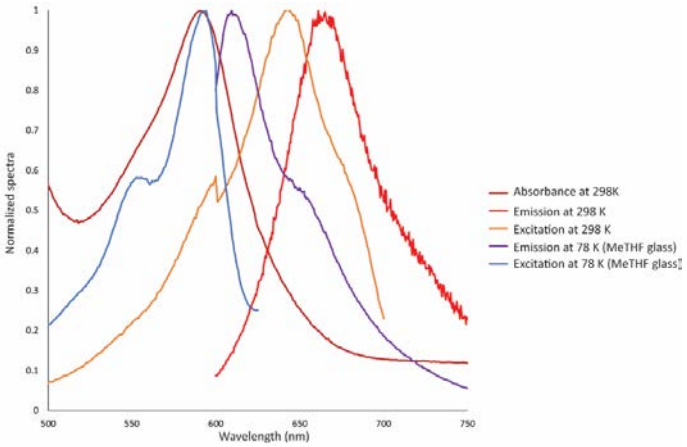


Figure 6: Emission and excitation Spectra of **1** in a methyl-tetrahydrofuan (MeTHF) glass at 78 K overlaid with room temperature spectra in MeTHF

Further, the fluorescence of the compound decreased at the same rate as absorbance, which would not be expected if aggregation was occurring. Addition of denaturants such as guanidinium ions or SDS did not change the shape of the absorbance band either. (FIgure 5c-d)

Interestingly, while the solution phase absorbances in a number of solvents at room temperature for **1** all showed the blue-shifted nonemissive band, when frozen in a methyl-THF glass, emission increased, and the color shifted from purple (containing the blue-shifted absorbance) to blue (cyanine absorbance alone) (Figure 6). While the exact mechanism of the anomalous absorbance of **1** is unknown, it is likely not aggregation.

5.5: Computational analysis of alkyne-aromatic cyanine dyes

Density functional theory (DFT) calculations were performed to seek factors that influence the dyes' relatively efficient, large Stokes shift emission. The ground and excited state rotation barriers of the synthesized dyes were measured using the CAM-B3LYP functional with a 6-31G(d,p) basis set, with time-dependent DFT used for the excited state

calculation.[35] This methodology was chosen because the addition of the CAM (Coulomb-attenuating method) to the common hybrid HF B3LYP functional is common in the assessment of red and NIR-absorbing chromophores.[36-39] This is because the CAM correction will build in a function for long-range Coulombic forces, which are important for a large chromophore, especially with a time-dependent application. Some computational methodologies can easily be aligned to experimental data. A good example are the cation-π interaction calculations detailed in Chapter 2, where gas phase computational results can easily be compared to gas phase empirical results. There is no equivalent one-to-one comparison for these calculations. Therefore, a highly validated and precedented method was chosen. In addition to DFT and TDDFT, these calculations were also performed using HF, which exhibited similar trends to the ground state DFT calculation, though had higher energies in general.

The rotation barriers or wells for the synthesized cyanine dyes are compiled in Table 2. The ability of CAM-B3LYP to accurately calculate rotation barriers was corroborated by also calculating the ground state, gas-phase rotation barrier of N,N-

Table 2: Rotation barriers of the aromatic group relative to the cyanine dye of dyes **1**-**9**. Ground states calculated at DFT//CAM-B3LYP/6-31G(d,p), and excited states using a time-dependent variant.

Dye	S_0 Rotation Barrier[1]	S_1 Rotation Barrier[1]	φ_{em}[2]
1	0.6	-1.8	0.0055
2	0.54	-2.0	0.069
3	0.58	-2.3	0.06
4	-0.9	-1.1	0.09
5	-0.97	-1.3	0.03
6	-0.49	-1.1	0.049
7	16.4	11.9	-
8	13.5	9.7	0.0036
9	11.0	9.7	0.0036

[1]Rotation barrier is the energy of the planar conformation relative to the perpendicular conformation. In kcal/mol
[2]Φ_{em} refers to the short wavelength, large Stokes shift emission.

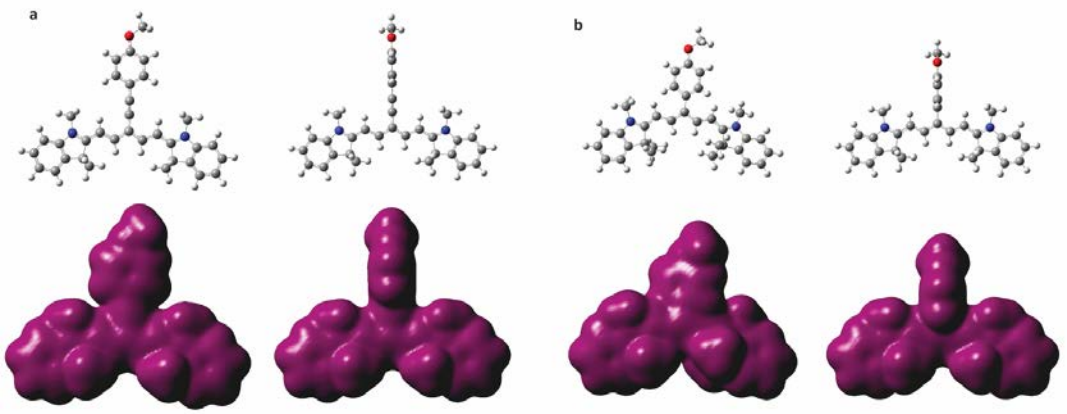
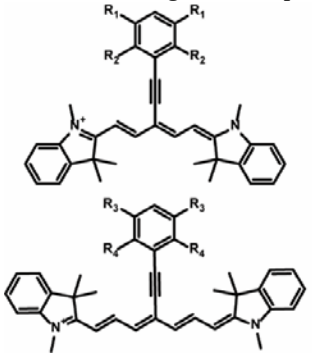


Figure 7: DFT//CAM-B3LYP/6-31G(d,p) calculated structures of (a) compound **2** and (b) compound **8** in parallel (left) and perpendicular (right) conformations. Below: electron density maps of compounds pictured in (a) and (b). Parallel conformation of **5** shows steric clash with methyl groups.

dimethylformamide. The rotation barrier calculated by this method was 18.05 kcal/mol. This is in good agreement with the empirical gas phase rotation barrier calculated using ^{13}C NMR, 19.4 kcal/mol.[40] In addition to the synthesized dyes, several other examples were investigated computationally (Table 3, Scheme 3). The other studied cyanine dyes were the unsubstituted phenyl aromatic as well as two dimethyl aromatics substituted to both Cy5 and Cy7 frameworks, to test the potential increase in rotation barrier caused by the addition of substituents *ortho* to the alkyne moiety. This study predicts that while the *ortho*-dimethyl compound does have a slightly higher rotation barrier, it will still have a much

Scheme 3, Table 3: Rotation barriers of the aromatic group relative to the cyanine dye of dyes **18-23** (right), described in the accompanying scheme (left). Ground states calculated at DFT//CAM-B3LYP/6-31G(d,p), and excited states using a time-dependent variant.



18: $R_1 = H$, $R_2 = H$
19: $R_1 = Me$, $R_2 = H$
20: $R_1 = H$, $R_2 = Me$
21: $R_3 = H$, $R_4 = H$
22: $R_3 = Me$, $R_4 = H$
23: $R_3 = H$, $R_4 = Me$

	S_0 Rotation Barrier[1]	S_1 Rotation Barrier[1]
18	0.5	-0.92
19	0.6	-0.93
20	0.01	-1.1
21	-0.35	-0.95
22	-0.24	-0.82
23	1.68	-0.24

[1]Rotation barrier is the energy of the planar conformation relative to the perpendicular conformation.

lower rotation barrier than compounds **7-9**, underscoring the importance of the acetylene group. The acetylene containing dyes exhibited much lower rotation barriers than **7-9** (Figure 7)**.**

In particular, the planar and perpendicular aromatic-cyanine structures were investigated further to determine the mechanism of the large Stokes shift emission (Figure 8). All the compounds investigated showed similar trends as observed in Figure 8: the higher-lying unoccupied orbitals that mix with the LUMO of the aromatic substituent are comparatively lower in energy than the same orbitals when unmixed (perpendicular). These data are detailed in Table 4. This energy lowering indicates orbital mixing between the LUMO of the aromatic and a higher lying LUMO of the cyanine dye. Because of the orbital topology of Cy5 versus Cy7 dyes, the identity of the higher lying orbital is different, but the energy stabilization is still present. The origins of these orbitals are described in orbital mixing diagrams at the end of this chapter (Figures 10-11)

5.6: Discussion

5.6.1: Mechanism of long Stokes shift fluorescence

Cyanine dyes are a widely studied class of organic dye in part because of their modular nature. Altering the size of the polymethine chain or the end groups predictably changes the excitation and emission wavelengths of the dye, and because of the facile synthesis of such dyes, changes can be made and investigated very easily.[10] In addition, independent alteration of end groups or the polymethine chain can introduce bioorthogonal reactivity and add utility for many applications. However, a significant limitation of all traditional cyanine dyes in fluorescence microscopy is their small Stokes shift. We have synthesized dyes **1-6** that feature, in addition to normal cyanine photophysics, an additional

high-energy excitation band to stimulate cyanine fluorescence with a large (>250 nm) Stokes shift. One strategy to increase the Stokes shift is to incorporate a second chromophore that, upon excitation, undergoes energy transfer to the Cy dye, which then emits. This requires a matching of the donor and acceptor properties.[6, 13, 19-20, 27, 41] We describe here an alternative, potentially general, strategy for creating large Stokes shift Cy dyes.

By connecting a second chromophore to the Cy dye via an acetylene linker, reasonably efficient emission from S1 – the Cy chromophore – can be seen on excitation of S2 – the linked chromophore. This efficiency can be represented numerically using a brightness measurement (Table 1, Figure 3). These brightness values are on the order of 300-1000. While they are significantly smaller than the extremely bright cyanine dyes, these dyes feature brightness values that are comparable to reported fluorescent proteins and commercial dyes. One example is a fluorescent protein (NirFP) that absorbs at 605 nm and emits at 670. It has a brightness of 470:[42] less than some of the synthesized dyes' S2 brightness, with a significantly smaller Stokes shift. A particularly relevant example is sandercyanin, a fluorescent protein which absorbs blue light (375 nm) and emits red light (635 nm). It has a brightness of 336; roughly the same as the S2 excitation from some of the synthesized dyes.[43-44]

Emission from S1 on excitation to S2 is, of course, expected, but it is typically not very efficient for cyanine dyes. For example, the dyes without the acetylene linker – **7-9** – show an order of magnitude drop in emission quantum yield for excitation to S2. One can imagine this is because the aromatic chromophore cannot achieve coplanarity with the polymethine chain. We have performed a number of calculations to support our

interpretation. The acetylene linked dyes **1-6** have rotation barriers < 1 kcal/mol in the ground state. In the S1 state, the barrier rises slightly, and the fully planar form is always preferred (Table 2). Without the acetylene, the twisted form is preferred over the planar by a large margin.

Looking at the relevant MOs (Figure 8), we see that in the LUMO is focused exclusively on the Cy unit in both **1** and **2**, regardless of whether the aromatic is coplanar

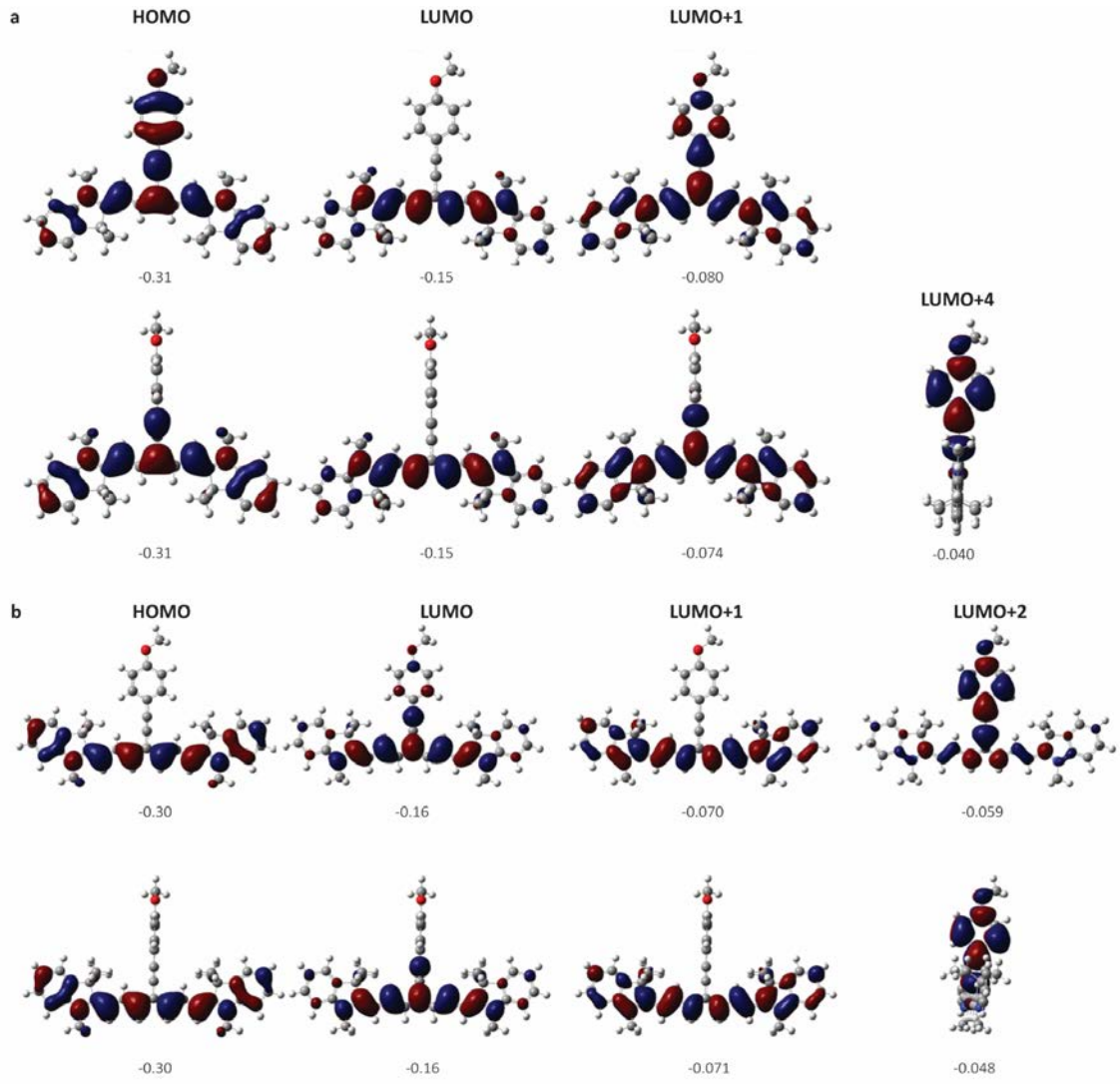


Figure 8: Kohn-Sham orbitals representing the HOMO, LUMO, LUMO+1, and higher lying orbital for (a) **2** and (b) **5** calculated in parallel and perpendicular conformations. Below each structure is the orbital energy in hartrees calculated at DFT//CAM-B3LYP/6-31G(d,p).

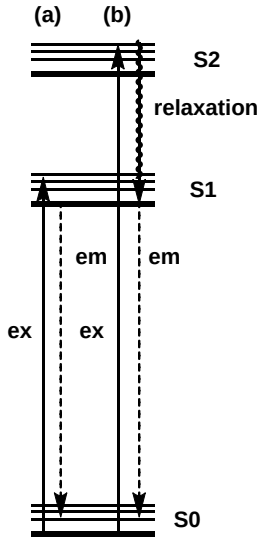
or 90° twisted. In the planar form, there is a low-lying empty MO (LUMO+1 for **1** and LUMO+2 for **2**) with large density on the aromatic, and if these MOs contribute significantly to S2, it could facilitate the efficient formation of S1. In contrast, when the aromatic is perpendicular to the Cy, only a very high energy empty MO has significant density on the aromatic.

Time-dependent DFT calculations support this analysis, in that S0-S2 transitions have increased oscillator strengths when the dye is in the planar conformation relative to perpendicular or to an unsubstituted cyanine dye (Table 3). In addition, while the transition dipole moment of the S0-S1 transition for cyanine dyes is strongly polarized along the polymethine chain (the x-axis in Table 4), the S0-S2 transition dipole moment is polarized along the orthogonal direction for dyes in the planar conformation to a greater magnitude than for the perpendicular conformation (the y axis). We conclude that by inserting an acetylene linker, we have allowed the second chromophore to couple strongly to the Cy fragment, and this increases both the extinction coefficient for S0-S2 absorption and the

Table 4: Transition dipole moment μ magnitudes and oscillator strength f for model acetylene-aromatic cyanine dyes. Dipole moments in the x direction refer to the polymethine chain, while the y direction is the acetylene group. Calculations performed at DFT//CAM-B3LYP/6-31G(d,p).

Dye	Dihedral Angle	$\mu_x^{(S0-S1)}$	$\mu_y^{(S0-S1)}$	$\mu_x^{(S0-S2)}$	$\mu_y^{(S0-S2)}$	$f^{(S0-S1)}$	$f^{(S0-S2)}$
Cy5	N/A	5.57	0.00	0.00	0.34	2.03	0.01
Cy7	N/A	6.62	0.00	0.00	0.38	2.57	0.01
1	0	4.97	0.03	0.01	2.54	1.51	0.63
	90	5.34	0.02	0.01	1.34	1.82	0.18
2	0	4.93	0.15	0.14	2.68	1.48	0.70
	90	5.34	0.03	0.00	1.40	1.82	0.20
3	0	4.76	0.00	0.00	3.19	1.38	0.95
	90	5.34	0.00	0.00	0.00	1.82	0.00
4	0	6.38	0.02	0.04	1.86	2.27	0.25
	90	6.54	0.03	0.00	0.03	2.40	0.00
5	0	6.37	0.06	0.07	2.00	2.27	0.27
	90	6.52	0.25	0.01	0.01	2.23	0.00
6	0	6.31	0.56	0.60	2.10	2.24	0.34
	90	6.53	0.06	0.05	1.10	2.4	0.10

quantum yield for emission from S1 on excitation to S2. Taken together, this creates a very large Stokes shift chromophore with adequate brightness for many applications.



Scheme 4: Proposed mechanism for short wavelength emission.

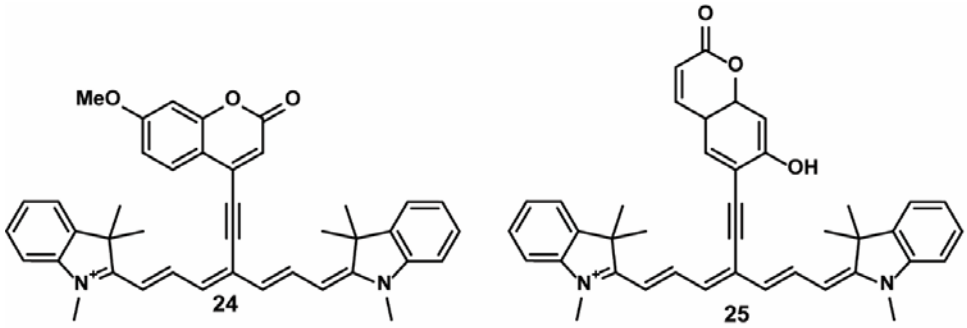
Changing the aromatic moiety from an anisyl to a naphthyl group changed the excitation wavelength as expected. This is a useful feature: the short wavelength excitation wavelength and cyanine emission wavelength are independent of each other. The synthesis of this class of dyes is highly convergent, so modification of the cyanine dye with groups to enable water solubility or biorthogonal attachment would not affect the performance of the Sonagashira reaction to complete dye synthesis. Because the normal photophysical properties of the dyes are virtually unchanged, they could feature dual use. With their long-wavelength readout but large Stokes shift, the dyes described in this work can be used with multiple long-wavelength dyes and can be differentiated in multi-dye systems via their excitation wavelengths, providing a potential new tool for multicolor fluorescence microscopy.

In a related example, a very large Stokes shift has been seen in a structure that contains a perylene linked to a BOPHY via an acetylene linker.[22] Excitation into the perylene absorption band leads to BOPHY emission through an energy transfer mechanism due to the “perfect spectral overlap between the emission of perylene and the absorption of the BOPHY framework”. In this system, large Stokes shifts were not seen in systems where the perylene was replaced by styryl derivatives, illustrating the necessity for energy matching between the two chromophores to achieve a large Stokes shift. A comparable

mechanism is clearly not operative here, as the simple anisyl and methoxynaphthyl groups of **1-6** do not display the necessary spectral overlap. While the perylene-BOPHY system shows a higher quantum yield for large Stokes shift emission, the present system benefits from the modular and general design, that could lead to broader applications.

5.6.2: Future applications: a coumarin-Cy hybrid

While molecules that have featured large Stokes shifts have been reported, this mechanism has to the best of our knowledge not been observed previously. While the brightness of the cyanine dye is very high, the large Stokes shift fluorescence brightness is



	24		25	
	0	90	0	90
HOMO:	-0.30	-0.30	-0.31	-0.31
LUMO:	-0.25	-0.22	-0.16	-0.16
LUMO+1:	-0.22	-0.20	-0.084	-0.080
Rotation Barrier (S1)		+1.16 kcal/mol	-1.82 kcal/mol	
Rotatoin Barrier (S2)		-0.26 kcal/mol	-2.50 kcal/mol	

Dye	Dihedral Angle	$\mu_x^{(S0-S1)}$	$\mu_y^{(S0-S1)}$	$\mu_x^{(S0-S2)}$	$\mu_y^{(S0-S2)}$	$f^{(S0-S1)}$	$f^{(S0-S2)}$
24	0	6.23	0.39	0.32	0.98	2.12	0.08
	90	6.53	0.06	0.05	1.1	2.4	0.1
25	0	6.36	0.33	0.34	2.01	2.24	0.3
	90	6.54	0.00	0.00	1.04	2.40	0.09

Figure 9: Top: chemical structure and orbital energies for parallel and perpendicular conformations of **24** and **25**. The LUMO+1 is the orbital which contains the coumarin/cyanine orbital mixing. Rotation barrier is the energy of the planar conformation relative to the perpendicular conformation. Bottom: Transition dipole moment μ magnitudes and oscillator strength f for proposed acetylene-coumarin cyanine dyes. Dipole moments in the x direction refer to the polymethine chain, while the y direction is the acetylene group. Calculations performed at DFT//CAM-B3LYP/6-31G(d,p).

relatively small. Increasing the brightness of a fluorescent dye can be accomplished in two ways: increasing the quantum yield of fluorescence or increasing the absorptivity of the transition. Unfortunately, the quantum yield of the S2-mediated fluorescence is likely bounded by the S1 cyanine fluorescence. Because the mechanism necessarily passes through the S1 state (Scheme 4), all normal transitions from that state must occur. Therefore, an increase in the absorptivity of the S2 state is a preferred strategy to produce an increase in brightness.

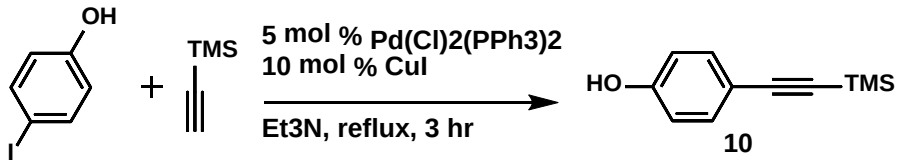
The absorptivity of the S2 fluorescence (Table 1) appears to be roughly on the same order of magnitude as the aromatic substituent in the acetylene-aromatic group. Therefore, an increase in the aromatic's absorptivity may increase the large Stokes shift fluorescence. We conducted preliminary computational studies on two such dyes, a 4-coumarin attached acetylene-Cy7 dye and a 7-attached acetylene Cyanine dye (**24** and **25** respectively). As expected, **25** featured a low rotation barrier, and **24** was slightly larger, though still not as high as **7-9**. TDDFT analysis on the planar form of all the dyes showed similar trends to what was observed with **1-6**.

5.7: Methods

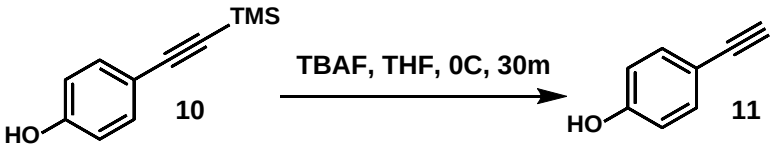
Organic Synthesis:

Unless otherwise stated, reactions were carried out under argon. Commercially available reagents were obtained from Sigma Aldrich, AK Scientific, Alfa Aesar, or Acros Organics and used without further purification. Solvents were used as received. Thin-layer chromatography with Sigma Aldrich silica gel coated plates with fluorescent indicator (0.25 mm) was used to monitor reactions. Silica gel chromatography was conducted as described by Still et al.[45], with silica gel purchased from Alfa Aesar (60 Å, 230-400 mesh).

NMR spectra were recorded on Varian (300, 400, 500, or 600 MHz) spectrometers. HRMS (ESI) were obtained with an Agilent 6200 Series TOF. Isolated yields are reported except for compounds purified by HPLC, where crude recovery yields are reported and only a portion of the crude was purified by HPLC.

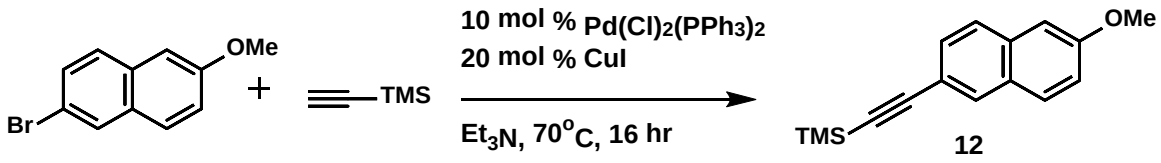


To a flame-dried 25 mL three-necked flask, 1.02 g (4.63 mmol) 4-iodophenol, 27.1 mg (0.142 mmol) copper (I) iodide, and 97.2 mg $Pd(Cl)_2(PPh_3)_2$ (0.138 mmol) were added. These were dissolved in NEt_3 (15 mL) and ethynyltrimethylsilane (0.95 mL, 13.9 mmol) was added dropwise. The reaction was heated to 80^0C and stirred for 3 hours. The reaction was then filtered, and concentrated. The brown residue was purified by silica gel column chromatography with a mobile phase of EtAc:Hexane 40:60. **10** was recovered as a light brown powder. ^{1}H NMR (300 MHz, Chloroform-d) δ 7.49 – 7.33 (m, 2H), 6.93 – 6.62 (m, 2H), 5.35 – 4.68 (br s, 1H), 0.35 – 0.14 (s, 9H). Yield: 87%. Characterization matched previously-synthesized compound.[46]

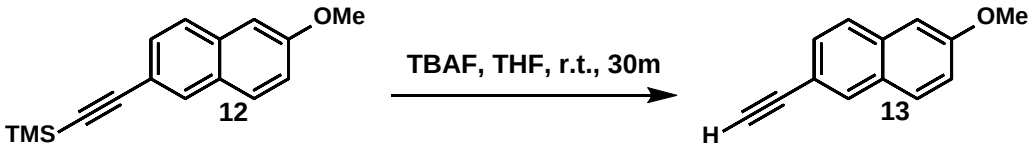


Add dry THF to a dry flask charged with **10** (0.7658 g, 4.02 mmol). Cool on ice and add 6.02 mL of 1.0 M tetrabutylammonium fluoride solution in THF. Let the reaction stir on ice for 30 minutes. Quench with aqueous ammonium chloride. Extract with ethyl acetate three times. Then collect the organic layers, dry on magnesium sulfate, concentrate, and dry under vacuum. Product is unstable on silica so was not purified further. **11** was recovered as a crude brown oil. Crude ^{1}H NMR (300 MHz, Chloroform-d) δ 7.68 – 7.63

(m, 2H), 7.31 – 7.25 (m, 2H), 3.29 (s, 1H). Crude yield: 95%. Matched previously-synthesized compound.[46]

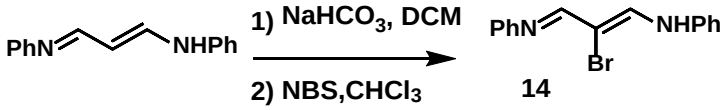


To a flame-dried 50 mL three-necked flask, 1.0 g (4.2 mmol) 6-bromo-2-methoxy-naphthalene, 24.3 mg (0.78 mmol) copper (I) iodide, and 90 mg (0.39 mmol) $Pd(Cl)_2(PPh_3)_2$ were added. These were dissolved in NEt_3 (25 mL), and ethynyltrimethylsilane (1.05 mL, 7.4 mmol) was added dropwise. The reaction was heated to 80^0C and stirred for 3 hours. The reaction was then filtered, and concentrated. The brown residue was purified by silica gel column chromatography with a mobile phase of EtAc:Hexane 40:60. **12** is a brown solid. 1H NMR (300 MHz, Chloroform-d) δ 7.96 – 7.90 (m, 1H), 7.66 (t, J = 8.7 Hz, 2H), 7.47 (d, J = 8.5, 1.6 Hz, 1H), 7.17 – 7.05 (m, 2H), 3.92 (s, 3H), 0.28 (d, J = 0.8 Hz, 9H). m/z=254.4, yield: 46%. Matched a previously-synthesized compound.[47]

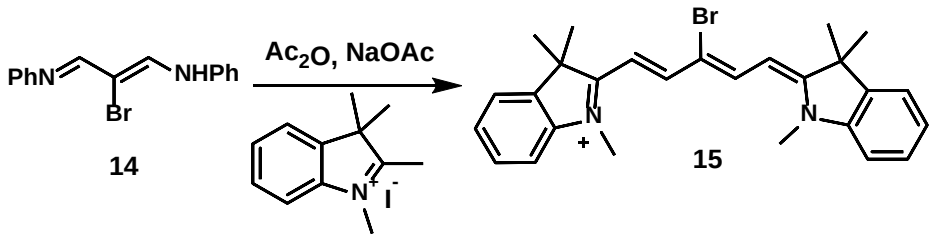


Add 20 mL dry THF to a dry flask charged with **12** (0.4966 g, 1.95 mmol). Add 2.04 mL of 1.0 M tetrabutylammonium fluoride solution in THF. Let the reaction stir on ice for 30 minutes. Quench with 1:1 aqueous ammonium chloride:water. Extract with ethyl acetate three times. Then collect the organic layers, dry on magnesium sulfate, concentrate, and dry under vacuum. **13** is unstable on silica so was not purified further. Crude ^{1}H NMR (300 MHz, Acetonitrile-d3) δ 7.98 (d, J = 1.4 Hz, 1H), 7.77 (s, 1H), 7.49 (dd, J = 8.5, 1.7

Hz, 1H), 7.26 (d, J = 2.5 Hz, 1H), 7.19 (dd, J = 9.0, 2.6 Hz, 1H), 3.91 (s, 4H), 3.44 (s, 1H). m/z = 182.2. Crude yield: 90%. Matches previous synthesis.[47]

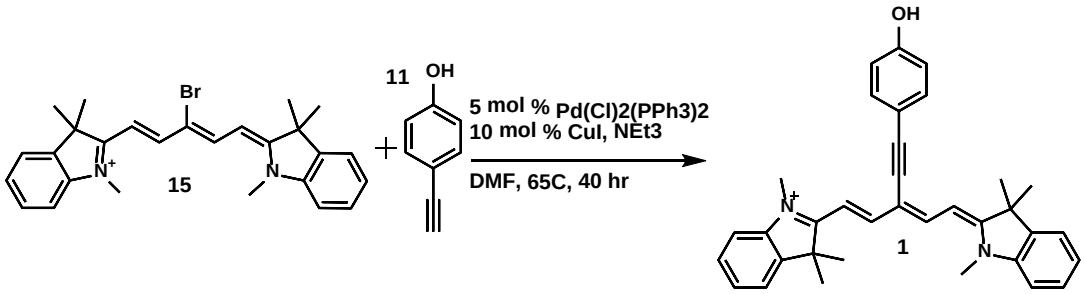


0.8 g (3.6 mmol) of malondialdehyde (bis)phenylimine HCl was freebased with sodium bicarbonate in dichloromethane. Organic layer was dried with magnesium sulfate, concentrated, and dried under vacuum to obtain the freebased malondialdehyde (0.671 g, 3.02 mmol). Chloroform was passed through silica gel into a flame-tried flask with molecular sieves. Added 1.3 g (7.24 mmol) of purified N-brom osuccinimide on ice. Reaction was brought to room temperature and stirred for two hours. The reaction was quenched with saturated aqueous sodium bicarbonate. This was extracted three times with chloroform. The organic layer was washed with water and brine. The organic layers were collected, dried under magnesium sulfate, concentrated, and dried under vacuum. The residue was purified by silica gel column chromatography with a mobile phase of EtAc/Hexanes 10:90. **14** is a bright yellow/orange oil. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.02 (s, 2H), 7.42 – 7.28 (m, 4H), 7.18 – 7.06 (m, 6H). Yield: 74%. m/z = 300.3. Matches a previous synthesis.[48]



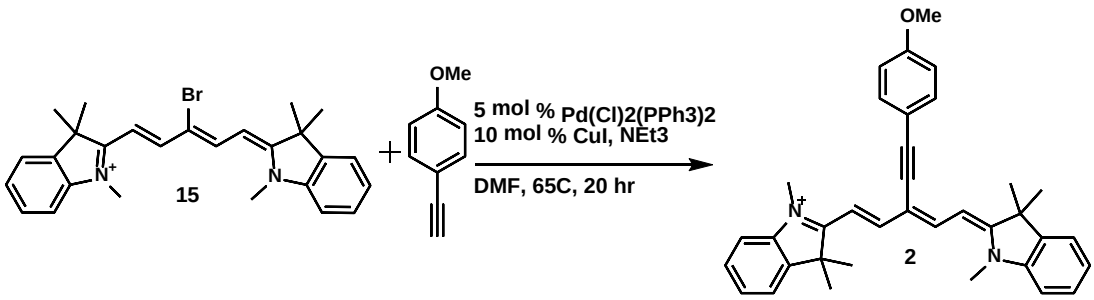
1.76 g (5.84 mmol) 1,2,2,3-tetramethylindolenium and **14** (0.80 g, 2.65 mmol) were dissolved in acetic anhydride in a flame-dried 10mL round bottom flask. To this mixture, 0.852 mg (10.6 mmol) sodium acetate was added. The reaction was stirred at room

temperature overnight. After completion, add brine and quench with dichloromethane three times. The organic fractions were collected, dried with magnesium sulfate, and dried under vacuum. The crude oil was purified by silica gel chromatography. The mobile phase is MeOH:DCM 2:98. **15** is a deep blue solid that when pure, gives a red/purple sheen. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.93 (d, J = 13.3 Hz, 2H), 7.56 – 7.48 (m, 5H), 7.38 (ddd, J = 7.4, 6.4, 1.4 Hz, 4H), 7.35 – 7.28 (m, 5H), 7.17 – 7.11 (m, 2H), 7.09 (d, J = 7.4 Hz, 2H), 6.34 (d, J = 13.3 Hz, 2H), 3.78 (s, 6H), 1.87 (s, 12H). m/z = 462.4. Yield: 97%. Characterization matches previous work.[48]



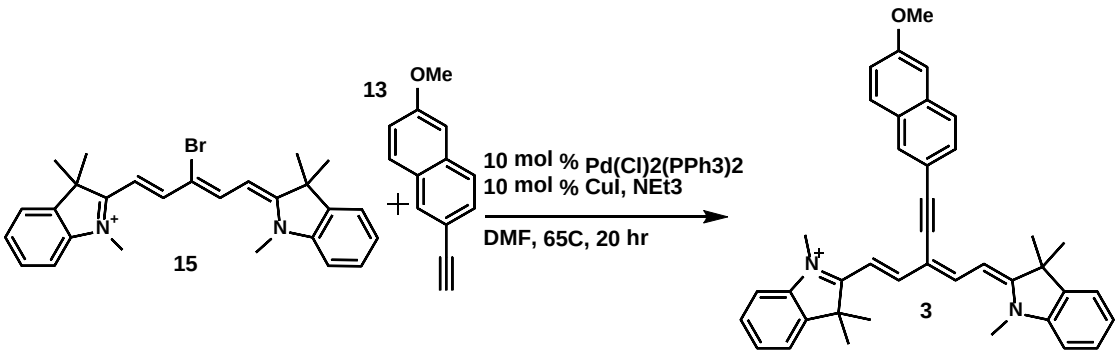
0.2 M stock solutions in dry DMF were made for the compounds **15** and **11**. 9.4 mg (0.013 mol) of $PdCl_2(PPh_3)_2$ and 5.1 mg (0.0253 mmol) of copper (I) iodide were added to a flame-dried 25 mL three necked flask. Triethylamine (0.5 mL) was then added. After the solid reagents are in suspension, 1.606 mL of **15** was added. Once the reagents are fully dissolved, 1.071 mL of **11** was added dropwise. The reaction was heated to 65^0C for 40 hours, monitored by analytical liquid chromatography-mass spectroscopy. After completion, the reaction was quenched with brine and extracted with DCM three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue/purple product was purified twice on silica gel chromatography with a mobile phase of MeOH:AcOH:DCM 9:1:90. The blue and purple fractions were then crystallized in a DCM/Hexanes mixture at 20^0C. **1** is a purple solid.

^{1}H NMR (600 MHz, DMSO-d6) δ 9.48 (s, 2H), 8.33 (d, J = 15.1 Hz, 2H), 7.74 (d, J = 7.0 Hz, 2H), 7.64 (d, J = 7.4 Hz, 2H), 7.45 (dt, J = 16.4, 7.9 Hz, 9H), 7.31 (t, J = 7.2 Hz, 2H), 6.96 (d, J = 8.0 Hz, 4H), 6.73 (d, J = 8.2 Hz, 4H), 3.02 (s, 6H), 1.70 (s, 12H). m/z = 499.4. Yield: 12%.

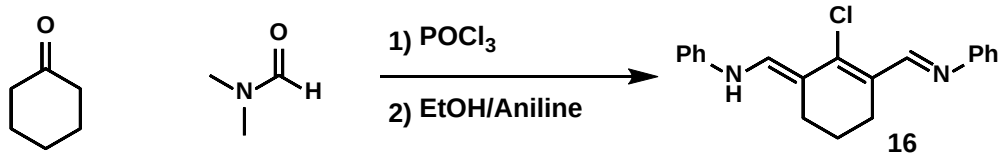


A 0.2 M stock solution in dry DMF were made for compound **15**. 68.2 mg (0.097 mol) of $PdCl_2(PPh_3)_2$ and 37.0 mg (0.194 mmol) of copper (I) iodide were added to a flame-dried 25 mL three necked flask. Triethylamine (0.5 mL) was then added. After the solid reagents are in suspension, 9.72 mL (1.94 mmol) of **15** was added. Once the reagents are fully dissolved, 0.26 mL (1.94 mmol) of 4-ethynylanisole was added dropwise. The reaction was heated to 65^{0}C for 20 hours, monitored by analytical liquid chromatography-mass spectroscopy. After completion, the reaction was quenched with brine and extracted with DCM three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of MeOH:DCM 5:95 followed by HPLC purification to yield **2** as a blue solid. ^{1}H NMR (400 MHz, Chloroform-d) δ 8.76 (d, J = 14.0 Hz, 2H), 7.54 – 7.48 (m, 2H), 7.39 (t, J = 8.2 Hz, 4H), 7.26 (d, J = 15.1 Hz, 1H), 7.17 (d, J = 8.0 Hz, 2H), 6.95 (d, J = 8.2 Hz, 2H), 6.54 (d, J = 14.0 Hz, 2H), 3.87 (d, J = 1.3 Hz, 3H), 3.74 (s, 6H), 1.80 (s, 12H). ^{13}C NMR (401 MHz, Chloroform-d) δ 175.32 (d, J = 29.7 Hz), 156.18 , 142.66 , 141.54 , 140.66 , 132.81 , 128.38 , 125.53 , 122.56 , 115.23 , 110.45

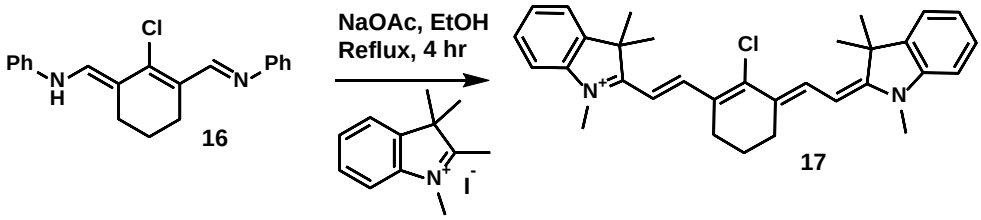
, 105.57 (d, J = 33.7 Hz), 102.13 , 81.70 , 55.39 , 50.13 , 27.66 , 24.36 . m/z = 513.7. Yield: 8.8%.



15 was dissolved in DMF and added to a flame dried 25 mL three-necked flask charged with 126 mg (0.180 mmol) of $PdCl_2(PPh_3)_2$ and 34.0 mg (0.18 mmol) of copper (I) iodide. Triethylamine (0.5 mL) was then added. After the solid reagents are in suspension, 16 mL (0.16 mmol) of **15** was added. Once the reagents are fully dissolved, 2.6 mL (2.6 mmol) of a 1.0 M solution of **13** was added dropwise. The reaction was heated to 65^0C for 20 hours, and monitored by analytical liquid chromatography-mass spectroscopy. After completion, the reaction was quenched with brine and extracted with DCM three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of MeOH:DCM 10:90 followed by HPLC purification to yield **3** as a blue-green solid. ^{1}H NMR (400 MHz, Chloroform-d) δ 9.06 (d, J = 14.0 Hz, 2H), 8.00 (s, 1H), 7.76 (dd, J = 8.7, 2.8 Hz, 2H), 7.58 (dd, J = 8.5, 1.6 Hz, 1H), 7.43 – 7.38 (m, 3H), 7.36 (s, 1H), 7.21 (dd, J = 8.9, 2.5 Hz, 1H), 7.15 (d, J = 8.4 Hz, 3H), 7.00 (s, 0H), 6.59 (d, J = 14.0 Hz, 2H), 3.96 (s, 3H), 3.77 (s, 6H), 1.86 (s, 12H). m/z=563.8. Yield: 10%.

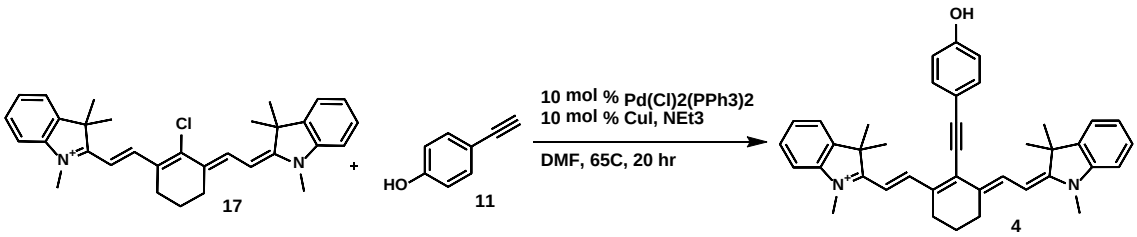


Phosphorus oxychloride (11 mL, 0.12 mol) was added dropwise to anhydrous DMF (13 mL, 0.17 mol) at 0 °C. After 1 h, cyclohexanone (5.5 mL, 0.053 mol) was added, and the mixture was refluxed for 1 h. After cooling to rt, an aniline/EtOH [1:1 (v/v), 18 mL] mixture was added dropwise, and the reaction was continued at rt for 30 min. The mixture was poured into ice cold H2O/HCl (10:1,110 mL), and the resulting residue was filtered, washed with cold H_2O and THF, and then dried *in vacuo* to obtain **16**. ^{1}H NMR (300 MHz, Chloroform-d) δ 11.41 (d, J = 14.6 Hz, 2H), 8.48 (d, J = 14.5 Hz, 2H), 7.71 – 7.63 (dz, 4H), 7.39 (t, J = 7.9 Hz, 4H), 7.25 – 7.17 (m, 2H), 3.02 (t, J = 6.1 Hz, 4H), 1.95 (q, 3H). Deep purple solid, m/z=322.8. Yield: 46%. Characterization matches previous synthesis.[49]

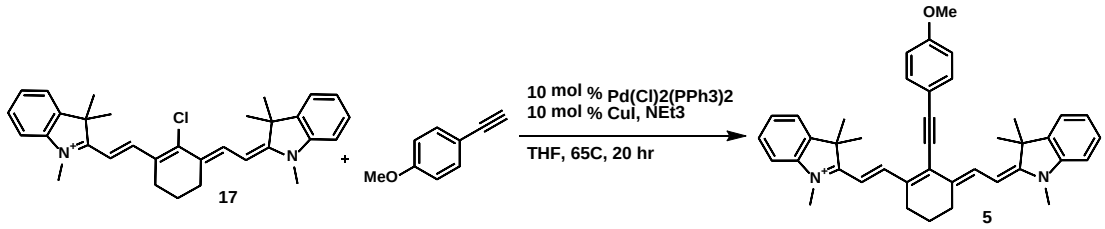


50 mg (0.14 mmol) **16** and (100.5 mg, 0.28 mmol) 1,2,2,3-tetramethylindolenium were dissolved in 4 mL ethanol in a flame-dried 10mL three-necked flask. To this mixture, 40 mg (0.56 mmol) sodium acetate was added. The reaction was stirred at reflux for four hours. After completion, 10 mL brine was added to quench and the residual ethanol was removed with a rotary evaporator. The aqueous mixture was extracted with dichloromethane three times. The organic fractions were collected, dried with magnesium sulfate, and dried under vacuum. The crude oil was purified by silica gel chromatography. The mobile phase is MeOH:DCM 10:90. **17** was obtained as a dark green solid, with a gold sheen when very pure. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.32 (d, J = 14.1 Hz, 2H),

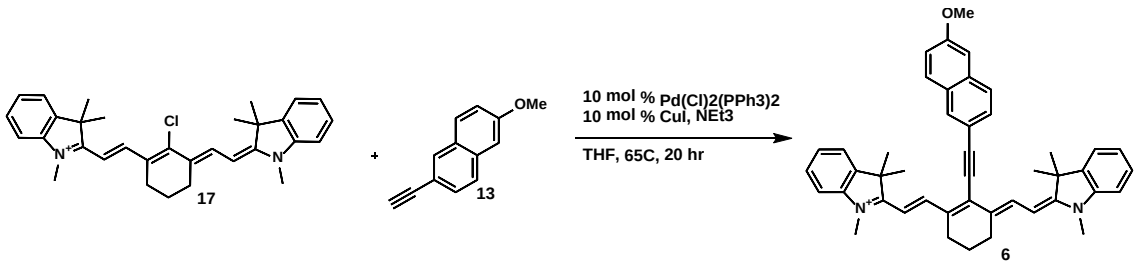
7.41 – 7.33 (m, 4H), 7.26 – 7.15 (m, 4H), 6.19 (d, J = 14.1 Hz, 2H), 3.74 (d, J = 0.5 Hz, 6H), 2.73 (t, J = 6.2 Hz, 4H), 2.02 – 1.83 (m, 2H), 1.70 (s, 12H). ^{13}C NMR (101 MHz, Chloroform-d) δ 172.80 , 150.57 , 144.30 , 142.75 , 140.87 , 128.81 , 127.76 , 125.31 , 122.11 , 110.80 , 101.69 , 49.21 , 32.65 (d, J = 2.4 Hz), 28.07 , 26.77 , 20.66 , -0.02 . m/z: 484.1. Yield: 92%.



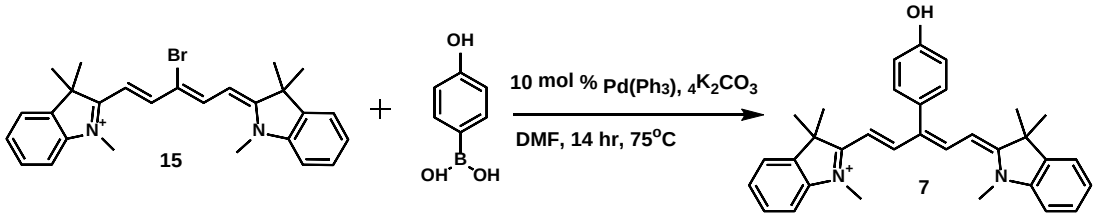
A 0.125 M stock solution in dry DMF were made for compound **17** and 0.125 M solution was made for **11**. 43.9 mg (0.062 mmol) of $PdCl_2(PPh_3)_2$ and 23.8 mg (0.125 mmol) of copper (I) iodide were added to a flame-dried 25 mL three necked flask. Triethylamine (0.5 mL) was then added. After the solid reagents are in suspension, 10 mL (1.25 mmol) of **17** was added. Once the reagents are fully dissolved, 15 mL (1.88 mmol) of **11** was added dropwise. The reaction was heated to 65^0C for 16 hours, monitored by analytical liquid chromatography-mass spectroscopy. After completion, the reaction was quenched with brine and extracted with DCM three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of MeOH:DCM 10:90 followed by HPLC purification to yield **4** as a green solid. ^{1}H NMR (500 MHz, Chloroform-d) δ 8.56 (d, J = 14.0 Hz, 2H), 7.49 (d, J = 6.3 Hz, 0H), 7.44 (dd, J = 8.3, 2.8 Hz, 3H), 7.39 (td, J = 7.7, 1.2 Hz, 2H), 7.35 (d, J = 8.6 Hz, 3H), 7.26 – 7.23 (m, 2H), 7.10 (d, J = 7.8 Hz, 3H), 6.05 (d, J = 14.0 Hz, 2H), 3.77 (s, 1H), 3.61 (s, 5H), 2.77 (d, J = 6.1 Hz, 2H), 2.62 (t, J = 6.2 Hz, 4H), 1.75 (s, 12H). m/z=565.8. Yield: 7.6%.



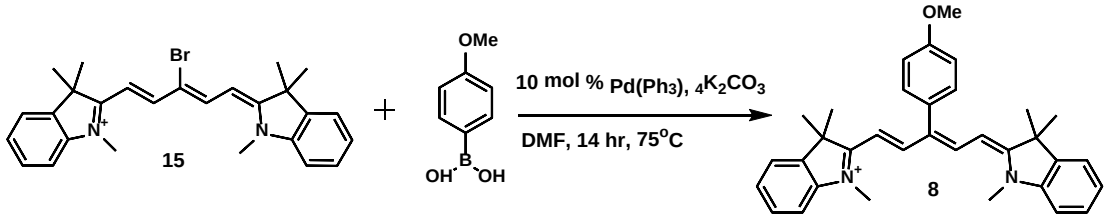
A 0.1 M stock solution in dry THF was made for compound **17** (3 mL total). 10.5 mg (0.015 mmol) of $PdCl_2(PPh_3)_2$ and 5.7 mg (0.03 mmol) of copper (I) iodide were added to a flame-dried 10 mL three necked flask. Triethylamine (0.5 mL) was then added. After the solid reagents are in suspension, add 3 mL of the **17** solution. Once the reagents are fully dissolved, 40 μL (0.45 mmol) of 4-ethynylanisole was added. The reaction was heated to 65^0C for 16 hours, and monitored by analytical liquid chromatography-mass spectroscopy. After completion, the reaction was quenched with brine and extracted with DCM three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of MeOH:DCM 10:90 followed by HPLC purification to yield **5** as yellow/green solid. ^{1}H NMR (500 MHz, Chloroform-d) δ 8.56 (d, J = 14.0 Hz, 2H), 7.49 (d, J = 6.3 Hz, 0H), 7.44 (dd, J = 8.3, 2.8 Hz, 3H), 7.39 (td, J = 7.7, 1.2 Hz, 2H), 7.35 (d, J = 8.6 Hz, 3H), 7.26 – 7.23 (m, 2H), 7.10 (d, J = 7.8 Hz, 3H), 6.05 (d, J = 14.0 Hz, 2H), 3.77 (s, 1H), 3.61 (s, 6H), 2.77 (d, J = 6.1 Hz, 2H), 2.62 (t, J = 6.2 Hz, 4H), 1.75 (s, 12H). m/z=579.8



A 0.1 M stock solution in dry DMF was made for compound **17** and 1.0 M solution was made for **13**. 40 mg (0.056 mmol) of $PdCl_2(PPh_3)_2$ and 20 mg (0.105 mmol) of copper (I) iodide were added to a flame-dried 25 mL three necked flask. Triethylamine (0.5 mL) was then added. After the solid reagents are in suspension, 6 mL (0.6 mmol) of **17** was added. Once the reagents are fully dissolved, 15 mL (0.15 mmol) of **13** was added dropwise. The reaction was heated to 65^0C for 16 hours, monitored by analytical liquid chromatography-mass spectroscopy. After completion, the reaction was quenched with brine and extracted with DCM three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of MeOH:DCM 10:90 followed by HPLC purification to yield **6** as a green solid. ^{1}H NMR (400 MHz, Chloroform-d) δ 8.52 (d, J = 14.1 Hz, 2H), 8.09 (s, 1H), 8.00 (d, J = 14.0 Hz, 1H), 7.85 (d, J = 8.5 Hz, 1H), 7.79 (d, J = 8.9 Hz, 1H), 7.64 – 7.58 (m, 1H), 7.38 (q, J = 7.3, 6.8 Hz, 3H), 7.35 – 7.27 (m, 4H), 7.21 (d, J = 6.3 Hz, 3H), 7.18 (d, J = 7.5 Hz, 3H), 7.14 – 7.05 (m, 2H), 6.27 (d, J = 14.1 Hz, 2H), 6.11 (d, J = 14.0 Hz, 1H), 3.73 (s, 6H), 3.62 (s, 3H), 3.01 (d, J = 0.7 Hz, 2H), 2.70 (d, J = 6.3 Hz, 5H), 1.74 (s, 12H). m/z=615.8.

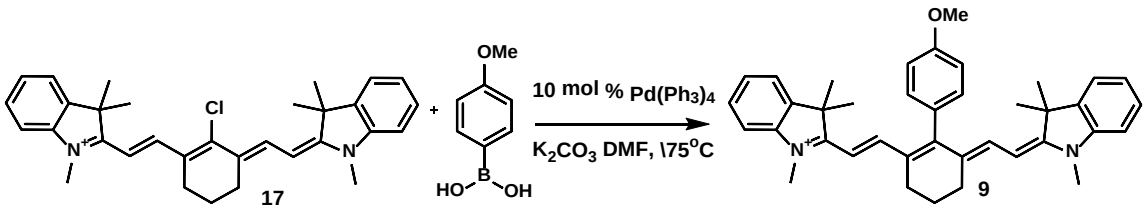


Dissolve 250 mg (0.49 mmol) **15** in 5 mL DMF and add 335 mg (2.43 mmol) K_2CO_3 and 80 mg (0.58 mmol) 4-hydroxyphenylboronic acid. Keep under argon, and purge/evacuate a few times. Weigh out 60 mg (0.052 mmol) palladium triphenylphosphine, and purge/evacuate in a vial. Add to the reaction flask and heat to 110°C overnight. Monitor using TLC and LC/MS. After completion, the reaction was quenched with brine and extracted with ethyl ether three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of MeOH:DCM 5:95 followed by HPLC chromatography. **7** was obtained as a blue/green solid. ^{1}H NMR (300 MHz, Chloroform-d) δ 7.86 (d, J = 14.0 Hz, 2H), 7.60 (s, 1H), 7.37 (d, J = 9.4 Hz, 6H), 7.32 (d, J = 2.6 Hz, 2H), 7.19 (d, J = 2.8 Hz, 5H), 7.07 (d, J = 7.6 Hz, 2H), 6.96 (d, J = 8.2 Hz, 4H), 6.91 (d, J = 0.6 Hz, 2H), 5.88 (d, J = 13.8 Hz, 2H), 3.64 (d, J = 2.0 Hz, 2H), 3.36 (s, 6H), 2.10 (d, J = 6.9 Hz, 2H), 1.75 (s, 12H) m/z=475.2. Yield: 4.4%.



Dissolve 250 mg (0.49 mmol) **15** in 5 mL DMF and add 335 mg (2.43 mmol) K_2CO_3 and 80 mg (0.58 mmol) 4-methoxyphenylboronic acid. Keep under argon, and purge/evacuate a few times. Weigh out 60 mg (0.052 mmol) palladium

triphenylphosphine, and purge/evacuate in a vial. Add to the reaction flask and heat to 110°C overnight. Monitor using TLC and LC/MS. After completion, the reaction was quenched with brine and extracted with ethyl ether three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of MeOH:DCM 5:95 followed by HPLC chromatography. **8** was obtained as a blue solid. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.74 (d, J = 13.9 Hz, 2H), 7.73 (dt, J = 7.7, 4.1 Hz, 2H), 7.65 – 7.47 (m, 2H), 7.33 (dd, J = 8.9, 7.4 Hz, 6H), 7.23 – 7.18 (m, 2H), 7.16 (s, 2H), 7.13 – 7.03 (m, 4H), 7.06 (d, J = 8.7 Hz, 2H), 6.48 (t, J = 8.1 Hz, 2H), 5.62 (d, J = 13.9 Hz, 2H), 3.92 (s, 3H), 3.49 (s, 6H), 1.78 (s, 12H). ^{13}C NMR (401 MHz, Chloroform-d) δ 173.91 , 154.85 , 148.36 , 142.67 , 141.03 , 131.15 , 128.43 , 125.23 , 122.34 , 114.59 , 110.27 , 101.21 , 55.33 , 49.57 , 27.72 , 24.36 , 20.81 . m/z=489.7. Yield: 6.5%.



Dissolve 250 mg (0.49 mmol) **17** in 5 mL DMF and add 335 mg (2.43 mmol) K_2CO_3 and 80 mg (0.58 mmol) 4-methoxyphenylboronic acid. Keep under argon, and purge/evacuate a few times. Weigh out 60 mg (0.052 mmol) palladium triphenylphosphine, and purge/evacuate in a vial. Add to the reaction flask and heat to 110°C overnight. Monitor using TLC and LC/MS. After completion, the reaction was quenched with brine and extracted with ethyl ether three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of

MeOH:DCM 5:95 followed by HPLC chromatography. **9** was obtained as a blue solid. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.75 (d, J = 14.1 Hz, 2H), 7.34 (td, J = 7.7, 1.3 Hz, 2H), 7.31 – 7.27 (m, 2H), 7.20 (dd, J = 7.5, 0.9 Hz, 2H), 7.09 (d, J = 7.9 Hz, 2H), 7.07 – 6.95 (m, 3H), 6.08 (d, J = 14.1 Hz, 2H), 3.61 (s, 6H), 2.68 (t, J = 6.1 Hz, 4H), 1.95 (t, J = 6.1 Hz, 2H), 1.50 (s, 12H). m/z=555.8. Yield: 8.2%.

Quantum Mechanical Calculations:

For all calculations, the Gaussian 09 software pack was used.[50] Visualization was performed using Gaussview 05.[51] Rotation barriers were calculated by setting dihedral angles frozen using modified redundant internal coordinates from a structure optimized using DFT/CAM-B3LYP/6-31G(d,p) and performing the restricted geometry optimization in HF/6-31G(d,p). Ground and excited-state energies were calculated using DFT//CAM-B3LYP/6-31G(d,p) and TDDFT//CAM-B3LYP/6-31G(d,p) respectively. Orbitals were visualized in Gaussview 05 from the DFT structures.

Photophysics:

All spectra were taken in Omnisolv purity spectroscopy-grade solvents and Millipore pure water. Absorbance spectra were taken using an Agilent™ Technologies Cary 60 UV-Vis with a pulsed xenon lamp as the excitation source. The error in reported wavelength values is at most 0.5 nm. Fluorescence and excitation spectra were recorded on a Jobin Yvon Spex Fluorolog-3-11. Sample excitation was achieved via a xenon arc lamp with a monochromator providing wavelength selection. Right angle light emission was sorted using a monochromator and fed into a Hamamatsu R928P photomultiplier tube with photon counting. Short and long pass filters were used where appropriate.

All emission and excitation spectra were taken as an average of two spectra subtracted from two blank spectra of the same excitation wavelength taken on the same day. Quantum yields of fluorescence were measured using Cy5 (φ_f=0.27) and *meso*-Chloro-Cy7 (φ_f=0.12) as references in acetonitrile. In addition to subtracting out background, anomalous features due to excitation and emission overtones were subtracted out of the spectra.

Spectra in methyl-tetrahydrofuran glass were taken using specialized glassware to take spectra using the same fluorimeter setup under liquid nitrogen. Samples were dissolved in methyl-tetrahydrofuran and flash-frozen in an NMR tube, and then immediately placed into a cavity containing $N_{2(l)}$ and emission and excitation spectra taken in the normal fashion.

5.8: Orbital mixing of cyanine/acetylene-aromatic chromophores

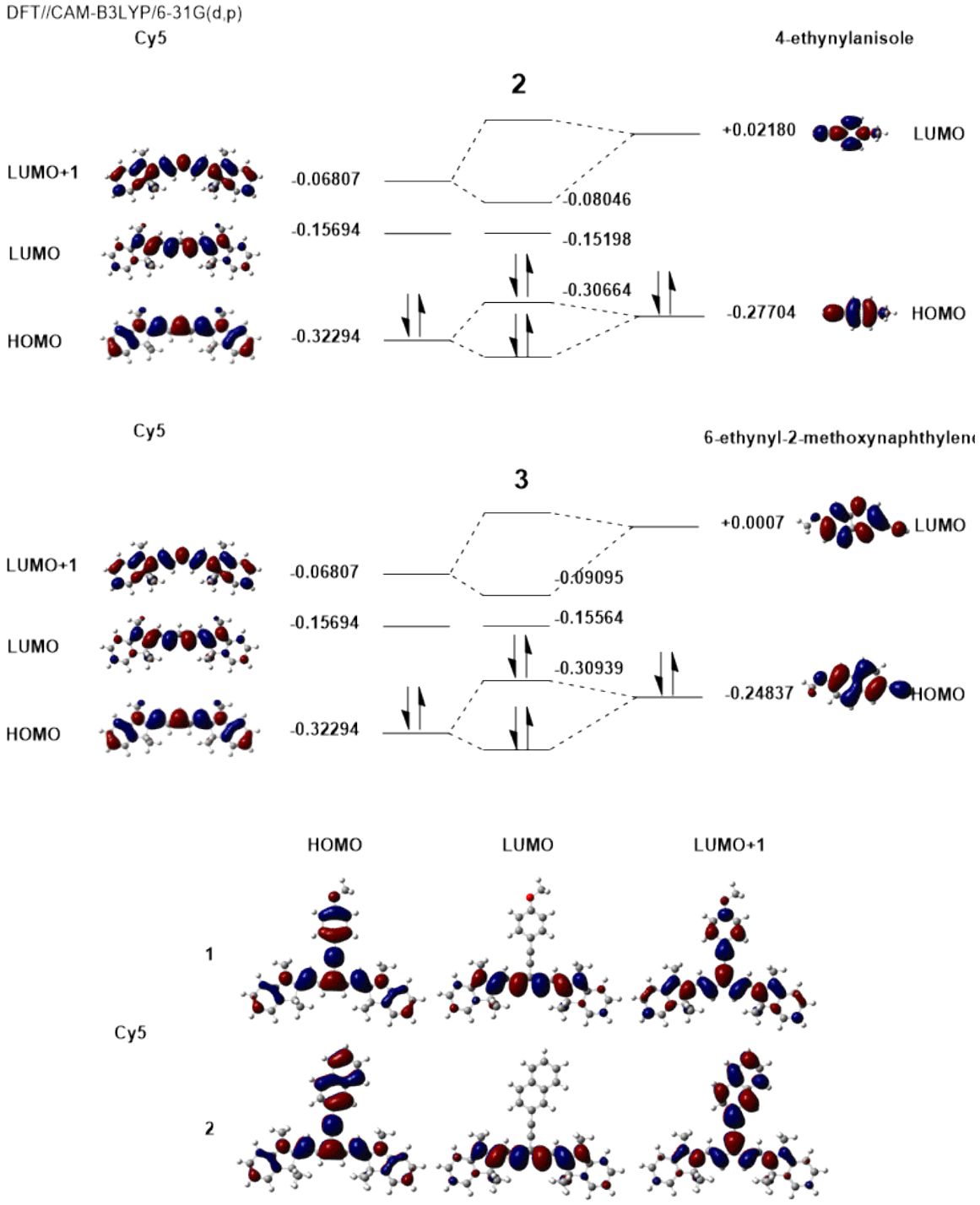


Figure 10: Orbital mixing of compounds **2** (top) and **3** (middle). Energies of all orbitals are in hartrees, calculated with DFT//CAM-B3LYP/6-31G(d,p). In each case, mixing of the LUMO+1 of Cy5 and the LUMO of the aromatic group results in the formation of the hybrid orbital, which results in S2 fluorescence (LUMO+1, bottom).

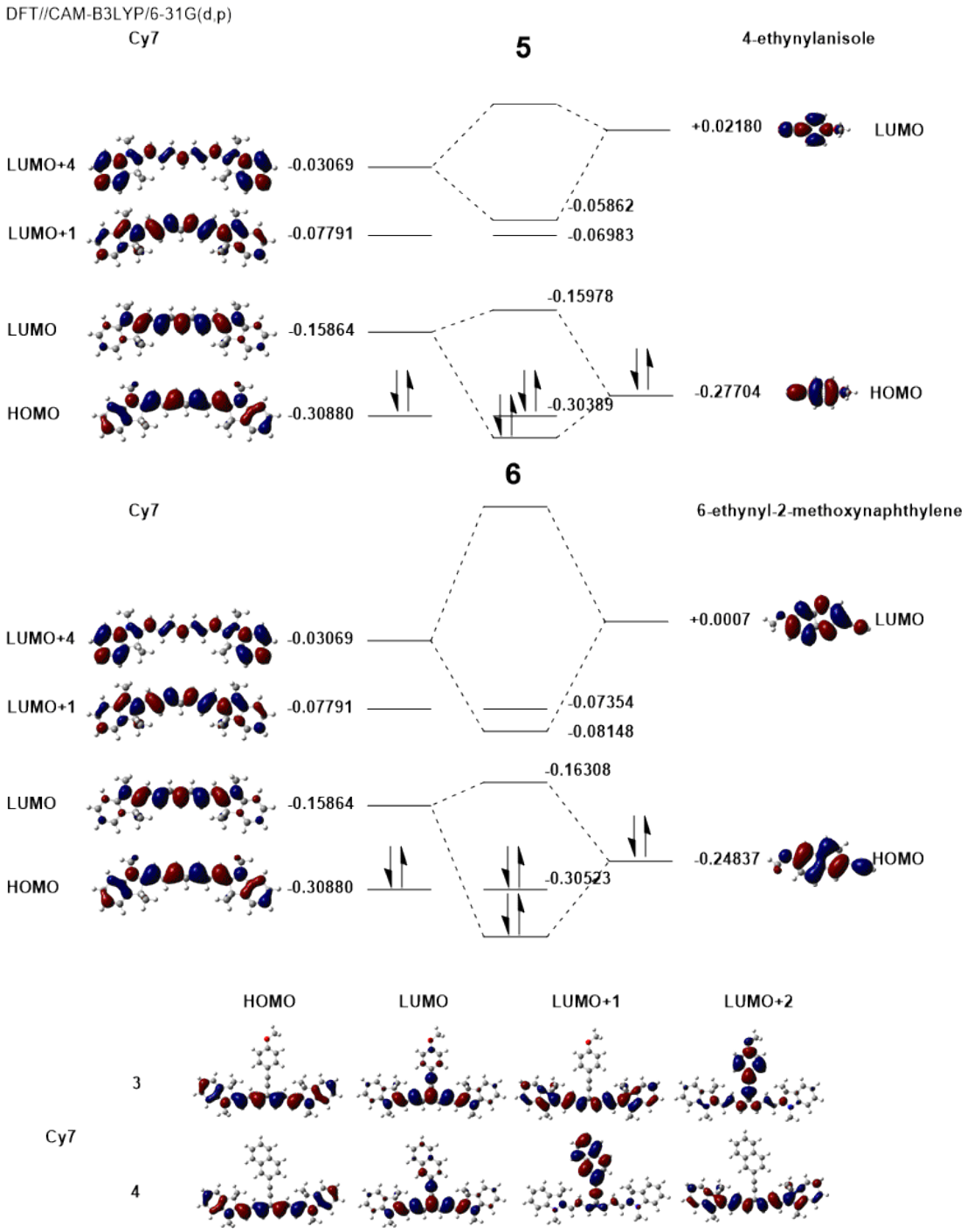


Figure 11: Orbital mixing of compounds **5** (top) and **6** (middle). Energies of all orbitals are in hartrees, calculated with DFT//CAM-B3LYP/6-31G(d,p). In each case, mixing of the LUMO+4 of Cy7 and the LUMO of the aromatic group results in the formation of the hybrid orbital which results in S2 fluorescence. Because of the relative energies of the phenol LUMO versus the naphthol LUMO, the hybrid orbital is either the LUMO+2 or LUMO+1 of **6** or **5** respectively. In any event, their energies are nearly identical.

5.9: References

1. Heilemann, M., Fluorescence microscopy beyond the diffraction limit. *J. Biotechnol.* **2010,** *149* (4), 243-251.
2. Leung, B. O.; Chou, K. C., Review of Super-Resolution Fluorescence Microscopy for Biology. *Appl. Spectrosc.* **2011,** *65* (9), 967-980.
3. Sarder, P.; Maji, D.; Achilefu, S., Molecular Probes for Fluorescence Lifetime Imaging. *Bioconjugate Chemistry* **2015,** *26* (6), 963-974.
4. Berlier, J. E.; Rothe, A.; Buller, G.; Bradford, J.; Gray, D. R.; Filanoski, B. J.; Telford, W. G.; Yue, S.; Liu, J.; Cheung, C. Y.; Chang, W.; Hirsch, J. D.; Beechem Rosaria P. Haugland, J. M.; Haugland, R. P., Quantitative Comparison of Long-wavelength Alexa Fluor Dyes to Cy Dyes: Fluorescence of the Dyes and Their Bioconjugates. *Journal of Histochemistry & Cytochemistry* **2003,** *51* (12), 1699-1712.
5. Liu, X.; Xu, Z.; Cole, J. M., Molecular Design of UV–vis Absorption and Emission Properties in Organic Fluorophores: Toward Larger Bathochromic Shifts, Enhanced Molar Extinction Coefficients, and Greater Stokes Shifts. *The Journal of Physical Chemistry C* **2013,** *117* (32), 16584-16595.
6. Richard, J.-A.; Massonneau, M.; Renard, P.-Y.; Romieu, A., 7-Hydroxycoumarin−Hemicyanine Hybrids: A New Class of Far-Red Emitting Fluorogenic Dyes. *Organic Letters* **2008,** *10* (19), 4175-4178.
7. Vogelsang, J.; Steinhauer, C.; Forthmann, C.; Stein, I. H.; Person-Skegro, B.; Cordes, T.; Tinnefeld, P., Make them Blink: Probes for Super-Resolution Microscopy. *ChemPhysChem* **2010,** *11* (12), 2475-2490.
8. Zhang, J.; Chen, R.; Zhu, Z.; Adachi, C.; Zhang, X.; Lee, C.-S., Highly Stable Near-Infrared Fluorescent Organic Nanoparticles with a Large Stokes Shift for Noninvasive Long-Term Cellular Imaging. *ACS Applied Materials & Interfaces* **2015,** *7* (47), 26266-26274.
9. Renikuntla, B. R.; Rose, H. C.; Eldo, J.; Waggoner, A. S.; Armitage, B. A., Improved Photostability and Fluorescence Properties through Polyfluorination of a Cyanine Dye. *Organic Letters* **2004,** *6* (6), 909-912.
10. Sun, W.; Guo, S.; Hu, C.; Fan, J.; Peng, X., Recent Development of Chemosensors Based on Cyanine Platforms. *Chemical Reviews* **2016,** *116* (14), 7768-7817.
11. Beppu, T.; Tomiguchi, K.; Masuhara, A.; Pu, Y.-J.; Katagiri, H., Single Benzene Green Fluorophore: Solid-State Emissive, Water-Soluble, and Solvent- and pH-Independent

Fluorescence with Large Stokes Shifts. *Angewandte Chemie International Edition* **2015,** *54* (25), 7332-7335.

12. Chen, S.; Hou, P.; Song, X., A red-emitting fluorescent probe for imaging hydrogen sulphide with a large Stokes shift. *Sensors and Actuators B: Chemical* **2015,** *221*, 951-955.
13. Jin, J.-L.; Li, H.-B.; Geng, Y.; Wu, Y.; Duan, Y.-A.; Su, Z.-M., Theoretical Insight into the Origin of Large Stokes Shift and Photophysical Properties of Anilido-Pyridine Boron Difluoride Dyes. *ChemPhysChem* **2012,** *13* (16), 3714-3722.
14. Liu, X.; Yang, L.; Gao, L.; Chen, W.; Qi, F.; Song, X., A phthalimide-based fluorescent probe for thiophenol detection in water samples and living cells with a large Stokes shift. *Tetrahedron* **2015,** *71* (43), 8285-8289.
15. Percivalle, C.; Sissi, C.; Greco, M. L.; Musetti, C.; Mariani, A.; Artese, A.; Costa, G.; Perrore, M. L.; Alcaro, S.; Freccero, M., Aryl ethynyl anthraquinones: a useful platform for targeting telomeric G-quadruplex structures. *Organic & Biomolecular Chemistry* **2014,** *12* (22), 3744-3754.
16. Qiao, Y.; Chen, J.; Yi, X.; Duan, W.; Gao, B.; Wu, Y., Highly fluorescent perylene dyes with large stokes shifts: synthesis, photophysical properties, and live cell imaging. *Tetrahedron Letters* **2015,** *56* (21), 2749-2753.
17. Zhou, Y.; Xiao, Y.; Chi, S.; Qian, X., Isomeric Boron−Fluorine Complexes with Donor−Acceptor Architecture: Strong Solid/Liquid Fluorescence and Large Stokes Shift. *Organic Letters* **2008,** *10* (4), 633-636.
18. Agmon, N., Elementary Steps in Excited-State Proton Transfer†. *The Journal of Physical Chemistry A* **2004,** *109* (1), 13-35.
19. Kim, T. G.; Topp, M. R., Ultrafast Excited-State Deprotonation and Electron Transfer in Hydroxyquinoline Derivatives. *The Journal of Physical Chemistry A* **2004,** *108* (46), 10060-10065.
20. Spry, D. B.; Goun, A.; Fayer, M. D., Deprotonation Dynamics and Stokes Shift of Pyranine (HPTS). *The Journal of Physical Chemistry A* **2006,** *111* (2), 230-237.
21. Tolbert, L. M.; Solntsev, K. M., Excited-State Proton Transfer: From Constrained Systems to "Super" Photoacids to Superfast Proton Transfer†. *Accounts of Chemical Research* **2001,** *35* (1), 19-27.
22. Huaulmé, Q.; Mirloup, A.; Retailleau, P.; Ziessel, R., Synthesis of Highly Functionalized BOPHY Chromophores Displaying Large Stokes Shifts. *Organic Letters* **2015,** *17* (9), 2246-2249.

23. Guarin, C. A.; Villabona-Monsalve, J. P.; Lopez-Arteaga, R.; Peon, J., Dynamics of the Higher Lying Excited States of Cyanine Dyes. An Ultrafast Fluorescence Study. *J. Phys. Chem. B* **2013,** *117* (24), 7352-7362.
24. Maus, M.; Rettig, W.; Bonafoux, D.; Lapouyade, R., Photoinduced Intramolecular Charge Transfer in a Series of Differently Twisted Donor−Acceptor Biphenyls As Revealed by Fluorescence. *The Journal of Physical Chemistry A* **1999,** *103* (18), 3388-3401.
25. Tanaka, H.; Shizu, K.; Nakanotani, H.; Adachi, C., Dual Intramolecular Charge-Transfer Fluorescence Derived from a Phenothiazine-Triphenyltriazine Derivative. *The Journal of Physical Chemistry C* **2014,** *118* (29), 15985-15994.
26. Zhang, Q.; Kuwabara, H.; Potscavage, W. J.; Huang, S.; Hatae, Y.; Shibata, T.; Adachi, C., Anthraquinone-Based Intramolecular Charge-Transfer Compounds: Computational Molecular Design, Thermally Activated Delayed Fluorescence, and Highly Efficient Red Electroluminescence. *Journal of the American Chemical Society* **2014,** *136* (52), 18070-18081.
27. Wang, L.; Zhang, Y.; Xiao, Y., meso-Alkoxy BODIPYs with a good balance between larger Stokes shifts and higher fluorescence quantum yields. *RSC Advances* **2013,** *3* (7), 2203-2206.
28. Johannes, H. H.; Grahn, W.; Reisner, A.; Jones, P. G., Ethynylated, Vinylated, and Hetarylated Indodicarbocyanines by Palladium-Catalyzed Cross-Coupling Reactions. *Tetrahedron Letters* **1995,** *36* (40), 7225-7228.
29. Shimizu, Y.; Temma, T.; Hara, I.; Yamahara, R.; Ozeki, E.-i.; Ono, M.; Saji, H., Development of Novel Nanocarrier-Based Near-Infrared Optical Probes for In Vivo Tumor Imaging. *J Fluoresc* **2012,** *22* (2), 719-727.
30. Samaniego Lopez, C.; Lago Huvelle, M. A.; Uhrig, M. L.; Coluccio Leskow, F.; Spagnuolo, C. C., Recognition of saccharides in the NIR region with a novel fluorogenic boronolectin: in vitro and live cell labeling. *Chemical Communications* **2015,** *51* (23), 4895-4898.
31. Prokhorov, V. V.; Perelygina, O. M.; Pozin, S. I.; Mal'tsev, E. I.; Vannikov, A. V., Polymorphism of Two-Dimensional Cyanine Dye J-Aggregates and Its Genesis: Fluorescence Microscopy and Atomic Force Microscopy Study. *The Journal of Physical Chemistry B* **2015**.
32. Chowdhury, A.; Wachsmann-Hogiu, S.; Bangal, P. R.; Raheem, I.; Peteanu, L. A., Characterization of chiral H and J aggregates of cyanine dyes formed by DNA templating using stark and fluorescence spectroscopies. *J. Phys. Chem. B* **2001,** *105* (48), 12196-12201.
33. Khairutdinov, R.; Serpone, N., Photophysics of Cyanine Dyes: Subnanosecond Relaxation Dynamics in Monomers, Dimers, and H- and J-Aggregates in Solution. *J Phys Chem B* **1997,** *101* (14), 9.

34. Ruedas-Rama, M. J.; Orte, A.; Martin-Domingo, M. C.; Castello, F.; Talavera, E. M.; Alvarez-Pez, J. M., Interaction of YOYO-3 with Different DNA Templates to Form H-Aggregates. *The Journal of Physical Chemistry B* **2014**.
35. Yanai, T.; Tew, D. P.; Handy, N. C., A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chemical Physics Letters* **2004,** *393* (1), 51-57.
36. Osman, O. I., DFT Study of the Structure, Reactivity, Natural Bond Orbital and Hyperpolarizability of Thiazole Azo Dyes. *International Journal of Molecular Sciences* **2017,** *18* (2).
37. Rachuta, K.; Bayda, M.; Majchrzak, M.; Koput, J.; Marciniak, B., Unusual emission properties of the selected organosilicon compounds containing a styrylcarbazole chromophore: inversion of the singlet excited states. *Physical Chemistry Chemical Physics* **2017,** *19* (18), 11698-11705.
38. Song, C.; Li, L.; Duan, S.; Luo, Y.; Tian, G. J., Theoretical simulations for vibrationally-resolved absorption spectra of naphthalenediimide cyclophane derivatives. *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2017,** *183*, 339-347.
39. Thorat, K. G.; Sekar, N., Pyrrole-thiazole based push-pull chromophores: An experimental and theoretical approach to structural, spectroscopic and NLO properties of the novel styryl dyes. *Journal of Photochemistry and Photobiology a-Chemistry* **2017,** *333*, 1-17.
40. LeMaster, C. B.; True, N. S., Gas-phase NMR study of torsional barriers in N,N-diethyl- and N,N-diisopropylformamide: medium effects on kinetic parameters. *The Journal of Physical Chemistry* **1989,** *93* (4), 1307-1311.
41. Bertolino, C.; Caputo, G.; Barolo, C.; Viscardi, G.; Coluccia, S., Novel Heptamethine Cyanine Dyes with Large Stokes' Shift for Biological Applications in the Near Infrared. *J Fluoresc* **2006,** *16* (2), 221-225.
42. Shcherbo, D.; Shemiakina, I. I.; Ryabova, A. V.; Luker, K. E.; Schmidt, B. T.; Souslova, E. A.; Gorodnicheva, T. V.; Strukova, L.; Shidlovskiy, K. M.; Britanova, O. V.; Zaraisky, A. G.; Lukyanov, K. A.; Loschenov, V. B.; Luker, G. D.; Chudakov, D. M., Near-infrared fluorescent proteins. *Nat Meth* **2010,** *7* (10), 827-829.
43. Yu, C.-L.; Ferraro, D.; Ramaswamy, S.; Schmitz, M. H.; Schaefer, W. F.; Gibson, D. T., Purification and properties of Sandercyanin, a blue protein secreted in the mucus of blue forms of walleye, Sander vitreus. *Environmental Biology of Fishes* **2008,** *82* (1), 51-58.
44. Ghosh, S.; Yu, C.-L.; Ferraro, D. J.; Sudha, S.; Pal, S. K.; Schaefer, W. F.; Gibson, D. T.; Ramaswamy, S., Blue protein with red fluorescence. *Proceedings of the National Academy of Sciences* **2016,** *113* (41), 11513-11518.

45. Still, W. C.; Kahn, M.; Mitra, A., Rapid chromatographic technique for preparative separations with moderate resolution. *The Journal of Organic Chemistry* **1978,** *43* (14), 2923-2925.
46. Potier, J.; Menuel, S.; Azaroual, N.; Monflier, E.; Hapiot, F., Limits of the Inversion Phenomenon in Triazolyl-Substituted β-Cyclodextrin Dimers. *European Journal of Organic Chemistry* **2014,** *2014* (7), 1547-1556.
47. Hachiya, S.; Asai, K.; Konishi, G.-i., Unique solvent-dependent fluorescence of nitro-group-containing naphthalene derivatives with weak donor–strong acceptor system. *Tetrahedron Letters* **2013,** *54* (14), 1839-1841.
48. Nanjunda, R.; Owens, E. A.; Mickelson, L.; Alyabyev, S.; Kilpatrick, N.; Wang, S.; Henary, M.; Wilson, W. D., Halogenated pentamethine cyanine dyes exhibiting high fidelity for G-quadruplex DNA. *Bioorganic & Medicinal Chemistry* **2012,** *20* (24), 7002-7011.
49. Mohammad, I.; Stanford, C.; Morton, M. D.; Zhu, Q.; Smith, M. B., Structurally modified indocyanine green dyes. Modification of the polyene linker. *Dyes and Pigments* **2013,** *99* (2), 275-283.
50. 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.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
51. Roy Dennington, T. K., and John Millam, Gaussview, Version 5. Semichem, Inc: Shawnee Mission, KS, 2009.

Chapter 6: Progress towards the synthesis and characterization of a NIR photocage

6.1: Abstract

Development of near infrared (NIR) photochemistry has been a long-standing goal of both basic science as well as for therapeutic purpose. A photocage that could release a bioactive molecule upon irradiation with NIR light – within the window of wavelengths that can pass through bone and tissue – would have immediate impact in many different fields of therapeutics. This chapter describes progress towards the development of a NIR photocage, mainly through a cyanine dye framework. These photocages take advantage of the *ortho*-quinone methide photoreaction, where an excited state charge transfer and subsequent proton transfer release a leaving group and generate an *ortho* quinone methide in the excited state. Detailed here are three potential photocages that are discussed theoretically and subsequently synthesized and investigated photophysically. The first proposed photocage is an acetylene-phenol-based photoacidic cyanine dye, where excited state proton transfer would initiate the *ortho* quinone methide generation. The second is a triaryl methane cation ethynologue, and the third takes advantage of PET from a pendant aryl group to Cy7 to initiate *ortho* quinone methide generation. Unfortunately, none of the three mechanisms resulted in a functional photocage. Finally, challenges to NIR photocage production are discussed, focusing on the difficulty with the short excited state lifetime of cyanine dyes.

6.2: Introduction

Spatiotemporal control of drug activation would be a useful tool for the treatment of complex disease targets.[1-2] Activation of drugs at the site of injury or invasive condition would avoid unwelcome side effects. This is particularly relevant in chemotherapeutics. The challenge of directing a drug or prodrug in the body has been addressed in a wide variety of ways. While peptide directing groups can be effective, they can also be, depending on their sequence, rapidly degraded *in vivo*.[3-5] PEGylation of a drug can increase its serum stability and circulation lifetime near a tumor, but requires many extra synthetic steps and complex formulation.[6-9]In each case, these solutions are specific rather than general. A general solution would be a useful addition to the toolkit of drug development. One crucial area of development for spatiotemporal control is treating disease of the brain, specifically traumatic brain injury.

Traumatic brain injury (TBI) is a currently unresolved public health issue. Roughly the same number of Americans suffer concussions and traumatic brain injuries every year as they do for cancer.[10-11] It is the leading cause of death among people less than 40 years old.[12-14] TBI has gained increased visibility recently with well-publicized cases in the military and sports such as the NFL.[15] There are few viable treatments for TBI, due to its quick progression.[16] After the initial injury, which includes brain bruising, follows a secondary injury that occurs minutes to days after the primary injury and involves a complex series of events that can cause significant damage to the brain.[17-18] This includes inflammation, generation of reactive oxygen species, and a cascade of neuronal activity by increased glutamate release referred to as excitotoxicity. These events lead to what is

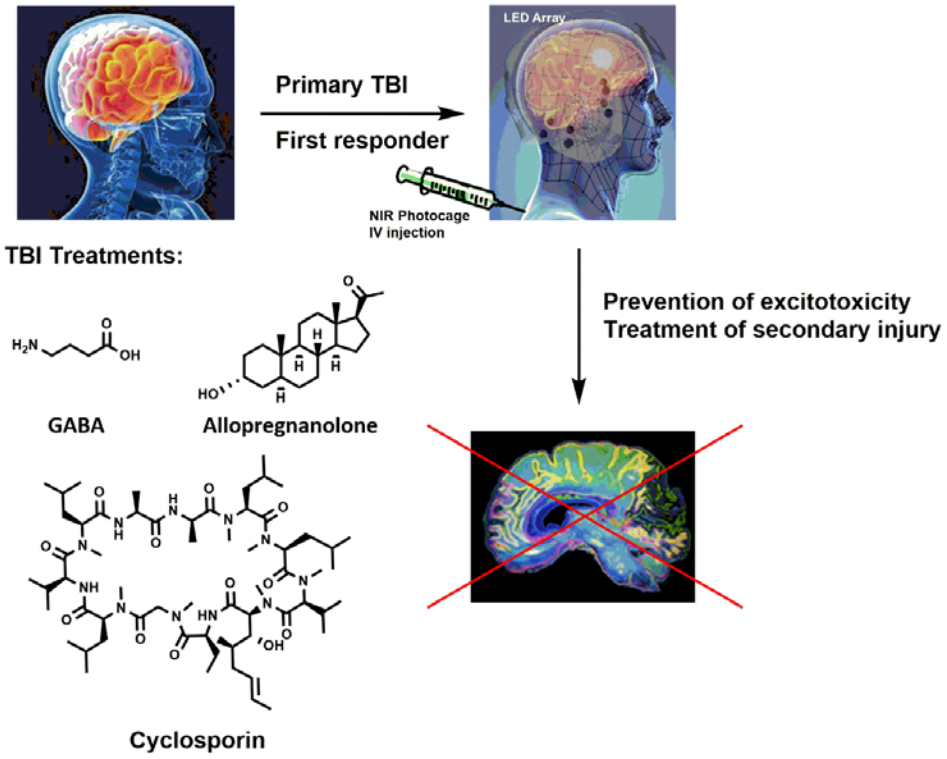
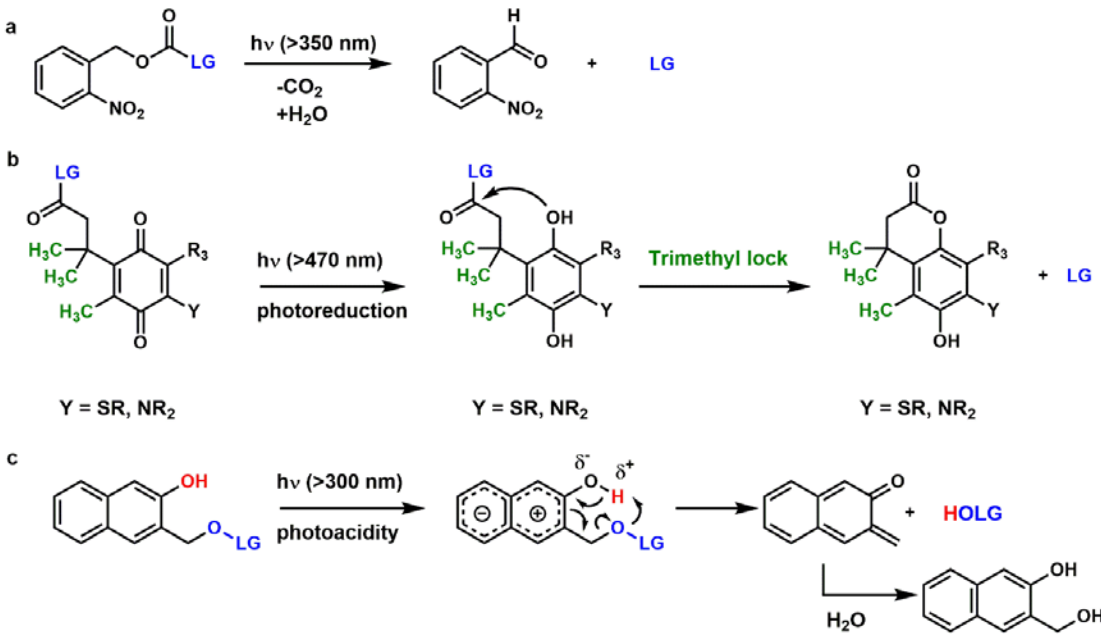


Figure 1: Proposed treatment of traumatic brain injury using a synthesized NIR photocage. After primary injury, a first responder can inject a photocage and irradiate NIR light into the brain. This will prevent excitotoxicity and secondary injury. Pictured are examples of potential TBI treatments. Top two images originally from a CDC report to congress,[12] the second edited. The bottom right image is a CT scan of a TBI patient.[14]

commonly known as "post-concussion syndrome," a common set of symptoms that can persist for months after the injury and affect cognition and awareness.[19-20]

Preventing excitotoxicity through suppression of neuronal activity has been proposed as one mechanism to treat secondary TBI. Where release of glutamate after the secondary injury activates neuroreceptors, application of a drug that will either inhibit neuronal function or activate inhibitory neuroreceptors such as the glycine or GABA receptor will prevent this excitatory activity. However, drugs that inhibit neurotransmitter function have potential problems. For example, cyclosporin, a cyclic peptide natural product, has been proposed as a potential treatment for TBI but has failed five separate



Scheme 1: Photodeprotection strategies. (a) ortho¬-nitrobenzyl protecting groups decage leaving groups or bioactive molecules (LG) with UV light via a radical reaction. (b) One effort has focused on the photoreduction of quinones and subsequent trapping of the unveiled nucleophile with the well-characterized trimethyl lock system. (c) Photochemical generation of an *ortho* quinone methide by excited state charge transfer and proton transfer.

Phase II clinical trials due to adverse side effects.[21] It stands to reason that a drug effective at shutting down neuronal function could pose issues in the peripheral nervous system as well as the central nervous system. Therefore, an effective TBI treatment would treat the secondary injury within the first 24 hours, but with activation restricted to inside the brain (Figure 1). Specific spatiotemporal activation of a drug within the brain is a difficult matter. The brain does not have a distinct chemical signature that could activate drugs, unlike, for example, the reducing anaerobic environment of tumors. Therefore, it is necessary for an external stimulus to activate a drug in situ inside the skull. This necessitates a methodology that can penetrate through the skull and brain tissue.

One potential avenue for affecting drug release *in vivo* through the skull or skin is by photorelease through photoremovable protecting groups.[22-23] Near infrared (NIR) light is among a limited set of wavelengths that can pass through bone and tissue, called the NIR

window. However, current state-of-the-art photochemistry does not extend to the NIR range. The most common photochemical deprotection is the *ortho*-nitrobenzyl group chemistry, which only extends at most to the far-UV (Scheme 1a). One recent example from our lab is the development of visible-absorbing quinone trimethyl lock (QTML) structures. This system features a decoupled photoreduction/decaging event, in which visible-light (~450 nm) triggers a photoreduction of the quinone to a hydroquinone followed by a trimethyl lock closing, releasing a leaving group (Scheme 1b).[24-25] However, these strategies are not yet close to the NIR threshold.

Another strategy is the use of an existing NIR-absorbing framework and adjusting it for use with cyanine dyes. However, many NIR-absorbing dyes have short lifetimes, making them difficult to work with photochemically. In addition, most photochemistry functions by homolysis or heterolysis of bonds. These energies are on the order of 50-100 kcal/mol, or in the UV/visible regime. Near infrared light has significantly less energy. This requires a type of photochemistry that can operate at comparatively low energy. Photoreduction could be an option, but a NIR chromophore which features photoreduction without the generation of singlet oxygen has not yet been discovered. One option is the photochemical generation of an *ortho*-quinone methide (*o*QM).

This photochemistry has been observed in a number of UV and visible-absorbing chromophores, mostly substituted naphthols.[26-27] The mechanism has not been definitively worked out, but likely involves an excited state charge transfer followed by proton transfer from the hydroxyl group to the leaving group (Scheme 1c).[28-30] This proceeds to the *ortho*-quinone methide, which can be trapped out either by solvent or by a Diels-Alder reaction. Excited-state proton transfer, or photoacidity, is a photophysical phenomenon in which a

molecule becomes more acidic in the excited state.[31-34] Unlike with *ortho*-nitrobenzyl photochemistry, which involves bond homolysis which in some cases can require 50-100 kcal/mol reaction barriers, photoacidity is has a lower intrinsic reaction barrier.[35] Changing the pKa of an acid by 10 orders of magnitude corresponds to an energy change of only 14 kcal/mol. As a reference, the energy of a 1010 nm laser is 28 kcal/mol, which is twice what is needed to produce this shift. This suggests that very long wavelength NIR light (800-1000 nm) could be used to irradiate these molecules and initiate photochemistry.

One such chromophore is the heptamethine cyanine dye (Cy7). As discussed previously, it is a well-understood dye framework that can absorb into the NIR window. Further, substitution of the dye at the *meso* position is facile, either by reaction with soft nucleophiles or Pd-catalyzed cross coupling reactions.[36-39] The synthetic ease of substitution of Cy7 as well as its well-understood photophysical properties make it a useful starting point to building in photochemistry. This work will describe progress towards the design of a NIR chromophore capable of *o*QM generation.

6.3: Synthesis and characterization of a potential *meso*-acetylene aromatic Cy7 photocage

6.3.1: Experiments to determine the viability of the acetylene-phenol moiety

The previous chapter detailed the synthesis of a *meso*-acetylene-phenol Cy7 dye (**2**, **4** in Chapter 5). In addition to featuring a short wavelength induced cyanine emission, the dye could be photoacidic. When deprotonated, a resonance form of the dye can be drawn which neutralizes the positive charge. This can be represented visually by comparison of the electrostatic potential maps for **1** compared to **2** (a Cy5 vs. a Cy7) in both the protonated and deprotonated state (Figure 2). The deprotonated state of **2** can much more easily delocalize negative charge. The photophysical data and computations from Chapter 5 suggest that higher-lying Cyanine orbitals mix with the LUMO of the aromatic-acetylene group. This may result in excited state charge transfer in S1 and photoacidity. To test this, a number of photophysical and photochemical experiments were carried out.

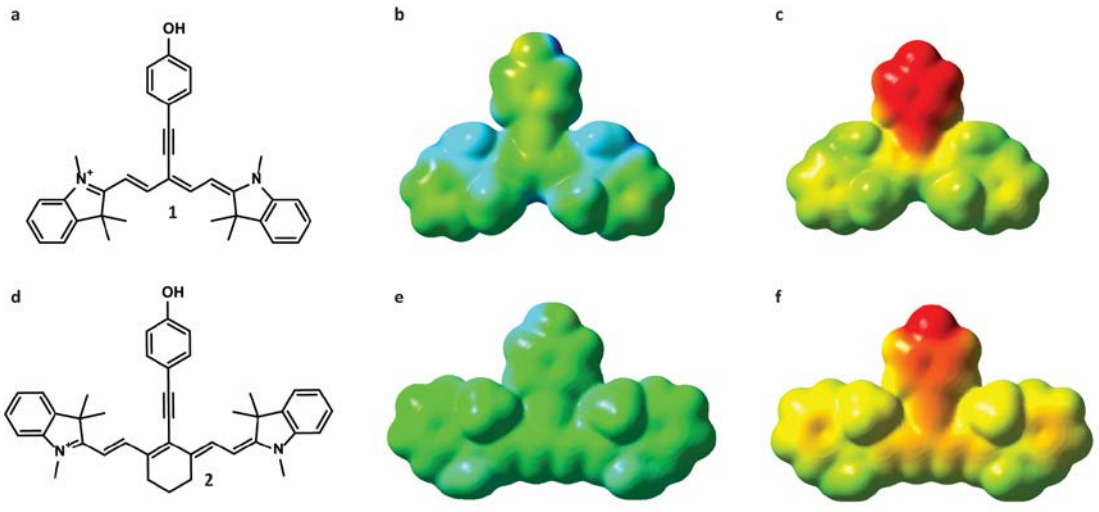


Figure 2: Electrostatic potential maps of (a) **1** in (b) protonated and (c) deprotonated states, as well as (d) **2** in (e) protonated and (f) deprotonated states. Energy surfaces calculated at DFT//CAM-B3LYP/6-31G(d,p). -150 to +460 kJ/mol.

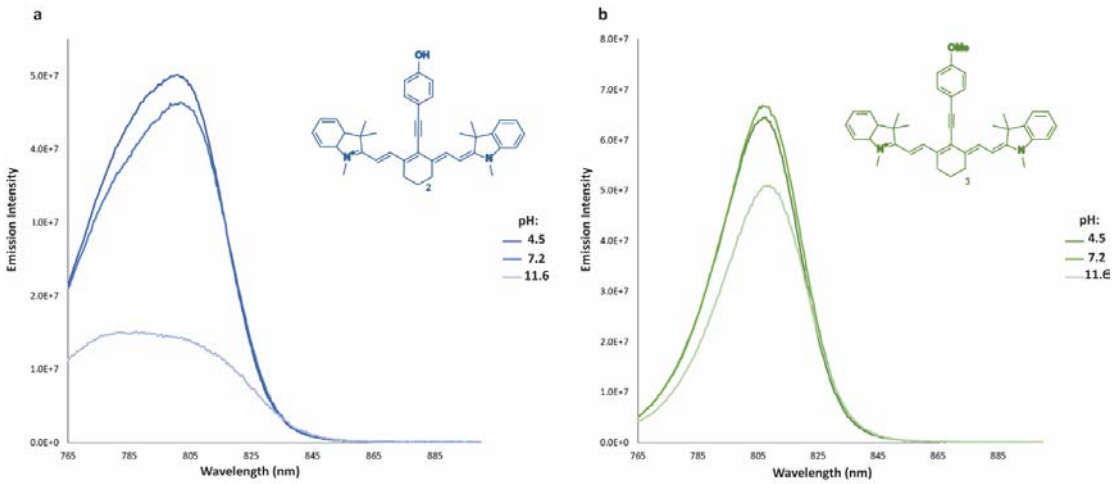


Figure 3: Emission of (a) **2** and (b) **3** in 85% buffer, 15% acetonitrile at various pHs.

First, the emission of the dye in both basic, neutral, and acidic media was measured (Figure 3). A photoacid such as 2-napthol will after excited state charge transfer and deprotonation emit at a red-shifted wavelength upon irradiation at the same absorbance.[40-41] The absorbance spectra will shift bathochromically upon addition of base as well. While **2** did not shift its absorbance or fluorescence, the overall fluorescence was significantly lower in basic conditions than for acidic or neutral, when compared to the methylated version **3** (**5** in Chapter 5). Because the fluorimeter used for these measurements could only record to 1000 nm, an extremely red-shifted emission peak cannot be ruled out. However, the reduction in emission at low pH suggests that **2** is photoacidic.

Solvatochromic analysis can also be used to measure photoacidity. The absorbance and emission of the studied dye can be taken in a number of different spectroscopically-pure solvents. The behavior of the solute (the dye) in these solvents can be related back to the properties of the solvents they were in. These solvent parameters include the solute's hydrogen bonding accepting ability (α), its hydrogen bond donating ability (β) and its polarizability (π*). Because these values are known for a number of different solvents, a

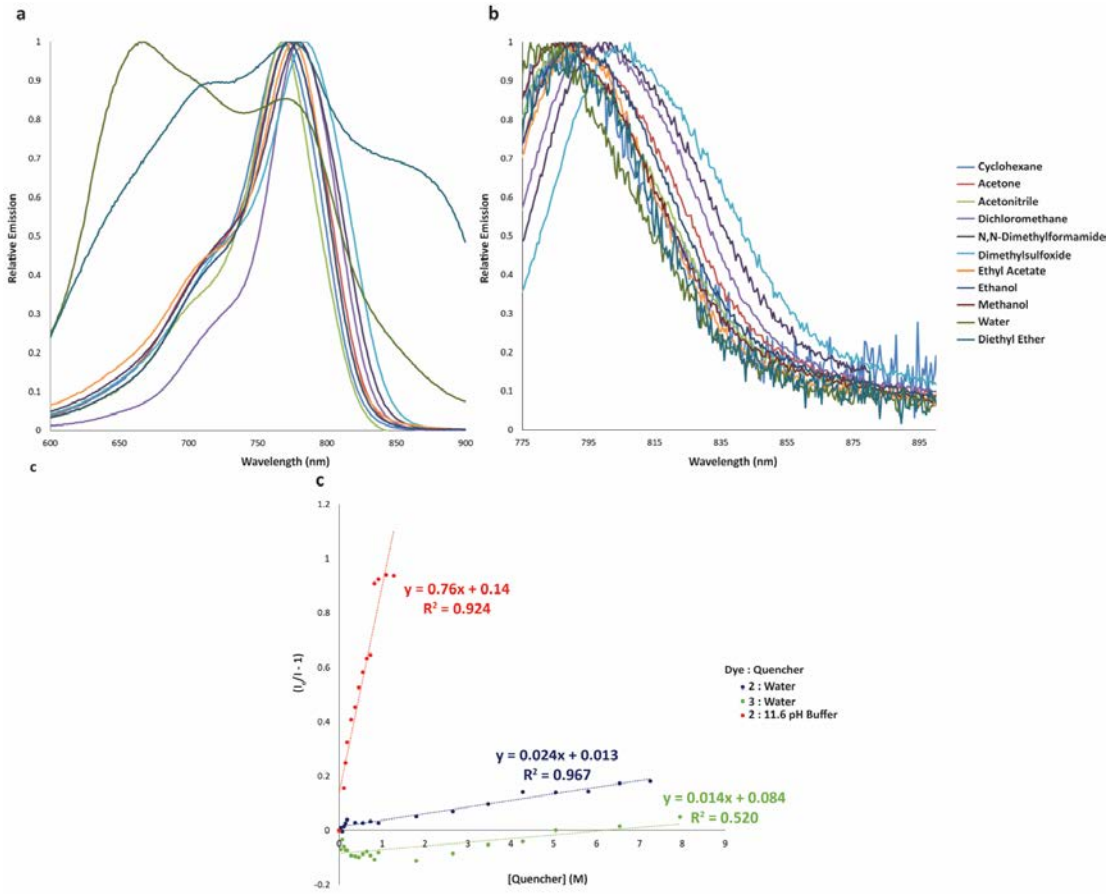


Figure 4: (a) Absorbance and (b) emission of **2** in a variety of different solvents. (c) Stern-Volmer quenching of **2** and **3** with water and pH 11.6 buffer.

four-variable regression (a Kamlet-Taft analysis) can be performed to determine the parameters the dye of interest.[41-46] This analysis was performed on Cy5, *meso*-chloro-Cy7, **1**, and **2** (Figure 4a-c). The data, however, show little correlation between any of the three parameters alone or any combination of the parameters for the dyes. The solvatochromic shifts of all studied cyanine dyes were relatively low, 10-15 nm at the maximum. The highest correlation was R = 0.82 (R^2 = 0.67) for the absorbance spectra of **2**.

An alternate measure of photoacidity is fluorescence quenching. Because the Frank-Condon state fluorescence of a photoacid like 2-naphthol is suppressed upon proton transfer, the addition of water or a proton acceptor to an organic solvent quenches that

Table 1: Kamlet-Taft analysis of the absorbance and emission data to yield a description of the solute in the ground and excited state. All parameters are in units of nm shifts.

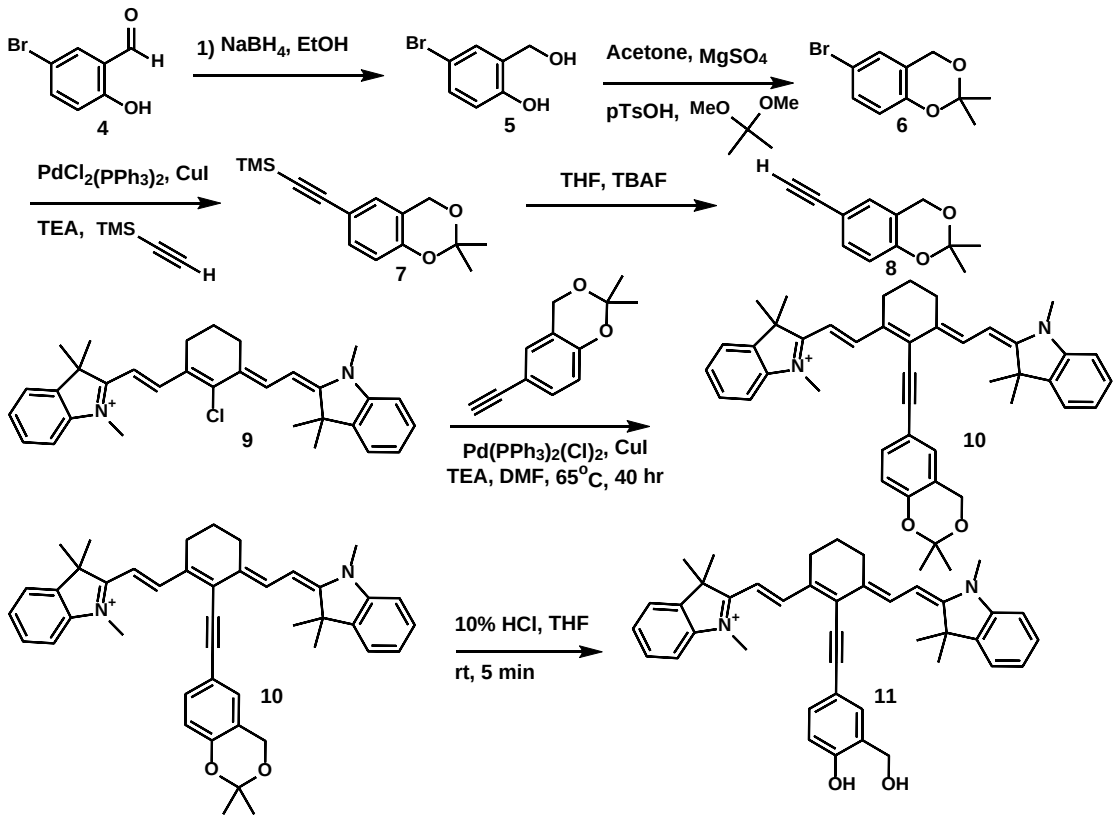
	Parameter	S0 (nm)	R	S1 (nm)	R	Change
1	π*	-2.77	0.13	3.09	0.26	5.86
	α	-7.82	0.53	-3	0.29	4.82
	β	3.11	0.01	0.78	0.06	-2.33
			0.55[1]		0.4[1]	
2	π*	7.9	0.56	15.4	0.54	7.5
	α	-6.5	0.46	-4.7	0.13	1.8
	β	4.9	0.3	2.8	0.17	-2.1
			0.82[1]		0.58[1]	
Cy5	π*	5.2	0.43	7.9	0.53	2.7
	α	-5.3	0.55	-5.3	0.45	-0.1
			0.71[2]		0.71[2]	
Cl-Cy7	π*	15.1	0.61	8.1	0.31	-7.0
	α	-6.9	0.34	-13.6	0.68	-6.6
			0.71[2]		0.76[2]	

[1] R value from 4-variable kamlet-taft analysis. [2] R value from 3-variable Kamlet Taft analysis

fluorescence. Similarly, because the emission of **4** is suppressed upon irradiation in basic solvent, differing amounts of basic buffer were added to a solution of **2** in acetonitrile. The relative reduction in fluorescence can be related to the concentration of the buffer.[34, 47-50] This is called a Stern-Volmer quenching, which can be quantified by a quenching constant K_{SV} found by the equation:

$$(2)\quad {}^{I_o}/_{I} = K_{SV} \times [Q]$$

Where I_o is the fluorescence intensity with no quencher and I is the intensity upon addition of a small amount of quencher. This procedure was repeated for **3** with water. While addition of pure water to both compounds did not result in a large Stern-Volmer quenching constant K_{SV}, addition of pH 11.6 buffered water produced significant quenching (Figure 4d). These data suggest that a photoacid-like mechanism is feasible for **2**.



Scheme 2: Synthesis of *meso* acetylene-*ortho*-methylhydroxylphenol Cy7 (**11**)

6.3.2: Synthesis of Cy7 ortho-quinone methide precursor

Scheme 2 describes the synthetic strategy to generate a first-generation photocage. The synthesis itself is similar to those described in the previous chapter. An ortho-hydroxymethyl phenol moiety is first generated through reduction of 5-bromo salicylic anhydride, then is protected with 2,2-dimethoxypropane to yield **6**. Then, dual Sonagashira reactions yield the protected Cy7 dye (**10**). These reactions are similar to those carried out for dyes in Chapter 5. **10** can be deprotected in relatively high yield with acid to yield the desired potential photocage **11**.

*6.3.3: Photophysical and photochemical characterization of **11***

The steady state absorbance and fluorescence of **10** and **11** are similar to **2**, though red-shifted in both absorbance and emission (Figure 5, Table 1). The absorbance of the protected compound and *o*QM precursor, however, were nearly identical. To test if this compound was a viable photocage, it and **11** were irradiated an arc lamp with a 686 nm cutoff for varying amounts of time. The first experiments, detailed in Scheme 3a, were in 50:50 water:methanol. If *o*QM generation was successful, the products would have been a 50:50 mixture of the **11** and its methylated derivative **12**. Irradiation was first performed under ambient conditions. These conditions resulted in decolorization and oxidative degradation of the Cy7 dye, a common photoreaction for dyes of this type.[51] After elimination of all oxygen in solution (via purging with argon or freeze/pump/thawing) and running the reaction under an argon atmosphere, no **12**

Table 1: Photophysical properties of synthesized dyes. λ_{max}: maximum excitation or emission wavelength. ε: extinction coefficient.

Dye	λ_{max} (nm)	ε x10^5 (M^{-1cm-1})	λ_{max} (nm)	φ_{em}
2	770	1	788	0.16
3	795	1	824	0.081
10	790	1	814	-
11	788	1	822	-

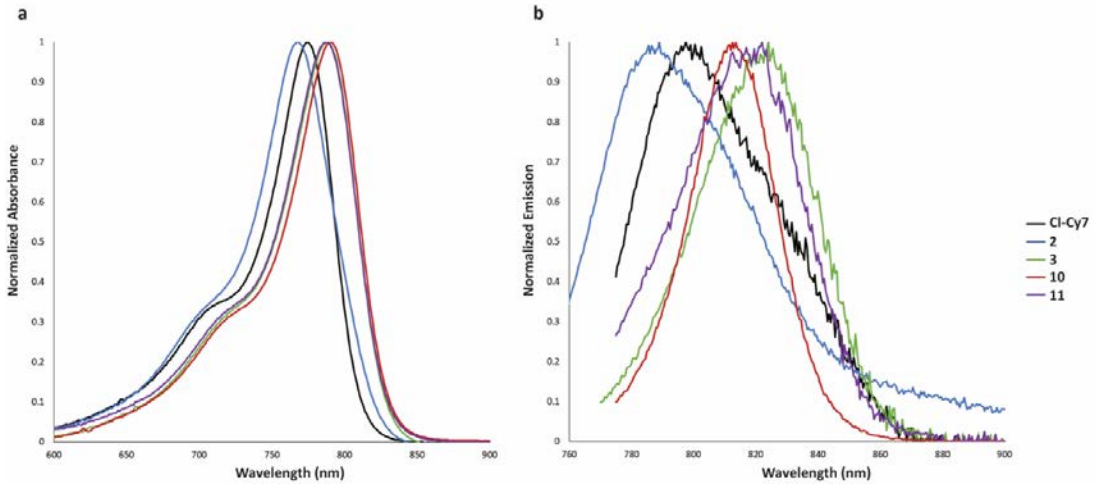
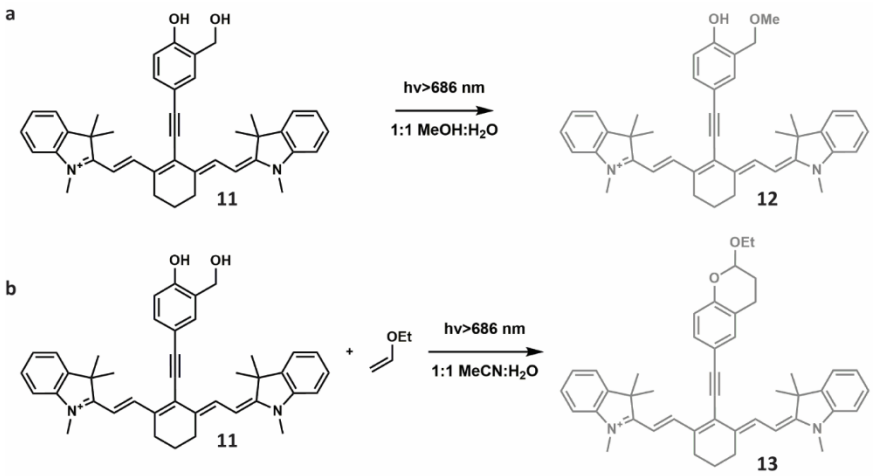


Figure 5: (a) Absorbance and (b) emission of relevant Cy7 dyes, including the potential photocage **11.**



Scheme 3: Attempted photochemical reactions of **11** with either (a) ethyl vinyl ether or (b) methanol:water with expected products **12** and **13**. Grey structures represent structures not observed with LC-MS monitoring.

was detected by analytical LC-MS. Even with irradiation times as long as 48 hours, no discernable *o*QM generation could be detected. The experiment was repeated with ethyl vinyl ether as the *o*QM quencher (Scheme 3b). Like the previous photoreaction, no **13** was detected after long irradiation times without oxygen. Because the chromophore did not exhibit the desired photochemistry, the photophysical characterization is left incomplete. The discussion section (**6.6**) will offer some conjecture as to why this photochemistry was unsuccessful.

6.4: Synthesis and characterization of a potential ethynyl-triphenylmethane cation derived NIR photocage

6.4.1: Theoretical justification

Triphenylmethane cation dyes are, like cyanine dyes, a well-studied dye framework, exemplified by fluorescein and rhodamine.[52] The structural requirement for dyes of this type is three aryl groups surrounding a carbocation. Addition of basic groups on two or more of the phenyl groups can polarize the cationic charge along one axis and

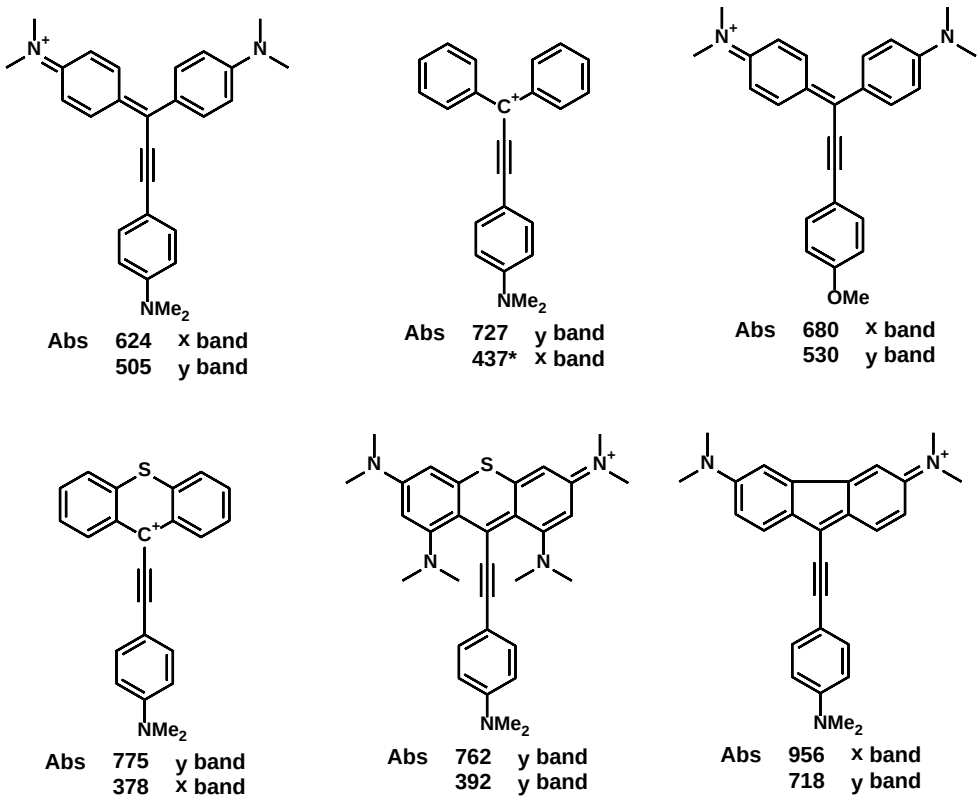
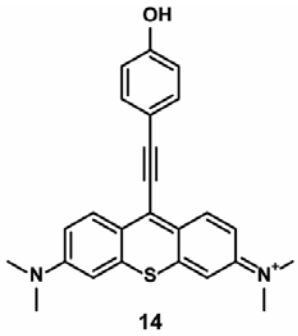


Figure 6: Previously-synthesized acetylene-triphenylmethane cation dyes. Below each is a calculated absorbance wavelength (PPP-CI), with the axis of excited state transition dipole moment.

cause a bathochromic shift, creating a push-pull chromophore. This has the added benefit of stabilizing the dye in non-organic solutions. A different way to red-shift the absorbance of these dyes is to increase planarity between the phenyl groups. This has been accomplished in two ways: the addition of bridge atoms such as an S or N-H group between two phenyl rings and extending the chromophore with an acetylene unit.[53-55] Interestingly, the addition of an acetylene can, depending on the substituents on the other two phenyl rings, change the direction of the transition state dipole moments of the first several excited singlet states. Figure 6 shows several previously-synthesized dyes and the absorbance and calculated polarizations of their first two absorbance bands.

This is relevant for the synthesis of photoacidic NIR dyes that can be adapted to *o*QM generators. As discussed in Chapter 5, the first excited state of cyanine dyes is

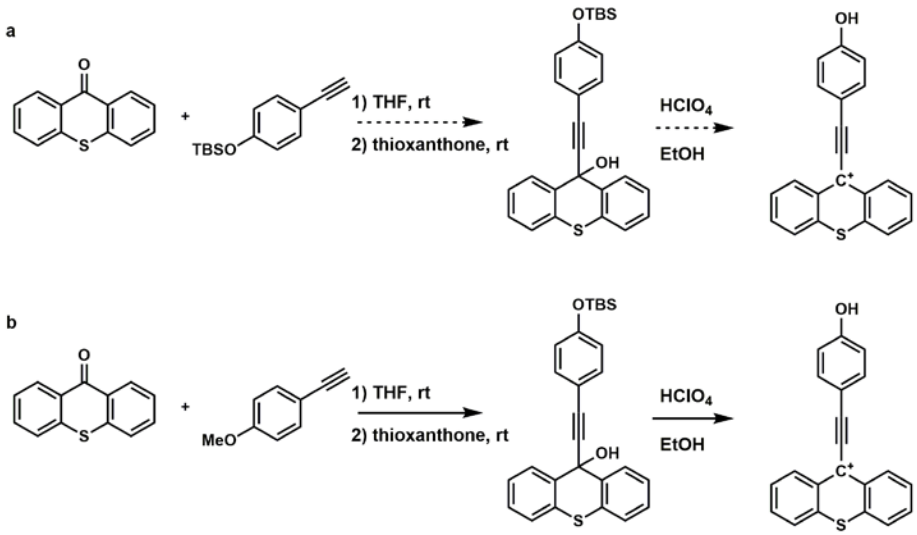


Dye	$\mu_x^{(S0\text{-}S1)}$	$\mu_y^{(S0\text{-}S1)}$	$\mu_x^{(S0\text{-}S2)}$	$\mu_y^{(S0\text{-}S2)}$	$f^{(S0\text{-}S1)}$	$f^{(S0\text{-}S2)}$
14	3.31	0.04	0.05	3.46	0.74	0.87
15	0.02	3.13	0.02	2.22	0.60	0.35

Figure 7, Table 2: Transition dipole moment μ magnitudes and oscillator strength f for **14** and **15**. Dipole moments in the x direction refer to the two adjacent phenyl rings, while the y direction is the acetylene group. Calculations performed at TDDFT//CAM-B3LYP/6-31G(d,p).

polarized along the x-axis (the axis of the polymethine chain). This may result in the cyanine excited state not including any branching chromophore such as with **4**, negating any photochemical use. An ethynyl-triphenylmethyl cation (eTPM) dye with a phenol group could be an effective photoacid in the first excited singlet state, provided the first singlet state is polarized in the direction of the acetylene group.

Time-dependant DFT calculations were performed on two dyes of this type, both of which contain an acetylene-phenol group on the methine cation (Figure 7). One (**14**) had two dimethylamino groups *para* to the central cation, where **15** did not. This was to test if the excited state polarization would change when stabilizing groups were introduced into the potential NIR dye. The results of the computations (Table 2) suggested that the dimethylamino groups would polarize the first singlet state of the dye along the axis of the two phenyl groups, or the x-axis. **15** would have a first excited singlet state that is polarized along the length of the potential photoacid. This dye could therefore potentially be a relevant starting point as a NIR photocage



Scheme 4: (a) Proposed synthesis of potential photoacidic acetylene-triphenylmethane cation dye **15**. (b) Synthesis of methylated version **18**.

6.4.2: Synthesis of a prototype photoacidic eTPM dye

A proposed synthesis of a prototype phenol dye is detailed in Scheme 4a and follows from literature precedent. The reaction would proceed through first deprotonating the protected 4-ethynylphenol, then reaction with thioxanthone. Subsequent one-pot deprotection and dehydration with acid would yield the putative photoacid **15**. To ensure that an eTPM dye without stabilizing basic groups would be stable in nonorganic solvents, a simplified anisole-based dye was first synthesized (Scheme 4b). Practically, the synthesis was the same as proposed, and proceeded efficiently to yield **18**, a purple-blue solid.

6.4.3: Photophysical characterization

While **18** did absorb in the red wavelength regime in organic solvents, upon solvation in ethanol or dimethylsulfoxide it quickly decolorized. This is likely due to nucleophilic addition into the cationic moiety. This is shown in part in Figure 7. Because the original design necessitates the removal of any solvating basic groups on either of the other two phenyl groups, this strategy was abandoned. Addition of solvating groups would

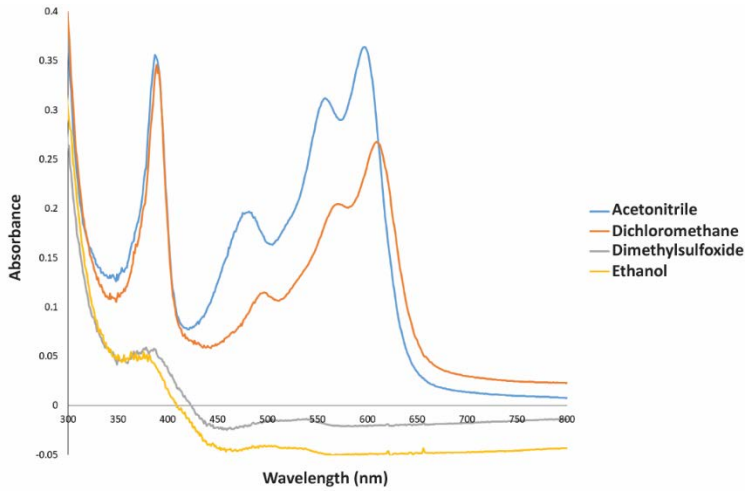


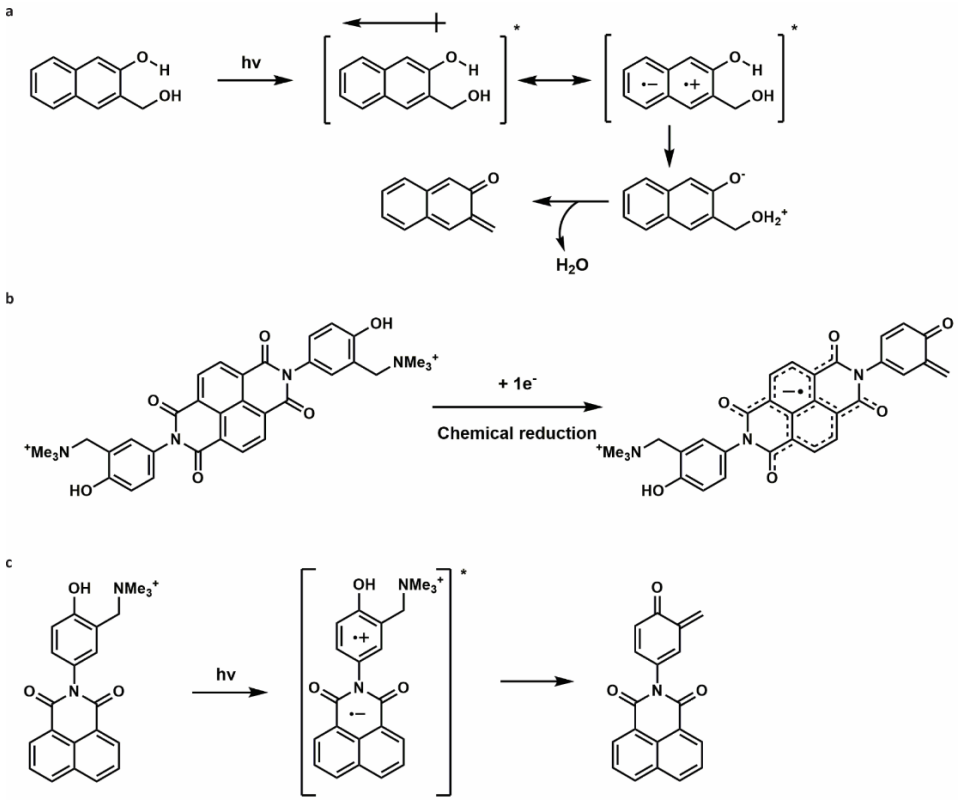
Figure 7: Absorbance of **18** in different solvents.

have switched the polarization from the y-axis to the x-axis. This would have defeated the original design of the potential photocage.

6.5: Progress towards development of an alternate Cy7-based PET *o*QM photocage

6.5.1: Theoretical and computational justification

While photoacidity is one accepted mechanism by which an *o*QM group can be generated photochemically, other mechanisms have been observed. For example, chemical and photochemical reduction has also been used to justify the generation of an *o*QM species (Scheme 5a). For example, a naphthalene diimide compound upon reduction with dithionite will release a trimethylammonium leaving group from its pendant *ortho*-methyltrimethylammonium phenol groups to an *o*QM intermediate (Scheme 5b).[56] Photolysis of a related compound (Scheme 5c) creates a radical cation/anion pair, which stimulates *o*QM formation as well.[30] Both of these examples point to an alternate



Scheme 5: *ortho* Quinone methide generation by chemical or photochemical reduction.

mechanism for this reaction (Scheme 5b-c). Namely, one electron reduction of a group next to an aryl *o*QM precursor can stimulate the photochemical reaction.

Photochemical electron transfer has been observed with NIR-absorbing Cy7 dyes (Figure 8a). These dyes are distinguished by their ability to perform photoinduced electron transfer (PET).[57-58] Upon irradiation at the cyanine wavelength, the thioaniline moiety will donate an electron into the vacant HOMO orbital, reducing its ability to fluoresce (Figure 8b). This phenomenon was suppressed upon addition of nitrogen chelators (Zn^{2+}) or at high pH, indicating that the presence of a lone pair is crucial to the mechanism. In addition, transition-state calculations (DFT//CAM-B3LYP/6-31G(d,p)) of a phenol *o*QM reaction show structurally that the proton transfer occurs before the water leaves (Figure 8c-e).

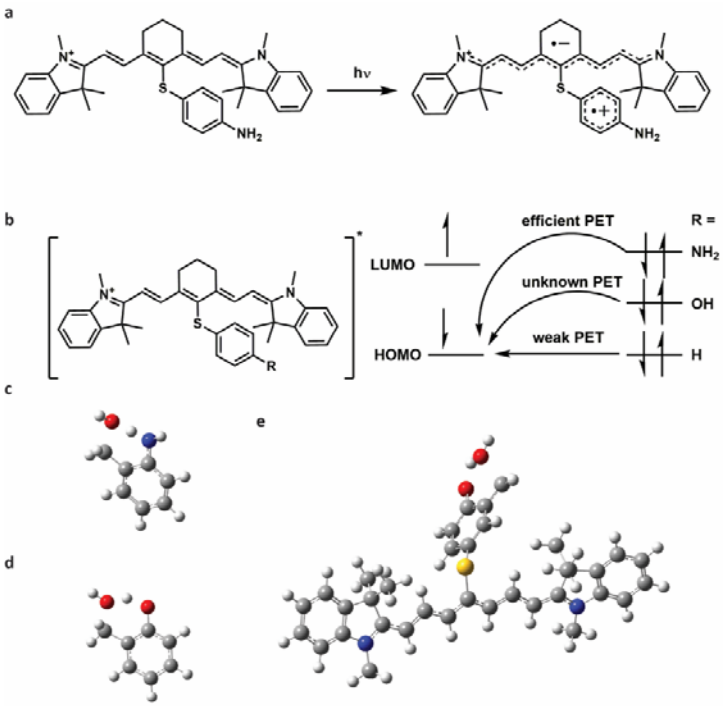
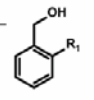


Figure 8: (a) Previously-reported PET Cy7 dye. (b) Proposed mechanism of PET, where single electron transfer from the thio-linked aniline occurs in the cyanine excited state. *o*QM transition state structures of (c) phenol and (d) aniline groups with water as the leaving group, as well as (e) a Cy7 – PET analogue.

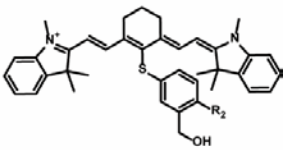
Therefore, it may not be the rate limiting step. A chromophore that absorbs in the NIR and can accept an electron from a pendant *o*QM precursor could be a viable photocage.

In the original study, the authors found that a *para*-amino thiophenol Cy7 dye performed best for PET fluorescence quenching (**24**). However, our calculations as well as previous studies suggest that an amino group is inferior to a hydroxyl group in performing an *o*QM photoreaction (Table 3, structures **19** and **20**). One aim of the preliminary studies was to determine the feasibility of PET from the phenol. We also studied the transition state of the *o*QM photogeneration in the context of the full Cy7 dye (Table 3, structures **21** and **22**). Fortunately, the addition of the full Cy7 chain did not affect

	S0		S1	
Compound	Transition State	Final Products	Transition State	Final Products
19	40.4	35.1	2.9	-15.0
20	49.6	43.9	14.1	9.2
21	41.2	36.7	40.8	36.6
22	50.7	45.9	45.4	46.0



19: R_1 = OH
20: R_1 = NH_2



21: R_2 = OH
22: R_2 = NH_2

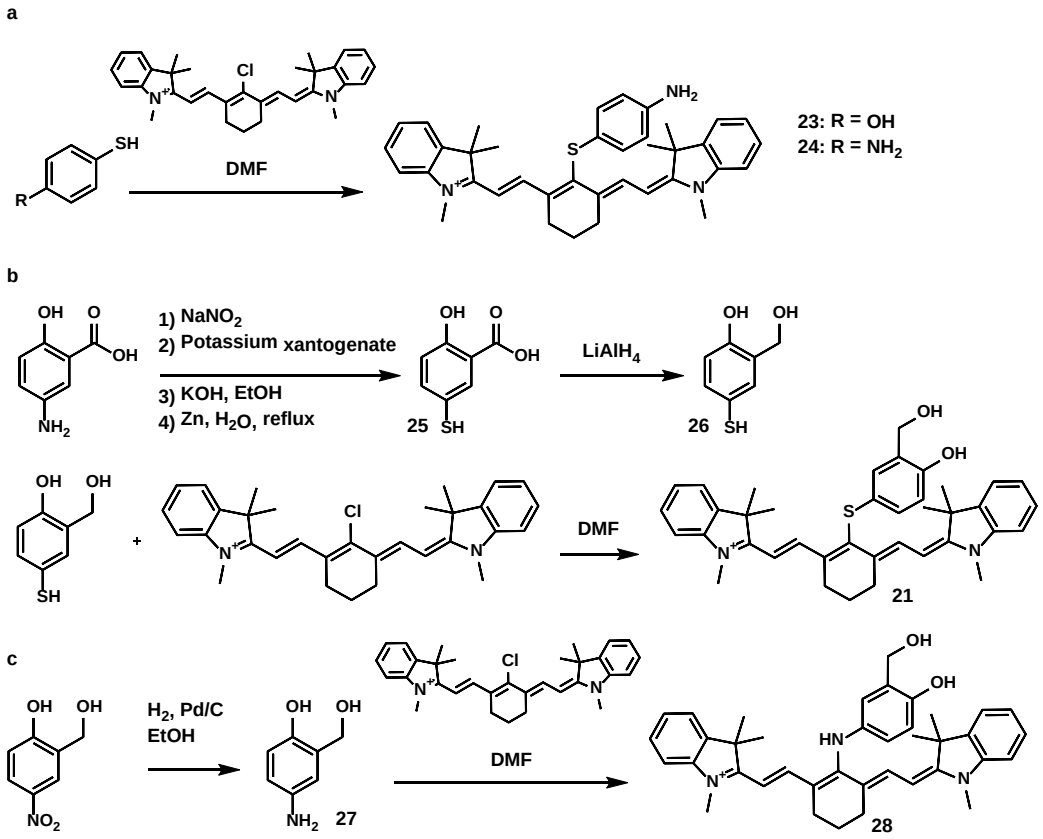
Table 3: Transition state calculations in the ground and excited states for *o*QM generation for the structures 19-22. Transition state structures found by DFT/CAM-B3LYP/6-31G(d,p) and TDDFT calculations performed as SPE calculations.

the relative energies of the transition states in the S0 state. The transition states for cyanine dyes in the S1 state were similar to the S0 counterparts. This is likely due to the fact that the S1 state of the cyanine dye is not mixed with the thiophenol LUMO, so will not by itself stabilize the transition state as seen with **19** and **20**.

In addition to thio-linked *o*QM precusers, amine-linked derivatives were studied as well. Because of the resonance form between the Cy7 and the amine, the dye is blue shifted to a large extent, but there is significant precedent that dyes of this type are useful in sensing.[36] It is very facile synthetically, with the aniline *o*QM precursor readily available (Scheme 6c).

6.5.2: Synthesis of Cy7 PET dyes

Scheme 6 shows the synthesis of the PET Cy7 dyes. Most of the relevant dyes could be synthesized in a manner similar to the literature precedent, where the thiophenol or aniline is added to the reaction and stirred in the dark for 1-3 hours at room temperature under argon (Scheme 6a). The mixture is then added to diethyl ether and filtered, yielding mostly product, which can then be further purified on silica gel. The intermediates to the synthesized of *o*QM precursors **21** and **28** were made using published procedures.[59] 4-hydroxy-3-methylhydroxy-thiophenol was synthesized through a multistep one-pot reaction followed by reduction with $LiAlH_4$. 4-hydroxy-3-methylhydroxyaniline was synthesized through a simple reduction from the aldehyde (Scheme 6b,c). Unlike with the cyanine dyes from the previous chapter, these do not require HPLC purification, and can be produced in much higher yields.



Scheme 6: Synthesis of (a) PET cyanine dyes, (b) potential PET Cy7 *o*QM generating dye, and (c) a dye with an amino linker.

6.5.3: Photophysical and photochemical studies of the synthesized dyes

The absorbance and emission spectra of the synthesized dyes are shown in Figure 10, along with the absorbance and emission of *meso*-(Cl)Cy7. Their absorbance and emission behavior is not qualitatively different from the other Cy7 dyes studied. The exception is **28**, whose absorbance in acetonitrile is noticeably blue shifted. Anecdotally, this absorbance shifts dramatically with different solvents. One measure of the dyes' ability to perform PET is their quantum yield of fluorescence. A lower quantum yield would mean more productive PET occurring. To that end, the quantum yields of fluorescence of **21**, **23**, **24**, and **28** were measured (Table 4). The fluorescence quantum yield of **23** was significantly higher than the other dyes studied. This supports the hypothesis that the

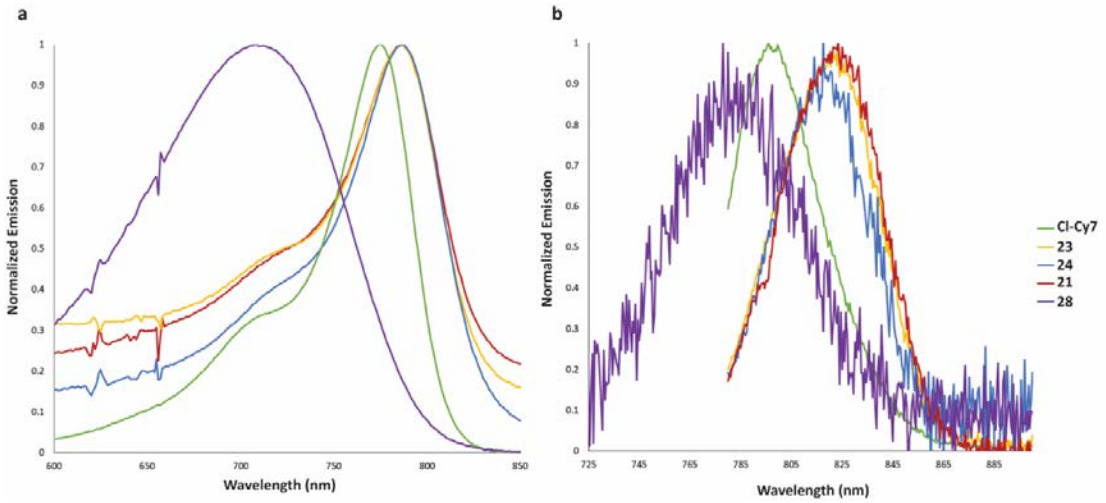


Figure 10: (a) Absorbance and (b) emission of relevant PET Cy7 dyes.

Table 4: Photophysical properties of synthesized dyes. λ_{max}: maximum excitation or emission wavelength. ε: extinction coefficient. φ_{em}: quantum yield of emission.

Compound	ε (10^{-5} M^{-1} cm^{-1})	λ_{max} (Absorbance)	λ_{max} (Emission)	$\varphi_{emission}$
23	2	785	823	0.12
24	2	785	823	0.015
21	2	785	823	0.037
28	0.34	710	791	0.00047

aniline is necessary for efficient PET. However, dye **21** had a quantum yield on the same order of magnitude as **23**. This could mean that it may feature PET in the excited state. While the quantum yield for **28** is very low, it cannot be compared to the other dyes measured. It is, however, within the range of similar compounds of its class.

Dye **23** was also investigated with Stern-Volmer analysis and it and **24** were subjected to Kamlet-Taft analysis. The results are shown in Figure 11a-b. Multivariate analysis of both dyes tracked with previous results (Figure 4a-c). However, the dyes were not especially solvatochromic, so it is difficult to draw any conclusions from the collected data. The collected Stern-Volmer quenching (Figure 11b) is similar: while **23** is weakly quenched by water, K_{SV} is very small as to be effectively zero.

a

	Parameter	S0 (nm)	R	S1 (nm)	R	Change
21	π*	3.54	0.24	4.04	0.22	0.5
	α	-10.15	0.67	-9.81	0.74	0.34
	β	4.8	0.13	-0.41	0.089	-5.21
		0.76		0.78		
22	π*	5.95	0.4	-1.81	0.0028	-7.76
	α	-10.89	0.81	4.51	0.54	4.51
	β	7.24	0.22	-6.12	0.3	-1.12
		0.88		0.63		
Cl-Cy7	π*	15.1	0.61	8.1	0.31	-7.0
	α	-6.9	0.34	-13.6	0.68	-6.6
		0.71		0.76		

b

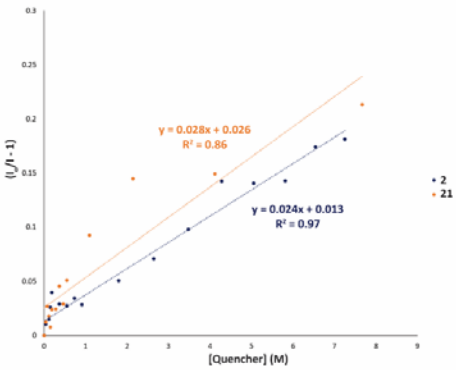


Figure 11: (a) Kamlet-Taft analysis of the absorbance and emission data to yield a description of the solute in the ground and excited state. (b) Stern-Volmer quenching of **2** and **21** with water.

To determine if the synthesized molecules are photochemically active, the compounds were irradiated with a 780 nm LED with and without O_2. As before, the compounds were irradiated in both a 1:1 mixture of methanol:water and with ethyl vinyl ether added. Unfortunately, even under long irradiation times no *o*QM generation was observed. With O_2 present in the solution, **21** degraded as before. However, **28** was stable for long periods of time when irradiated open to air with 780 nm and 660 nm LEDs.

6.6: Discussion

This chapter describes progress towards a NIR photocage. The approach to the photocage design was to start at validated NIR-absorbing dyes and build in photoreactivity. The intention was to design a dye that could initiate *ortho* quinone methide generation via either PET or photoacidity, both of which involve large charge reorganization in the excited state. While both photoacidic (section **6.3**) and PET (section **6.5**) dyes were synthesized, neither dye exhibited the desired photoreactivity. This could be for several reasons.

First, cyanine dyes have a very short excited state lifetime. Prototype dyes have fluorescence lifetimes measured as at most 1.28 ns (in DMSO) and 0.36 ns (in water).[60] While proton transfer is also very fast,[61-62] the kinetics of proton transfer are roughly on

the same order as the cyanine excited state lifetime. Likewise, while the fluorescence of **21** is less efficient because of PET, the lifetime of the transient radical anion/cation pair may be too short for the photochemistry to effectively occur. Cyanine-based photocages have been synthesized recently, but these photocages oxidatively decage, releasing singlet oxygen and destroying the dye in the process.[63]

While the kinetics of any potential cyanine photoacid may not lend themselves well to useful photochemistry, the thermodynamics also prove challenging. The only photophysically available evidence for photoacidity of **2** was reduced fluorescence while in a pH 11.6 buffer. This would imply that the increased excited state acidity only brings the pK_a of the phenol down to a point in which it is deprotonated, or activates an alternate fluorescence reduction pathway, in highly basic conditions. This implies that the change in pK_a is small enough that the reduction in fluorescence is only observable upon a fully deprotonated state or that the lifetime of the excited state is small enough that a pK_a change (and therefore a transiently deprotonated state) is not observable with steady-state spectroscopy.

In addition, the first excited state of cyanine dyes is polarized along the axis of the polymethine chain. All of the excited state calculations we have performed, in addition to those in the literature, corroborate this. The photoacidic mechanism proposed in **6.3** does not fall neatly in this category. The excited state of the *o*QM precursor is not included in the first cyanine excited state, as the ethynylphenol moiety is orthogonal to the axis of the transition dipole moment. The mechanism of the proposed PET-based photocage **6.5** attempts to surmount this obstacle by decoupling the excited state transition dipole from

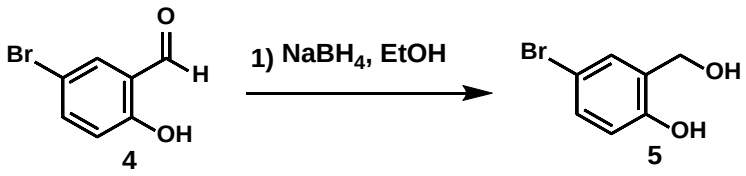
the photochemistry itself through PET. However, the lifetime of the cyanine dye limits the useful timescale of PET.

That being said, there are potential unexplored avenues for these dyes. For example, *o*QM generation proceeds more rapidly when there is a good leaving group in the *ortho* position. Substitution of the hydroxyl group with a trimethylammonium or other, better leaving group could improve any photochemistry that is limited by the current system. Alternatively, the *o*QM precursor could be designed in the direction of the excited state polarization.

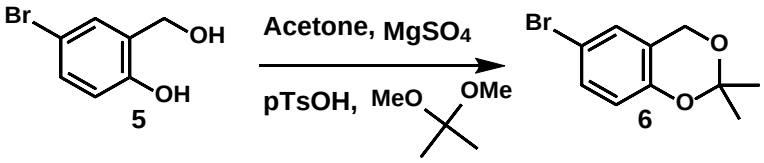
6.7: Methods

Organic Synthesis:

Unless otherwise stated, reactions were carried out under argon. Commercially available reagents were obtained from Sigma Aldrich, AK Scientific, Alfa Aesar, or Acros Organics and used without further purification. Solvents were used as received. Thin-layer chromatography with Sigma Aldrich silica gel coated plates with fluorescent indicator (0.25 mm) was used to monitor reactions. Silica gel chromatography was conducted as described by Still et al.,[64] with silica gel purchased from Alfa Aesar (60 Å, 230-400 mesh). NMR spectra were recorded on Varian (300, 400, 500, or 600 MHz) spectrometers. Mass spectra (ESI) were obtained with an Agilent liquid chromatography-mass spectrometry system. Isolated yields are reported except for compounds purified by HPLC, where crude recovery yields are reported and only a portion of the crude was purified by HPLC.

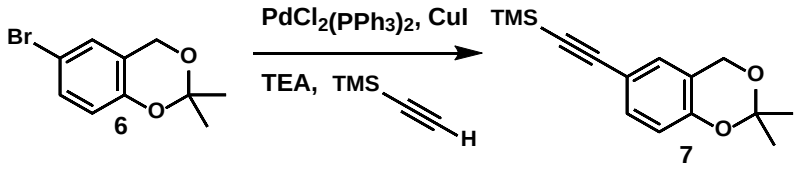


Add 50 mL dry ethanol to a dry flask containing 5-bromosalicaldehyde (**4**) (1.0 g, 4.97 mmol). Add 0.282 g (7.46 mmol) pellets of sodium borohydride and let stir at room temperature for 30 minutes or until all pellets dissolve. After the half-hour has elapsed, quench the reaction with saturated aqueous ammonium chloride. Extract this with ethyl ether (3x 50 mL). Collect the organic layers and extract 1x with brine (100 mL). Dry with magnesium sulfate and concentrate on a rotovap. This product was taken on to the next reaction. **5** is a white solid. ^{1}H NMR (400 MHz, Methanol-d4) δ 7.40 (d, J = 2.5 Hz, 1H), 7.18 (dd, J = 8.5, 2.5 Hz, 1H), 6.67 (d, J = 8.5 Hz, 1H), 4.59 (s, 2H). m/z = 203.1. Yield: 96%. Data match previously synthesized compound.[65]

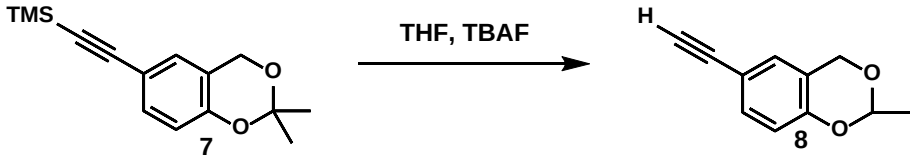


A mixture of 3-bromosalicyl alcohol (**5**, 0.97g, 4.77 mmol), 2,2-dimethoxypropane(6.1 mL, 47.7 mmol), para-toluenesulfonic acid monohydrate (0.0991 g, 0.47 mmol) and 1.8 g of anhydrous magnesium sulfate in 60 mL of acetone was heated at 40 °C for 3 d. The solvent was removed under reduced pressure and the residue was dissolved in ethyl acetate and water. The aqueous phase was extracted twice with ethyl acetate (40 mL). The combined organic phases were washed twice with 1 M sodium hydroxide and water, dried with sodium sulfate, and concentrated under reduced pressure. **6** is a pale yellow solid. ^{1}H NMR (400 MHz, Chloroform-d) δ 7.30 – 7.21 (m, 1H), 7.10

(dt, J = 2.1, 1.0 Hz, 1H), 6.70 (dd, J = 8.7, 1.0 Hz, 1H), 4.80 (d, J = 1.2 Hz, 2H), 1.53 (d, J = 0.9 Hz, 5H). m/z = Yield: 68%. Matches previously synthesized compound.[66]

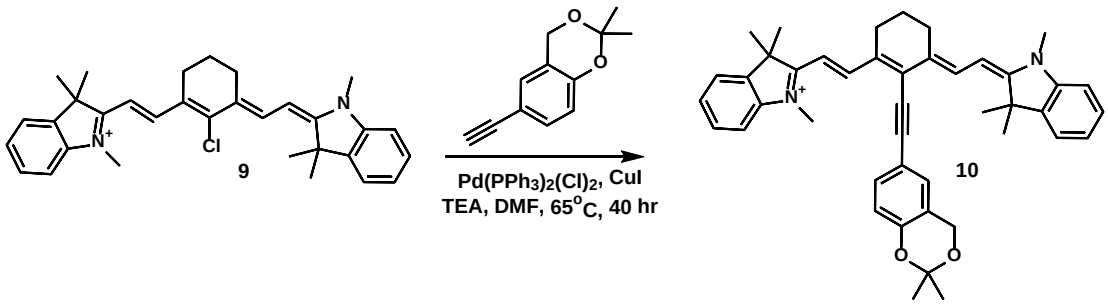


To a flame-dried 50 mL three-necked flask, 0.791 g (3.25 mmol) **6**, 125 mg (0.65 mmol) copper (I) iodide, and 235 mg (0.325 mmol) $Pd(Cl)_2(PPh_3)_2$ were added. These were dissolved in NEt_3 (30 mL), and ethynyltrimethylsilane (1.38 mL, 9.75 mmol) was added dropwise. The reaction was heated to 80^0C and stirred for 3 hours. The reaction was then filtered, and concentrated. The brown residue was purified by silica gel column chromatography with a mobile phase of EtAc:Hexane 40:60. **7** is a light brown solid. ^{1}H NMR (300 MHz, Chloroform-d) δ 7.32 – 7.19 (m, 1H), 7.12 – 7.08 (m, 1H), 6.72 (ddd, J = 10.0, 8.6, 0.4 Hz, 1H), 4.87 – 4.59 (m, 2H), 1.59 – 1.47 (m, 6H), 0.19 (s, 10H). m/z = 260.3. Yield: 98%. Previously synthesized and matches characterization data.[67]

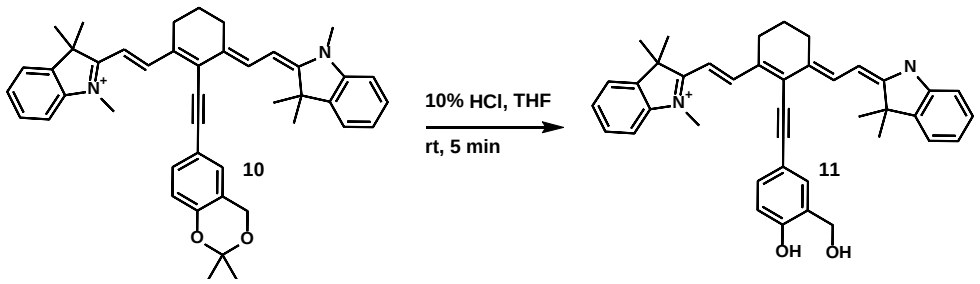


Add 30 mL dry THF to a dry flask charged with **7** (0.83 g, 3.18 mmol). Add 6.11 mL of 1.0 M tetrabutylammonium fluoride solution in THF. Let the reaction stir at room temperature for 30 minutes. Quench with 1:1 aqueous ammonium chloride:water. Extract with ethyl acetate three times. Then collect the organic layers, dry on magnesium sulfate, concentrate, and dry under vacuum. **8** is unstable on silica so was not purified further. ^{1}H NMR (300 MHz, Chloroform-d) δ 7.31 – 7.20 (m, 1H), 7.10 (dt, J = 2.5, 0.9 Hz, 1H), 6.70

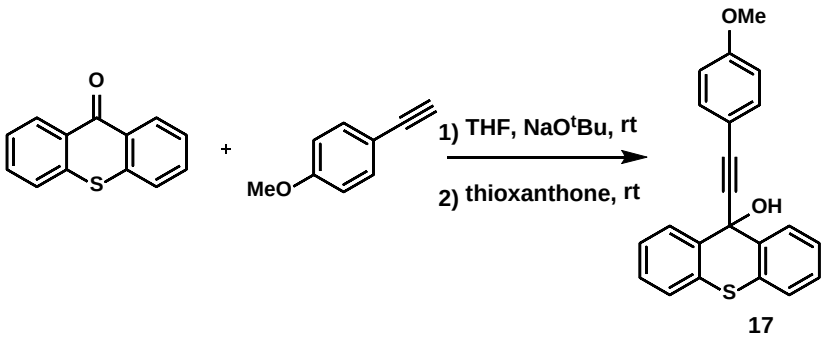
(d, J = 8.7 Hz, 1H), 4.81 (d, J = 0.9 Hz, 3H), 3.30 – 3.08 (m, 1H), 2.98 (s, 0H), 1.53 (s, 5H). m/z = 188.23. Yield: 85%. Matches previous characterization data.[67]



Dissolve 0.4947 g of **9** (synthesized in the same way as in Chapter 5) in 10 mL of dry DMF to achieve a 0.1 M solution. Then, dissolve 0.6907 g of **8** in 0.39 mL of dry DMF to achieve a 1.0 M solution. Weigh out the copper iodide and $PdCl_2(PPh_3)_2$ and put into a flame-dried three necked flask with a stir bar. Add the cyanine dye solution in DMF and heat to 50C. Subsequently, add dropwise the ethynylphenol, and the TEA. The reaction was monitored by analytical liquid chromatography-mass spectroscopy. After completion, the reaction was quenched with brine and extracted with DCM three times. The organic layers were collected and dried over magnesium sulfate, then concentrated and dried under vacuum. The blue mixture was purified first by silica gel chromatography with a mobile phase of MeOH:DCM 10:90 followed by HPLC purification to yield **10** as a green solid. ^{1}H NMR (400 MHz, Chloroform-d) δ 8.42 (d, J = 14.1 Hz, 2H), 7.43 (dd, J = 8.6, 2.0 Hz, 1H), 7.41 (s, 0H), 7.38 (dd, J = 7.7, 1.2 Hz, 2H), 7.34 – 7.30 (m, 2H), 7.24 – 7.15 (m, 4H), 6.93 (d, J = 8.4 Hz, 1H), 6.24 (d, J = 14.1 Hz, 2H), 4.91 (s, 2H), 3.73 (s, 6H), 2.69 (t, J = 6.2 Hz, 4H), 1.98 – 1.93 (m, 6H), 1.70 (s, 12H), 1.62 (s, 6H). ^{13}C NMR (101 MHz, Chloroform-d) δ 171.76 , 145.54 , 142.95 , 142.20 , 140.66 , 133.60 , 128.80 , 128.22 , 124.92 , 121.97 , 117.93 , 113.78 , 110.57 , 101.19 , 100.52 , 84.24 , 60.58 , 48.92 , 32.15 , 28.24 , 24.88 , 20.53. m/z = 635.4 Yield: 4.6%

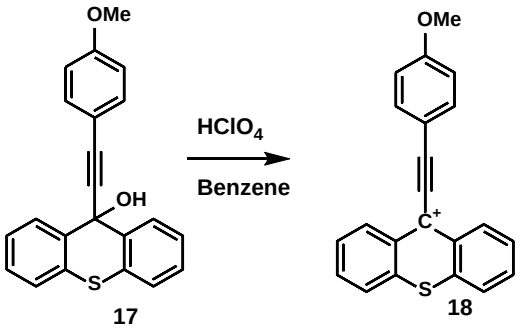


Dissolve **10** (93.6 mg, 0.15 mmol) in 1.5 mL of dry THF to achieve a 0.1 M solution. Then, add 150 uL of this to 15 mL of dry THF. Stir under argon for one half hour. Then, add 1.5 mL of HCl and continue stirring. Monitor using analytical liquid chromatography-mass spectroscopy. Quench with sodium bicarbonate (20mL). Remove the THF with a rotovap, then extract the remaining aqueous solution with $CHCl_3$. Dry the chloroform with sodium sulfate, then take solvent off and dry under vacuum. Purify by silica gel chromatography with a mobile phase of MeOH:DCM 10:90. **11** is a green solid. ^{1}H NMR (400 MHz, Chloroform-d) δ 8.55 (d, J = 14.0 Hz, 2H), 8.02 (s, 2H), 7.53 (d, J = 7.5 Hz, 2H), 7.38 (d, J = 11.7 Hz, 6H), 7.14 (d, J = 8.1 Hz, 1H), 7.10 (d, J = 8.1 Hz, 2H), 6.05 (d, J = 14.0 Hz, 2H), 4.95 (s, 2H), 3.71 (d, J = 7.5 Hz, 2H), 3.61 (s, 6H), 2.66 – 2.56 (m, 4H), 2.22 (t, J = 7.6 Hz, 2H), 1.76 (s, 12H). m/z = 595.7. Yield: 95%.

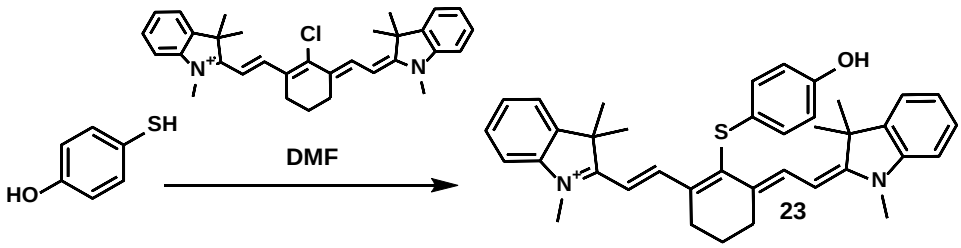


A mixture of 4-ethynylanisole (0.181 g, 1.4 mmoL) and sodium *tert*-butyl alkoxide (0.386 g, 3.5 mmol) in tetrahydrofuran (15 mL) was stirred at room temperature for 5 min under argon and a solution of the thioxanthone (0.234g, 1.12 mmol) in THF (10 ml) was

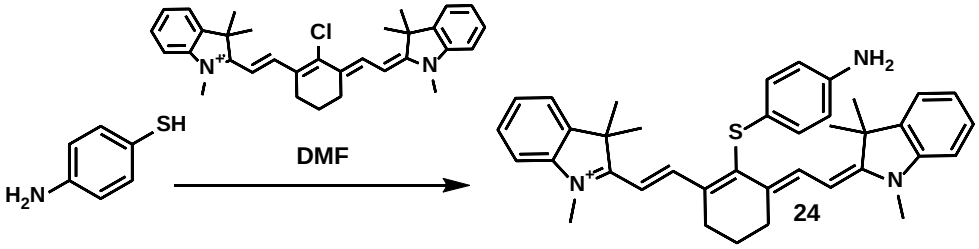
then added. The mixture was stirred for 30 min at room temperature. The solution was then poured into water (20 mL). The organic layer was extracted with benzene (75 ml x 2), and the extract was washed with water, and dried. Evaporation of the extract resulted in pale yellow crystals, and these were recrystallized from ether to give **17**. ^{1}H NMR (500 MHz, Chloroform-*d*) δ 8.30 – 8.19 (m, 1H), 7.59 – 7.55 (m, 1H), 7.54 – 7.51 (m, 1H), 7.44 – 7.39 (m, 1H), 7.37 (dd, *J* = 7.5, 1.6 Hz, 1H), 6.96 – 6.81 (m, 2H), 3.86 (d, *J* = 4.2 Hz, 3H). m/z = 344.3



Perchloric acid (70%; 3 drops) was added to a solution of the crude product **17** in 15 mL of benzene. The mixture was stirred for 1 h. The resulting precipitates were filtered off, washed with ether, and dried to give the salt as deep green crystals. Recrystallization from $CHCl_3$-Diethyl ether gave **18**. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.52 – 8.42 (m, 2H), 8.26 – 8.16 (m, 4H), 8.09 (d, J = 8.8 Hz, 2H), 7.69 – 7.54 (m, 2H), 7.13 (d, J = 8.7 Hz, 2H), 3.97 (s, 3H). m/z = 327.4.

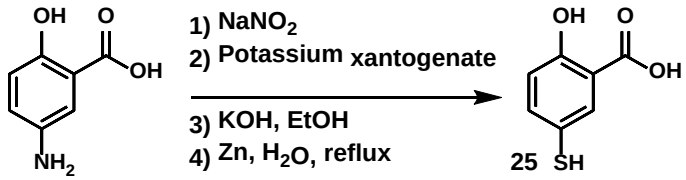


To a flask charged with 0.0705 mg **9** (0.136 mmol), 3 mL DMF purged with argon was added and stirred under argon. 22 mg 4-mercaptophenol (0.16 mmol) was added to the flask and monitored with analytical liquid chromatography-mass spectroscopy. Reaction proceeded for one hour in the dark. Following completion, ether was added to the reaction, precipitating **23**. The reaction was filtered off, then **23** dissolved in dichloromethane and extracted with brine. The green mixture was purified by silica gel chromatography with a mobile phase of MeOH:DCM 10:90. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.74 (d, J = 14.1 Hz, 2H), 7.35 (ddt, J = 8.2, 7.5, 1.1 Hz, 2H), 7.32 – 7.28 (m, 2H), 7.19 (tt, J = 7.4, 0.9 Hz, 2H), 7.12 – 7.08 (m, 2H), 7.08 – 6.97 (m, 4H), 6.09 (d, J = 14.1 Hz, 2H), 5.30 (d, J = 0.8 Hz, 0H), 3.62 (d, J = 0.8 Hz, 6H), 2.69 (t, J = 6.4 Hz, 4H), 2.06 – 1.90 (m, 2H), 1.51 (d, J = 0.7 Hz, 12H). m/z = 573.8. Yield: 84%



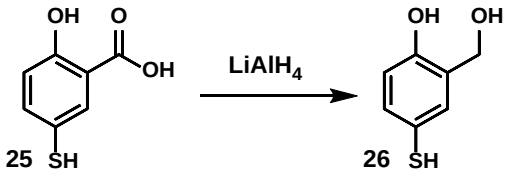
To a flask charged with 0.0953 mg **9** (0.183 mmol), 3 mL DMF purged with argon was added and stirred under argon. 28 mg 4-mercaptoaniline (0.22 mmol) was added to the flask and monitored with analytical liquid chromatography-mass spectroscopy. Reaction proceeded for one hour in the dark. Following completion, ether was added to the reaction, precipitating **24**. The reaction was filtered off, and then **24** was dissolved in

dichloromethane and extracted with brine. The green mixture was purified by silica gel chromatography with a mobile phase of MeOH:DCM 10:90. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.74 (d, J = 14.2 Hz, 2H), 7.40 – 7.33 (m, 2H), 7.30 (dd, J = 7.6, 1.2 Hz, 2H), 7.19 (dd, J = 7.9, 7.0 Hz, 2H), 7.15 – 7.09 (m, 2H), 7.05 – 6.94 (m, 2H), 6.65 – 6.58 (m, 2H), 6.18 (d, J = 14.1 Hz, 2H), 3.68 (d, J = 0.7 Hz, 6H), 2.74 (t, J = 6.2 Hz, 4H), 1.99 (p, J = 6.3 Hz, 2H), 1.55 (d, J = 0.7 Hz, 12H). m/z = 573.8. Yield: 87%. Matches previously synthesized compound.[57]



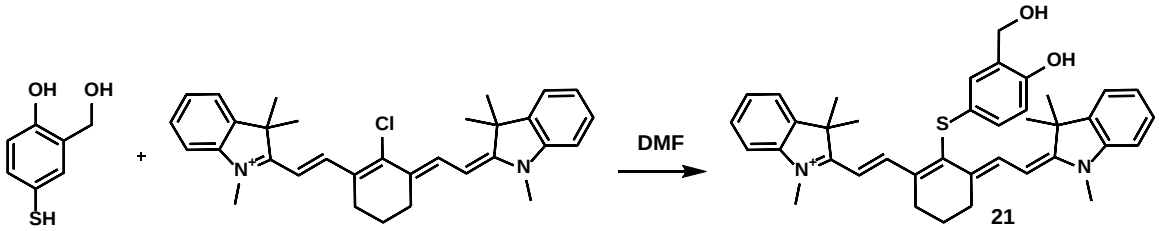
In a three necked round bottomed flask, 5-amino-salicylic acid (2 g, 13.06 mmol) was dissolved in 20 mL Water and 2.5 mL concentrated sodium hydroxide solution (50% w/w). Sodium Nitrite (0.902, 1 eq) was added and the mixture was cooled to 5°C using an ice bath to obtain a solution of deep purple color. In a separate flask, equipped with a thermometer, water (5 mL) and hydrochloric acid (4.6 mL, 32%) were cooled to 5°C. The nitrite and aniline solution was added dropwise to the acidic solution, keeping the temperature below 5°C. The reaction pH was then raised to 5 using potassium acetate buffer solution. Potassium xantogenate (6.32 g) was dissolved in water (8 mL) under nitrogen at room temperature. The cooled diazonium solution was added dropwise to this solution. The solution turned red and gas was released. The reaction mixture was stirred for an additional hour at room temperature. Dichloromethane was added and the mixture was acidified using hydrochloric acid 1N. The organic phase was separated, extracted with brine, filtered and dried under reduced pressure. The xantogenate product (3.16 g) was used without further purification. Product was dissolved in ethanol (22 mL) to give a red colored

solution. Potassium hydroxide (2.32 g, 3eq) was added with stirring to this solution. Stirring was continued for 5 hours. The reaction mixture was acidified using HCl 1N and extracted using ethyl acetate. This crude product was dissolved in acetic acid (6 mL) and Zinc powder (110 mg) was added. The mixture was heated to reflux for 4 hours. After cooling, the product precipitated and was filtered and washed with small portions of water. The crude product was dissolved in basic water and the zinc was filtered from the mixture. The product was extracted using diethylether. The organic phase was washed with brine and dried over magnesium sulfate. The solvent was removed under reduced pressure to yield **25**. Yellow powder. ^{1}H NMR (300 MHz, DMSO-d_6) 7.80 (d, J = 2.4 Hz, 1H), 7.00 (d, J = 8.6 Hz, 1H), 6.92 (d, J = 8.7 Hz, 0H). m/z = 170.2 Yield (three steps): 16%. Matches previous synthesis.[68]

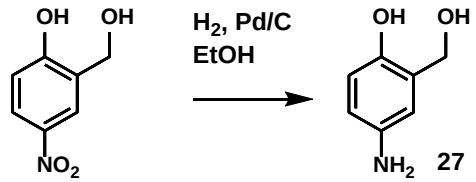


Under cooling (5-10 °C) a solution of 0.351 g (2.06 mmol) **25** in THF (30 mL) was slowly added to $LiAlH_4$ in THF (15 mL). The reaction was stirred at room temperature for 0.5 h until the evolution of H_2 had ceased and stirred at 60 °C for another 17 h. The reaction was quenched by addition of ice, the precipitate dissolved with 10% aqueous H_2SO_4 and the solution extracted with diethyl ether (3x). The combined organic extracts were extracted with 10% aqueous NaOH (2x) and the combined alkaline extracts neutralized with 10% hydrochloric acid. The precipitate was collected by filtration and the aqueous filtrate extracted with diethyl ether (3x). The combined organic extracts were washed with brine (2x) and dried over Na_2SO_4, and the solvent was evaporated. The product (**26**) is a

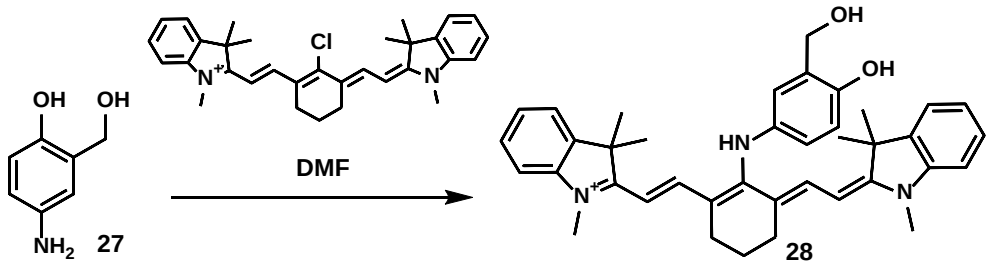
pale yellow solid. ^{1}H NMR (300 MHz, Chloroform-d) δ 7.11 – 6.99 (m, 2H), 6.83 (dd, J = 25.9, 2.4 Hz, 1H), 4.82 (d, J = 9.8 Hz, 2H). m/z = 156.2. Yield: 35%.



To a flask charged with 70.5 mg **9** (0.136 mmol), 3 mL DMF purged with argon was added and stirred under argon. 22 mg **26** (0.16 mmol) was added to the flask and monitored with analytical liquid chromatography-mass spectroscopy. Reaction proceeded for one hour in the dark. Following completion, ether was added to the reaction, precipitating **21**. The reaction was filtered off, and then **21** was dissolved in dichloromethane and extracted with brine. The green mixture was purified by silica gel chromatography with a mobile phase of MeOH:DCM 10:90. Green solid. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.77 (d, J = 14.2 Hz, 2H), 7.36 – 7.32 (m, 3H), 7.21 (td, J = 7.3, 0.9 Hz, 2H), 7.09 (d, J = 7.8 Hz, 2H), 7.06 – 7.02 (m, 1H), 6.94 (d, J = 2.5 Hz, 1H), 6.91 (d, J = 1.7 Hz, 1H), 6.04 (d, J = 14.1 Hz, 2H), 4.72 (s, 2H), 3.58 (s, 6H), 3.56 (s, 1H), 2.67 (t, J = 6.1 Hz, 4H), 1.97 (t, J = 6.1 Hz, 2H), 1.53 (s, 12H). m/z = 603.4. Yield: 72%.



Evacuate a flame-dried round bottom flask and backfill with argon five times. Weigh out 0.1 g of Pd/C and add to the flask quickly. Add DCM to wash, then purge/backfill. Add 5-nitro-2-methylhydroxyphenol (1.0 g, 7.2 mmol) dissolved with 75 mL ethanol. Subsequently, evacuate the flask and backfill with argon several times. Then, evacuate and attach a hydrogen balloon to fill the flask. Evacuate a second time, and refill with hydrogen three times. Stir at room temperature for 24 hours. Remove the balloon and fill with argon. Filter the reaction mixture through a bed of Celite, making sure that cake doesn't go to dryness. Evaporate the ethanol. The white solid is the product **27**. ^{1}H NMR (300 MHz, Acetonitrile-d_3) δ 6.58 (d, J = 8.4 Hz, 1H), 6.55 – 6.49 (m, 1H), 6.46 (dd, J = 8.4, 2.8 Hz, 1H), 4.57 (s, 2H), 3.74 (s, 2H). m/z = 139.2. Yield: 97%. Matches previous synthesis of this compound.[69]



To a flask charged with 70.5 mg **9** (0.136 mmol), 3 mL DMF purged with argon was added and stirred under argon. 22 mg **27** (0.16 mmol) was added to the flask and monitored with analytical liquid chromatography-mass spectroscopy. Reaction proceeded for one hour in the dark. Following completion, ether was added to the reaction, precipitating **28**. The reaction was filtered off, and then **28** was dissolved in

dichloromethane and extracted with brine. The blue mixture was purified by silica gel chromatography with a mobile phase of MeOH:DCM 10:90. Blue solid. ^{1}H NMR (300 MHz, Chloroform-*d*) δ 8.04 (d, *J* = 13.5 Hz, 2H), 7.25 – 7.19 (m, 3H), 7.08 (dd, *J* = 7.4, 0.9 Hz, 2H), 7.06 (d, *J* = 10.8 Hz, 1H), 6.89 (d, *J* = 8.3 Hz, 3H), 5.67 (d, *J* = 13.4 Hz, 2H), 4.77 (d, *J* = 13.7 Hz, 3H), 3.44 (s, 6H), 3.24 – 3.10 (m, 2H), 2.52 (t, *J* = 6.3 Hz, 4H), 1.89 (t, *J* = 6.3 Hz, 1H), 1.39 (s, 12H). m/z = 586.8. Yield: 72%.

Photophysics:

All spectra were taken in Omnisolv purity spectroscopy-grade solvents and Millipore pure water. Absorbance spectra were taken using an Agilent™ Technologies Cary 60 UV-Vis with a pulsed xenon lamp as the excitation source. The error in reported wavelength values is at most 0.5 nm. Fluorescence and excitation spectra were recorded on a Jobin Yvon Spex Fluorolog-3-11. Sample excitation was achieved via a xenon arc lamp with a monochromator providing wavelength selection. Right angle light emission was sorted using a monochromator and fed into a Hamamatsu R928P photomultiplier tube with photon counting. Short and long pass filters were used where appropriate. All emission and excitation spectra were taken as an average of two spectra subtracted from two blank spectra of the same excitation wavelength taken on the same day. Quantum yields of fluorescence were measured using Cy5 (φ_f=0.27) and *meso*-Chloro-Cy7 (φ_f=0.12) as references in acetonitrile. In addition to subtracting out background, anomalous features due to excitation and emission overtones were subtracted out of the spectra.

Kamlet-Taft analysis was performed by taking the absorbance and emission spectra of a dye at the same concentration (Absorbance of ~0.2) in a number of different solvents.

The absorbance and emission maxima were recorded, and then the parameters were solved using the following equation:

(2) $\Delta\lambda = \alpha_s \times \alpha_D + \beta_s \times \beta_D + \pi^*{}_s \times \pi^*{}_D$

Where Δλ refers to the wavelength of either absorption or emission relative to that of the dye in cyclohexane, α is the hydrogen bond accepting parameter, β is the hydrogen bond donating parameter, and π* is the parameter for the dipole moment. The subscript refers to if the parameter refers to the solvent (s) or dye (D). The resulting X_D parameters are in the units of nm.

Stern-Volmer quenching is carried out in a similar way to the literature. To a 0.2 absorbance sample of a dye, water is added and the fluorescence measured at various concentrations. Because near the end of the analysis, a significant volume of water was added, the fluorescence of the dye was corrected based on the changing absorbance after repeated additions of the quencher. The data were fit to equation **1** and analyzed as discussed above.

Photochemistry

Samples were irradiated with a 300 W Hg arc lamp with glass filter (>686 nm), or a M780L3 (785 nm, 200 mW) mounted LED purchased from Thor Labs. Photolysis was conducted in a custom made 3mL quartz cuvette with freeze/pump thaw capability. Oxygen exclusion was accomplished by freezing the sample in an attached round bottom flask to the cuvette in liquid nitrogen under vacuum, then thawing under vacuum, freezing, and thawing again. After 5 cycles, the system was backfilled with argon.

Quantum Mechanical Calculations:

For all calculations, the Gaussian 09 software pack was used. Visualization and generation of surfaces was performed using Gaussview 05.[70] Ground and excited-state energies were calculated using DFT//CAM-B3LYP/6-31G(d,p) and TDDFT//CAM-B3LYP/6-31G(d,p) respectively. Orbitals were visualized in Gaussview 05 from the DFT structures.[71]

6.8: References

1. Brudno, Y.; Silva, E. A.; Kearney, C. J.; Lewin, S. A.; Miller, A.; Martinick, K. D.; Aizenberg, M.; Mooney, D. J., Refilling drug delivery depots through the blood. *Proceedings of the National Academy of Sciences* **2014,** *111* (35), 12722-12727.
2. Levskaya, A.; Weiner, O. D.; Lim, W. A.; Voigt, C. A., Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **2009,** *461* (7266), 997-1001.
3. Bonifacino, J. S.; Traub, L. M., Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu Rev Biochem* **2003,** *72* (1), 395-447.
4. Dyrløv Bendtsen, J.; Nielsen, H.; von Heijne, G.; Brunak, S., Improved Prediction of Signal Peptides: SignalP 3.0. *Journal of Molecular Biology* **2004,** *340* (4), 783-795.
5. Nielsen, H.; Engelbrecht, J.; Brunak, S.; von Heijne, G., Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering, Design and Selection* **1997,** *10* (1), 1-6.
6. Bailon, P.; Won, C.-Y., PEG-modified biopharmaceuticals. *Expert Opinion on Drug Delivery* **2009,** *6* (1), 1-16.
7. Greenwald, R. B., PEG drugs: an overview. *Journal of Controlled Release* **2001,** *74* (1–3), 159-171.
8. Hatakeyama, H.; Akita, H.; Harashima, H., A multifunctional envelope type nano device (MEND) for gene delivery to tumours based on the EPR effect: A strategy for overcoming the PEG dilemma. *Advanced Drug Delivery Reviews* **2011,** *63* (3), 152-160.
9. Maruyama, K., Intracellular targeting delivery of liposomal drugs to solid tumors based on EPR effects. *Advanced Drug Delivery Reviews* **2011,** *63* (3), 161-169.

10. Finkelstein, E. A.; Corso, P. S.; Miller, T. R., *Incidence and Economic Burden of Injuries in the United States*. Oxford University Press: New York, NY, 2006.

11. Siegel, R.; Miller, K.; Jemal, A. *Cancer Facts and Figures: 2016*; American Cancer Society: Atlanta, GA, 2016.

12. Centers for Disease Control and Prevention (CDC), National Center for Injury Prevention and Control. Report to Congress on mild traumatic brain injury in the United States: steps to prevent a serious public health problem. Prevention, C. f. D. C. a., Ed. Atlanta, GA, 2003.

13. Faul, M.; Xu, L.; Wald, M.; Voronado, V., Traumatic brain injury in the United States: emergency department visits, hospitalizations, and deaths. Centers for Disease Control and Prevention, N. C. f. I. P. a. C., Ed. Atlanta, GA, 2010.

14. Nigrovic, L. E.; Schunk, J. E.; Foerster, A.; Cooper, A.; Miskin, M.; Atabaki, S. M.; Hoyle, J.; Dayan, P. S.; Holmes, J. F.; Kuppermann, N., The Effect of Observation on Cranial Computed Tomography Utilization for Children After Blunt Head Trauma. *Pediatrics* **2011**.

15. McKee, A. C.; Daneshvar, D. H.; Alvarez, V. E.; Stein, T. D., The neuropathology of sport. *Acta Neuropathologica* **2014,** *127* (1), 29-51.

16. Bernard, S. A.; Buist, M., Induced hypothermia in critical care medicine: A review. *Critical Care Medicine* **2003,** *31* (7), 2041-2051.

17. *Textbook of Traumatic Brain Injury*. American Psychiatric Publishing: Arlington, VA, 2005; Vol. 27-39.

18. Chesnut, R. M.; Marshall, L. F.; Klauber, M. R.; Blunt, B. A.; Baldwin, N.; Eisenberg, H. M.; Jane, J. A.; Marmarou, A.; Foulkes, M. A., The role of secondary brain injury in determining outcome from severe head injury. *J Trauma* **1993,** *34* (2), 216-222.

19. Shively, S. B.; Horkayne-Szakaly, I.; Jones, R. V.; Kelly, J. P.; Armstrong, R. C.; Perl, D. P., Characterisation of interface astroglial scarring in the human brain after blast exposure: a post-mortem case series. *The Lancet Neurology 15* (9), 944-953.

20. Staff, M. C. Diseases and Conditions: Post-concussion syndrome. http://www.mayoclinic.org/diseases-conditions/post-concussion-syndrome/basics/definition/con-20032705 (accessed 23 February).

21. Lulic, D.; Burns, J.; Bae, E. C.; van Loveren, H.; Borlongan, C. V., A Review of Laboratory and Clinical Data Supporting the Safety and Efficacy of Cyclosporin A in Traumatic Brain Injury. *Neurosurgery* **2011,** *68* (5), 1172-1186.

22. Bochet, C. G., Photolabile protecting groups and linkers. *Journal of the Chemical Society, Perkin Transactions 1* **2002,** (2), 125-142.

23. Falvey, D. E.; Sundararajan, C., Photoremovable protecting groups based on electron transfer chemistry. *Photochemical & Photobiological Sciences* **2004,** *3* (9), 831-838.

24. Regan, C. J.; Walton, D. P.; Shafaat, O. S.; Dougherty, D. A., Mechanistic Studies of the Photoinduced Quinone Trimethyl Lock Decaging Process. *Journal of the American Chemical Society* **2017,** *139* (13), 4729-4736.

25. Walton, D. P.; Dougherty, D. A., A General Strategy for Visible-Light Decaging Based on the Quinone Trimethyl Lock. *Journal of the American Chemical Society* **2017,** *139* (13), 4655-4658.

26. Givens, R. S.; Heger, D.; Hellrung, B.; Kamdzhilov, Y.; Mac, M.; Conrad, P. G.; Cope, E.; Lee, J. I.; Mata-Segreda, J. F.; Schowen, R. L.; Wirz, J., The photo-Favorskii reaction of p-hydroxyphenacyl compounds is initiated by water-assisted, adiabatic extrusion of a triplet biradical. *Journal of the American Chemical Society* **2008,** *130* (11), 3307-+.

27. Song, Y.; Tian, T.; Wang, P.; He, H. P.; Liu, W. L.; Zhou, X.; Cao, X. P.; Zhang, X. L.; Zhou, X., Phenol quaternary ammonium derivatives: charge and linker effect on their DNA photo-inducible cross-linking abilities. *Organic & Biomolecular Chemistry* **2006,** *4* (17), 3358-3366.

28. Basaric, N.; Doslic, N.; Ivkovic, J.; Wang, Y. H.; Malis, M.; Wan, P., Very Efficient Generation of Quinone Methides through Excited State Intramolecular Proton Transfer to a Carbon Atom. *Chemistry-a European Journal* **2012,** *18* (34), 10617-10623.

29. Wang, Y. H.; Wan, P., Excited state intramolecular proton transfer (ESIPT) in dihydroxyphenyl anthracenes. *Photochemical & Photobiological Sciences* **2011,** *10* (12), 1934-1944.

30. Percivalle, C.; La Rosa, A.; Verga, D.; Doria, F.; Mella, M.; Palumbo, M.; Di Antonio, M.; Freccero, M., Quinone Methide Generation via Photoinduced Electron Transfer. *Journal of Organic Chemistry* **2011,** *76* (9), 3096-3106.

31. Tolbert, L. M.; Solntsev, K. M., Excited-State Proton Transfer: From Constrained Systems to "Super" Photoacids to Superfast Proton Transfer†. *Accounts of Chemical Research* **2001,** *35* (1), 19-27.

32. Spry, D. B.; Goun, A.; Fayer, M. D., Deprotonation Dynamics and Stokes Shift of Pyranine (HPTS). *The Journal of Physical Chemistry A* **2006,** *111* (2), 230-237.

33. Pines, D.; Pines, E., Solvent Assisted Photoacidity. In *Hydrogen-Transfer Reactions*, Wiley-VCH Verlag GmbH & Co. KGaA: 2007; pp 377-415.

34. Agmon, N., Elementary Steps in Excited-State Proton Transfer†. *The Journal of Physical Chemistry A* **2004,** *109* (1), 13-35.

35. Blanksby, S. J.; Ellison, G. B., Bond Dissociation Energies of Organic Molecules. *Accounts of Chemical Research* **2003,** *36* (4), 255-263.

36. Sun, W.; Guo, S.; Hu, C.; Fan, J.; Peng, X., Recent Development of Chemosensors Based on Cyanine Platforms. *Chemical Reviews* **2016,** *116* (14), 7768-7817.

37. Sarder, P.; Maji, D.; Achilefu, S., Molecular Probes for Fluorescence Lifetime Imaging. *Bioconjugate Chemistry* **2015,** *26* (6), 963-974.

38. Johannes, H. H.; Grahn, W.; Reisner, A.; Jones, P. G., Ethynylated, Vinylated, and Hetarylated Indodicarbocyanines by Palladium-Catalyzed Cross-Coupling Reactions. *Tetrahedron Lett* **1995,** *36* (40), 7225-7228.

39. Berlier, J. E.; Rothe, A.; Buller, G.; Bradford, J.; Gray, D. R.; Filanoski, B. J.; Telford, W. G.; Yue, S.; Liu, J.; Cheung, C. Y.; Chang, W.; Hirsch, J. D.; Beechem Rosaria P. Haugland, J. M.; Haugland, R. P., Quantitative Comparison of Long-wavelength Alexa Fluor Dyes to Cy Dyes: Fluorescence of the Dyes and Their Bioconjugates. *Journal of Histochemistry & Cytochemistry* **2003,** *51* (12), 1699-1712.

40. Wan, P.; Shukla, D., Utility of acid-base behavior of excited states of organic molecules. *Chemical Reviews* **1993,** *93* (1), 571-584.

41. Solntsev, K. M.; Huppert, D.; Agmon, N., Solvatochromism of beta-naphthol. *Journal of Physical Chemistry A* **1998,** *102* (47), 9599-9606.

42. Taft, R. W.; Kamlet, M. J., SOLVATOCHROMIC COMPARISON METHOD .2. ALPHA-SCALE OF SOLVENT HYDROGEN-BOND DONOR (HBD) ACIDITIES. *Journal of the American Chemical Society* **1976,** *98* (10), 2886-2894.

43. Kamlet, M. J.; Taft, R. W., SOLVATOCHROMIC COMPARISON METHOD .1. BETA-SCALE OF SOLVENT HYDROGEN-BOND ACCEPTOR (HBA) BASICITIES. *Journal of the American Chemical Society* **1976,** *98* (2), 377-383.

44. Kamlet, M. J.; Abboud, J. L. M.; Abraham, M. H.; Taft, R. W., LINEAR SOLVATION ENERGY RELATIONSHIPS .23. A COMPREHENSIVE COLLECTION OF THE SOLVATOCHROMIC PARAMETERS, PI-STAR, ALPHA AND BETA, AND SOME METHODS FOR SIMPLIFYING THE GENERALIZED SOLVATOCHROMIC EQUATION. *Journal of Organic Chemistry* **1983,** *48* (17), 2877-2887.

45. Kamlet, M. J.; Abboud, J. L.; Taft, R. W., SOLVATOCHROMIC COMPARISON METHOD .6. PI-STAR SCALE OF SOLVENT POLARITIES. *Journal of the American Chemical Society* **1977,** *99* (18), 6027-6038.

46. Catalan, J.; Lopez, V.; Perez, P.; Martinvillamil, R.; Rodriguez, J. G., PROGRESS TOWARDS A GENERALIZED SOLVENT POLARITY SCALE - THE SOLVATOCHROMISM OF 2-(DIMETHYLAMINO)-7-NITROFLUORENE AND ITS HOMOMORPH 2-FLUORO-7-NITROFLUORENE. *Liebigs Annalen* **1995,** (2), 241-252.

47. Kim, J. M.; Chang, T. E.; Kang, J. H.; Park, K. H.; Han, D. K.; Ahn, K. D., Photoacid-induced fluorescence quenching: A new strategy for fluorescent imaging in polymer films. *Angewandte Chemie-International Edition* **2000,** *39* (10), 1780-+.

48. Solntsev, K. M.; Huppert, D.; Agmon, N.; Tolbert, L. M., Photochemistry of "super" photoacids. 2. Excited-state proton transfer in methanol/water mixtures. *Journal of Physical Chemistry A* **2000,** *104* (19), 4658-4669.

49. Pohlers, G.; Scaiano, J. C.; Sinta, R., A novel photometric method for the determination of photoacid generation efficiencies using benzothiazole and xanthene dyes as acid sensors. *Chemistry of Materials* **1997,** *9* (12), 3222-3230.

50. Htun, T., Transient effects in the fluorescence quenching of a strong photoacid. *Journal of Luminescence* **2013,** *144*, 122-127.

51. Samanta, A.; Vendrell, M.; Das, R.; Chang, Y.-T., Development of photostable near-infrared cyanine dyes. *Chemical Communications* **2010,** *46* (39), 7406-7408.

52. Gessner, T.; Mayer, U., Triarylmethane and Diarylmethane Dyes. In *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA: 2000.

53. Nakatsuji, S. i.; Nakazumi, H.; Nakashima, K.; Akiyama, S., Synthesis and Properties of Triphenylmethane Dye Ethynologs. *NIPPON KAGAKU KAISHI* **1992,** *1992* (11), 1279-1293.

54. Shuzo, A.; Keiko, Y.; Misuzu, H.; Kenichiro, N.; Shin'ichi, N.; Masahiko, I., ETHYNOLOGS OF TRIPHENYLMETHANE DYES. SYNTHESES AND PROPERTIES OF ACETYLENIC ANALOGS OF MALACHITE GREEN, CRYSTAL VIOLET, AND THEIR RELATED COMPOUNDS. *Chemistry Letters* **1981,** *10* (3), 311-314.

55. Shin'ichi, N.; Naoko, O.; Kenichiro, N.; Shuzo, A., SYNTHESIS AND ELECTRONIC ABSORPTION SPECTRA OF MONOETHYNOLOGS OF MALACHITE GREEN WITH p-SUBSTITUENT ON PHENYL RING. *Chemistry Letters* **1986,** *15* (3), 329-332.

56. Di Antonio, M.; Doria, F.; Mella, M.; Merli, D.; Profumo, A.; Freccero, M., Novel Naphthalene Diimides as Activatable Precursors of Bisalkylating Agents, by Reduction and Base Catalysis. *The Journal of Organic Chemistry* **2007,** *72* (22), 8354-8360.

57. Song, F.; Peng, X.; Lu, E.; Wang, Y.; Zhou, W.; Fan, J., Tuning the photoinduced electron transfer in near-infrared heptamethine cyanine dyes. *Tetrahedron Lett* **2005,** *46* (28), 4817-4820.

58. Peng, X.; Song, F.; Lu, E.; Wang, Y.; Zhou, W.; Fan, J.; Gao, Y., Heptamethine Cyanine Dyes with a Large Stokes Shift and Strong Fluorescence: A Paradigm for Excited-State Intramolecular Charge Transfer. *Journal of the American Chemical Society* **2005,** *127* (12), 4170-4171.

59. Gospodova, T.; Rashkova, J.; Ivanova, D.; Viteva, L.; Duprat, C.; Mazières, M.-R.; Bakalova, S.; Kaneti, J., Synthetic Potentials of Heptamethine Merocyanine Dyes Containing an Active Chlorine Atom: Reactivity towards Nucleophiles. *European Journal of Organic Chemistry* **2009,** *2009* (29), 5063-5071.

60. Lee, H.; Berezin, M. Y.; Henary, M.; Strekowski, L.; Achilefu, S., Fluorescence lifetime properties of near-infrared cyanine dyes in relation to their structures. *Journal of photochemistry and photobiology. A, Chemistry* **2008,** *200* (2-3), 438-444.

61. Ault, A., General Acid and General Base Catalysis. *Journal of Chemical Education* **2007,** *84* (1), 38.

62. Kiefer, P. M.; Hynes, J. T., Adiabatic and nonadiabatic proton transfer rate constants in solution. *Solid State Ionics* **2004,** *168* (3), 219-224.

63. Gorka, A. P.; Nani, R. R.; Zhu, J.; Mackem, S.; Schnermann, M. J., A Near-IR Uncaging Strategy Based on Cyanine Photochemistry. *Journal of the American Chemical Society* **2014,** *136* (40), 14153-14159.

64. Still, W. C.; Kahn, M.; Mitra, A., Rapid chromatographic technique for preparative separations with moderate resolution. *The Journal of Organic Chemistry* **1978,** *43* (14), 2923-2925.

65. Kitschke, P.; Walter, M.; Ruffer, T.; Seifert, A.; Speck, F.; Seyller, T.; Spange, S.; Lang, H.; Auer, A. A.; Kovalenko, M. V.; Mehring, M., Porous Ge@C materials via twin polymerization of germanium(ii) salicyl alcoholates for Li-ion batteries. *Journal of Materials Chemistry A* **2016,** *4* (7), 2705-2719.

66. Gisch, N.; Balzarini, J.; Meier, C., Enzymatically Activated cycloSal-d4T-monophosphates: The Third Generation of cycloSal-Pronucleotides. *Journal of Medicinal Chemistry* **2007,** *50* (7), 1658-1667.

67. Doria, F.; Gallati, C. M.; Freccero, M., Hydrosoluble and solvatochromic naphthalene diimides with NIR absorption. *Organic & Biomolecular Chemistry* **2013,** *11* (45), 7838-7842.

68. Saphier, S.; Haft, A.; Margel, S., Bacterial Reduction as Means for Colonic Drug Delivery: Can Other Chemical Groups Provide an Alternative to the Azo Bond? *Journal of Medicinal Chemistry* **2012,** *55* (23), 10781-10785.

69. Mudarisova, R. K.; Badykova, L. A.; Borisov, I. M.; Fatykhov, A. A.; Medvedeva, E. N.; Babkin, V. A.; Monakov, Y. B., The interaction of poly- and oligosaccharides based on arabinogalactan with 5-aminosalicylic acid. *Russian Journal of Physical Chemistry A, Focus on Chemistry* **2008,** *82* (8), 1393-1396.

70. 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.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd,

J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

71. Roy Dennington, T. K., and John Millam, Gaussview, Version 5. Semichem, Inc: Shawnee Mission, KS, 2009.

Appendix A: Synthesis and characterization of γ-hydroxy-Cy5

A.1: Introduction

As discussed at length in Chapter 6, a major goal of this thesis was to develop a photochemically active cyanine dye. Chapter 6 detailed several putative photocages which, with the exception of the ethynyl-triphenylmethane cation dyes, progressed to a photochemistry step. One early design of a photoacidic cyanine dye was a *meso*-hydroxy-Cy5 dye. Because the dye is on the γ carbon of the polymethine chain, it was named γ(OH)Cy5. The reasoning behind the synthesis of such a dye is that, due to the known large intramolecular charge transfer of cyanine dyes, placement of a hydroxyl group on the polymethine chain could result in a photoacidic dye.

Photoacidity, as discussed in Chapter 6, is a photophysical phenomenon in which a chromophore has a lower pKa in the excited state than the ground state.[1-2] Naphthol is the first widely studied photoacid. In its ground state, its hydroxyl group has a pKa similar to phenols: 9.5. However, in its excited state, its pKa has been measured to be 2.8.[3-4] This large shift in pKa is due to a large intramolecular charge transfer event in its excited state. Determination of the excited state pKa (pKa*) can be done in several ways. The first is a Förster analysis: the pKa* can be calculated from the absorbance and emission maxima of the protonated and deprotonated states:

(1) $$pKa^* = pKa - {}^{(h\upsilon_{AH} - h\upsilon_{A^-})}/_{2.3RT}$$

where $h\nu_{AH}$ is the energy of the ground state acid measured by the absorbance of the protonated state and $h\nu_{A^-}$ is the energy of the excited state base measured by the emission of the deprotonated state. This excited state pKa measurement is an

approximation, and represents the theoretical maximum shift in pKa upon excitation. In addition to a photophysical measurement at different pH values, Kamlet-Taft analyses has been used as partial evidence for photoacidity.[5-6]

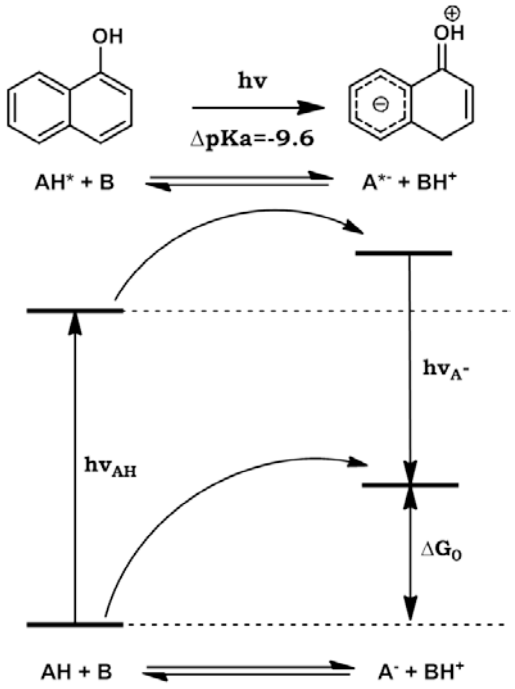


Figure 1: Photoacidity of 1-Napthol. The deprotonated form emission is significantly red-shifted, proportional to the pKa*

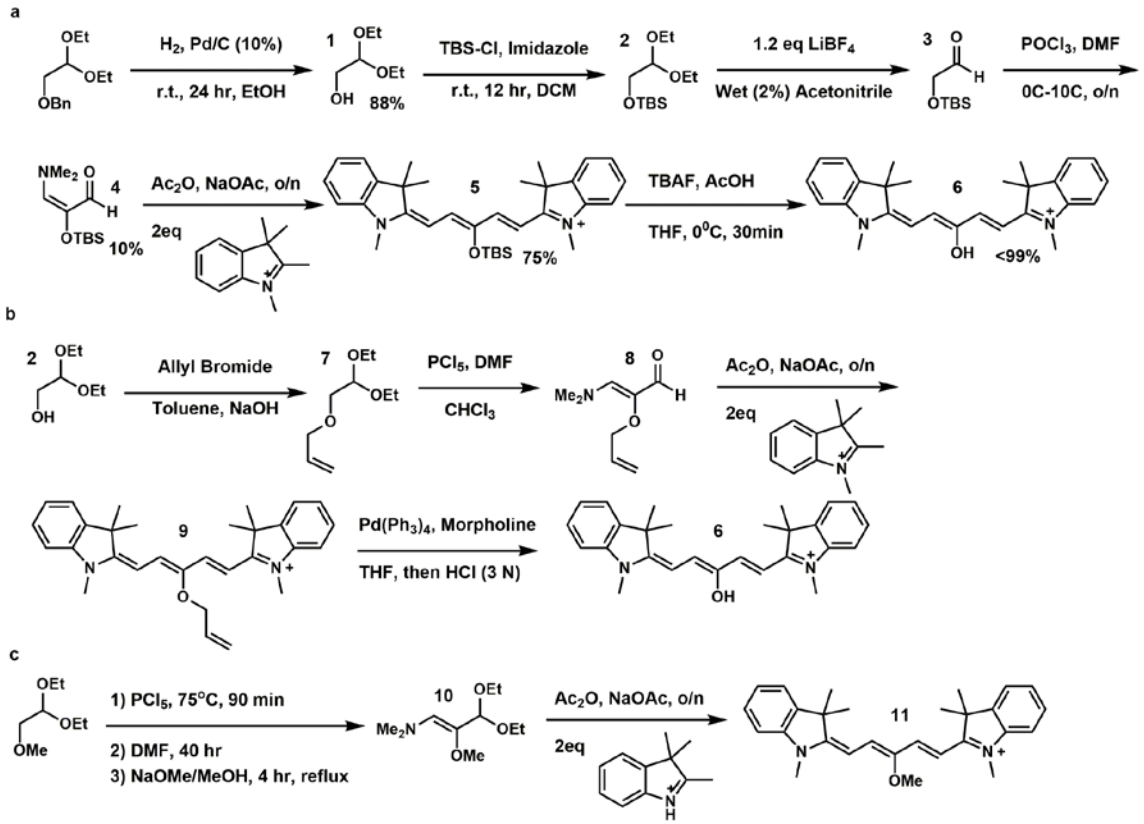
This technique, as discussed previously, utilizes absorbance and emission data in a variety of solvents. Then, the spectroscopic shifts of the dye in different solvents are correlated to three solvent parameters: α; the ability of a solvent to donate hydrogen bonds; β, the solvent's ability to accept hydrogen bonds, and π*, the solvent's plolarizability/dipolarity.[7-11] Photoacids were observed to have a much higher excited state dipole moment and a greatly increased ability to donate hydrogen bonds.

However, dyes of the type synthesized in this appendix are unique in that they overcome the difficulty of substitution of functional groups on the *meso* position of Cy5 dyes. While both Cy5 and Cy7 dyes can be synthesized with a *meso*-halogen group, the Cy7 dye (due to its resonance form) can perform a nucleophilic substitution reaction with soft nucleophiles, whereas Cy5 cannot. The use of C-C bond coupling reactions has precedent[12] (and has been utilized in Chapter 5), but inclusion of heteroatoms had little precedent. This appendix describes a general synthetic strategy to include heteroatoms at this position.

A.2: Synthesis and characterization of *meso*-substituted Cy5 dyes

A.2.1: Synthesis of relevant Cy5 dyes

Unlike Cy7 dyes, the heteroatoms of Cy5 dyes must be incorporated at an earlier stage. In order to do that, protected malondialdehyde derivatives were synthesized and immediately taken forward to the Cy5 dyes. Because the malondialdehyde group is so reactive, it could not be purified by column chromatography so was taken directly on to dye formation reactions in all cases. The desired dye **6** could be generated in two ways. The first (Scheme 1a) was via TBS protection of an aldehyde synthon (**1**) to yield **2** and after deprotection of the *gem*-diether formylation of the protected aldehyde **3** to yield **4** followed by dye condensation. The TBS group could then be removed from **5** to yield the desired dye. Alternatively (Scheme 1b), an allyl protecting group scheme could be used,



Scheme 1: Synthesis of *meso*-substituted Cy5 dyes. Two separate syntheses of **6** involve protection with either a (a) TBS or (b) allyl group. (c) Synthesis of a methylated version **11** is much simpler.

with the final deprotection step utilizing $Pd(PPh_3)_4$ to yield the same dye with fewer steps. Each strategy utilized protection of an aldehyde before formylation to yield the malondialdehyde synthon. Synthesis of a methylated dye **11** was much simpler (Scheme 1c) and followed many of the same steps previously outlined.

A.2.2: Photophysical studies of γ(OH)Cy5

What follows are several photophysical experiments to determine if **6** is a photoacid. **11** forms a valuable point of comparison. First, the absorbance and emission of **6** was measured in a variety of pH buffers (Figure 2). Unfortunately, the absorbance or emission did not change when in basic buffer. This would suggest that the deprotonated form of **6** was not achieved in either the ground or excited state at pH 11.59. However, there is precedent to a photoacid not exhibiting a changed absorbance or emission. In order to test for this, the photophysics of **6**, **11**, and Cy5 were measured in a number of solvents and a Kamlet-Taft analysis performed. This analysis, as for other cyanine dyes tested in this work (Chapter 6), does not suggest that **6** is a photoacid, and the correlation (R) values for the four-variable regression are relatively low.

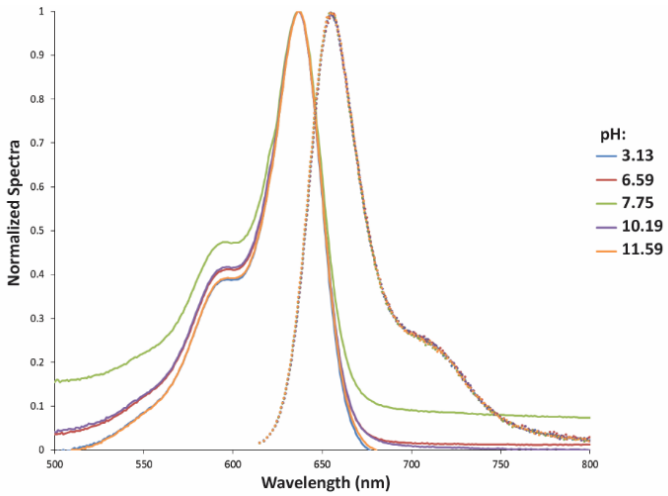


Figure 2: Normalized absorbance (solid lines) and emission (dotted lines) of **6** in various pH buffers. No additional absorbance or fluorescence bands were observed that corresponded to the deprotonated state. Emission was observed out to 1000 nm.

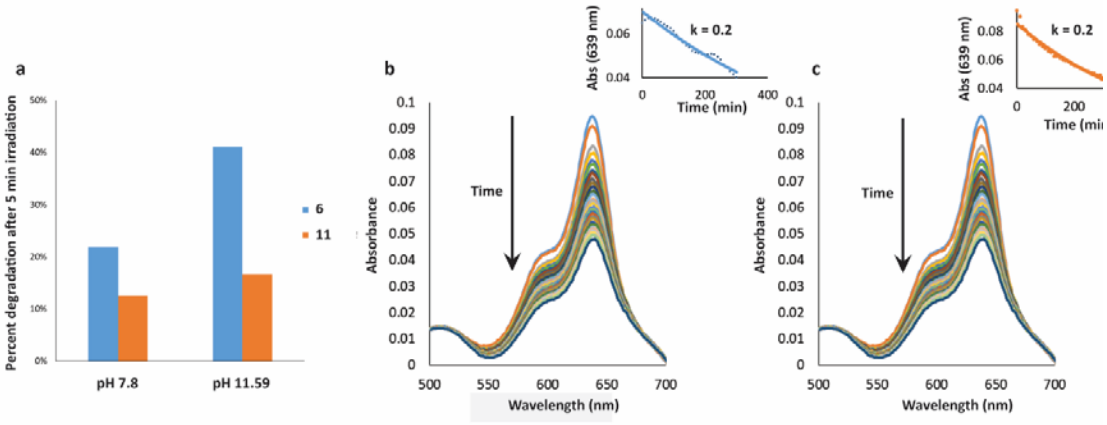


Figure 4: (a) pH-dependent degradation of **6** and **11** in buffered solutions of the indicated pH values. At basic pH, the (b) **6** and (c) **11** degraded at the same rate in the dark (inset), where k is the exponent of the decay function.

6	π*	10.89	0.63	10.85	0.76	-0.04
	α	-4.72	0.39	-2.25	0.20	2.47
	β	-3.97	0.07	-2.3	0.054	1.68
			0.77[1]		0.64[1]	
11	π*	1.33	0.026	2.6	0.17	1.27
	α	-9.07	0.68	-6.92	0.64	2.15
			0.69[1]		0.67[1]	
Cy5	π*	5.2	0.43	7.9	0.53	2.7
	α	-5.3	0.55	-5.3	0.45	-0.1
			0.71[1]		0.71[1]	

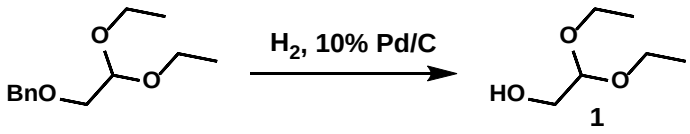
[1]R value from multivariate Kamlet-Taft regression analysis.

Irradiation of **6** and **11** were performed under both oxygen free and air-saturated conditions. Oxygen was excluded in similar ways as outlined in Chapter 6. Naphthol and other aromatic phenol-containing molecules are prone to oxidative degradation in the excited state to a much greater extent.[13] Upon irradiation with 600-650 nm light, **6** degrades significantly more rapidly in base than in neutral solution, where **11** degrades a similar rate at neutral and basic pH. At pH 11.59, **6** degrades at roughly the same rate as at pH 7.8 in the dark (Figure 12b). **11** degrades at a similar rate in the dark (Figure 12c). This suggests that the hydroxycyanine dye is unstable upon excitation, and moreso in basic solution. This suggests that the pH dependence of the degradation is tied to the presence of a deprotonated

hydroxyl group. In oxygen-free solutions, degradation of γ(OH)Cy5 and γ(OMe)Cy5 proceeds much more slowly, as before.

A.3: Methods

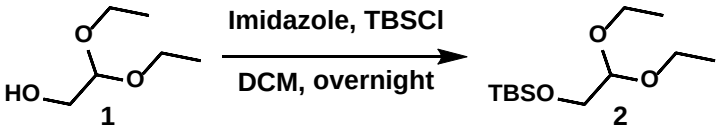
Unless otherwise stated, reactions were carried out under argon. Commercially available reagents were obtained from Sigma Aldrich, AK Scientific, Alfa Aesar, or Acros Organics and used without further purification. Solvents were used as received. Thin-layer chromatography with Sigma Aldrich silica gel coated plates with fluorescent indicator (0.25 mm) was used to monitor reactions. Silica gel chromatography was conducted as described by Still et al.,[14] with silica gel purchased from Alfa Aesar (60 Å, 230-400 mesh). NMR spectra were recorded on Varian (300, 400, 500, or 600 MHz) spectrometers. Mass spectra (ESI) were obtained with an Agilent liquid chromatography-mass spectrometry system. Isolated yields are reported.



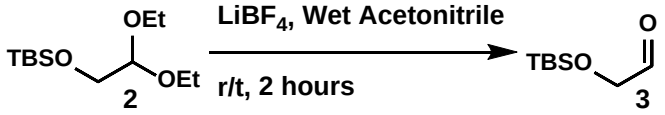
1

Evacuate a 250 mL reaction vessel and backfill with Argon. Weigh 200 mg of Pd/C out on clean, dry weigh paper and transfer to the reaction vessel. Add DCM to wash, then purge and backfill. Add 150 mL of dry ethanol to the reaction vessel via a canula. Then add 2.0 g (8.92 mmol) of benzyloxyacetylaldehyde diethyl acetal with a syringe. Purge and backfill the vessel with argon, then purge and backfill with hydrogen balloon three times. Stir vigorously for 24 hours. After completion, detach balloon and fill argon. Filter reaction mixture through a bed of Celite. Ethanol was removed under vacuum and taken onwards to the next step. 1.1885 g (8.86 mmol) of **1** obtained (99.3% yield). ^{1}H NMR

(300 MHz, Chloroform-d) δ 4.51 (t, J = 5.4 Hz, 1H), 3.81 – 3.37 (m, 4H), 2.17 (s, 2H), 1.19 (td, J = 7.1, 0.8 Hz, 6H). Yield: <99%.

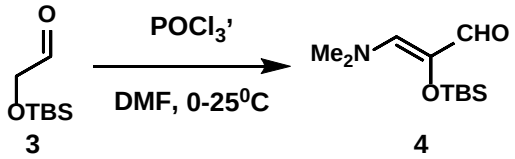


Flame dry a 125 mL round bottom flask. Add to that 1.1885 g (8.86 mmol) **1**, diluted in 80 mL dry dichloromethane. 2.01 g (13.29 mmol) *tert*-butyl dimethylsilyl chloride and 0.904 g (13.29 mmol) of Imidazole were added and the reaction was stirred at room temperature overnight. Reaction was worked up by washing with water 3 times, then with brine. Organic fractions were dried with magnesium sulfate and solvent removed under vacuum. The product **2** was taken directly on to the next step. 1H NMR (300 MHz, Chloroform-d) δ 4.48 (t, J = 5.3 Hz, 1H), 3.79 – 3.65 (m, 2H), 3.64 – 3.60 (m, 2H), 3.59 – 3.49 (m, 2H), 1.21 (tt, J = 7.1, 0.4 Hz, 6H), 0.89 (t, J = 0.4 Hz, 9H), 0.08 – 0.05 (m, 6H). Yield: 88%.



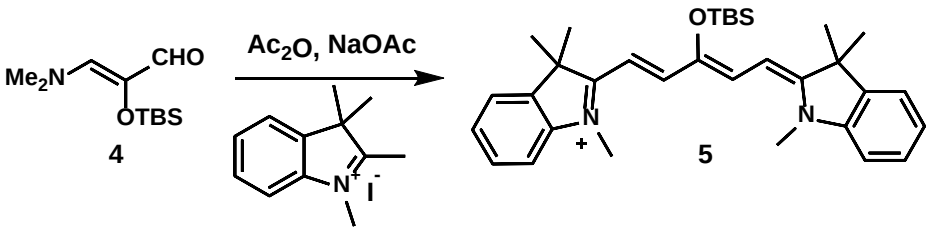
Add 2 mL wet acetonitrile (2% water) and 37.6 mg (0.482 mmol) lithium tetrafluoroborate to a flame dried three-nicked 10 mL round bottom flask under argon. Add 0.1 g (0.402 mmol) **2** dropwise while keeping the reaction mixture on ice, being careful to maintain the temperature below 10^0C in argon. Stir at room temperature for two hours. The reaction was quenched with saturated $NaHCO_3$, then the organic phase was extracted with ether three times. The ether was concentrated under vacuum, and the residue added to a silica gel column. The eluent was MeOH/DCM 3:97 to give **3** as a yellow oil. ^{1}H NMR

(400 MHz, Chloroform-d) δ 9.68 (q, J = 0.8 Hz, 0H), 5.29 (d, J = 0.8 Hz, 0H), 1.73 (s, 4H), 0.93 – 0.86 (m, 9H), 0.13 – 0.02 (m, 6H). Yield: 70%.



3 **4**

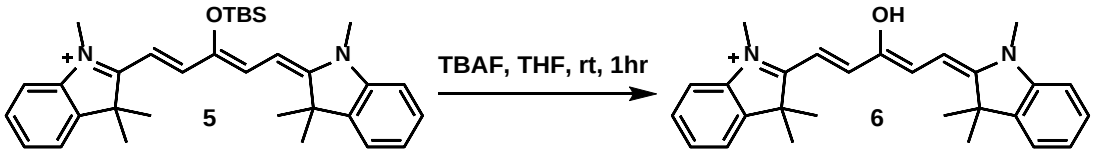
DMF was added to a flame-dried 25 mL round bottom flask. To this, 215 μL (1.37 mmol) phosphoryl chloride was added and reacted with solvent reagent for fifteen minutes on ice. **3** (0.2 mL, 1.14 mmol) was added dropwise and kept on ice for 30 min. The reaction mixture was then slowly brought to room temperature and stirred overnight. The reaction was quenched with aqueous sodium bicarbonate on ice. The reaction was diluted with aqueous sodium bicarbonate until the mixture is less than 10% DMF, then was extracted three times with diethyl ether. The organic layers were collected and dried with magnesium sulfate. The ether was concentrated under vacuum. Because the product is unstable on silica gel, **4** was taken directly on to the next step. Crude ^{1}H NMR (400 MHz, DMSO-d6) δ 9.61 (s, 1H), 5.26 (s, 1H), 1.35 (d, J = 1.1 Hz, 6H), 0.84 (d, J = 1.0 Hz, 9H), -0.05 (d, J = 1.0 Hz, 6H).



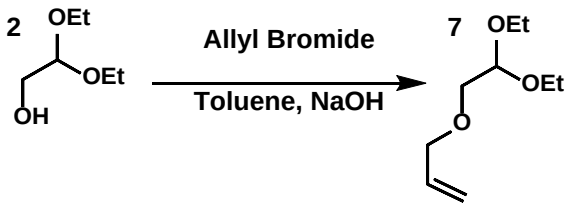
4 **5**

100 mg (0.33 mmol) 1,2,2,3-tetramethylindolenium and **5** (34.4 mg, 0.15 mmol) were dissolved in acetic anhydride in a flame-dried 10mL round bottom flask. To this mixture, 50 mg (0.61 mmol) sodium acetate was added. The reaction was stirred at room temperature overnight. After completion, add brine and quench with dichloromethane

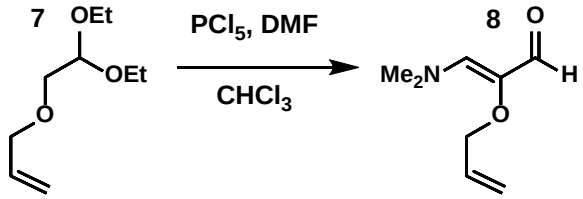
three times. The organic fractions were collected, dried with magnesium sulfate, and dried under vacuum. The crude oil was purified by silica gel chromatography. The mobile phase is MeOH:DCM 2:98. ^{1}H NMR (400 MHz, Chloroform-d) δ 8.97 (d, J = 13.4 Hz, 2H), 8.54 (d, J = 13.9 Hz, 2H), 7.38 (q, J = 9.0, 8.3 Hz, 4H), 7.16 – 7.03 (m, 4H), 6.85 (d, J = 8.2 Hz, 2H), 6.31 (d, J = 13.2 Hz, 2H), 3.79 (s, 6H), 1.89 (s, 12H), 1.25 (s, 9H), 0.07 (s, 6H). m/z = 439.0. Yield: (two steps) 5%



Add dry THF to a dry flask charged with **6** (31.8 mg, 0.062 mmol). Add 1.0 M tetrabutylammonium fluoride solution in THF (93 μL, 0.093 mmol) along with 100 uL acetic acid and let stir at room temperature for 1 hour. Add saturated aqueous ammonium chloride to quench. Extract with Ethyl Acetate and wash with 1:1 deionized water:saturated ammonium chloride. Collect organic layers, dry with magnesium sulfate, concentrate, and dry under vacuum. Purified on silica chromatography with the mobile phase of methanol:dichloromethane 2:98. This was then further purified by high-performance liquid chromatography (HPLC). ^{1}H NMR (400 MHz, DMSO-d6) δ 8.44 (d, J = 13.3 Hz, 2H), 7.68 (d, J = 7.6 Hz, 2H), 7.51 (d, J = 8.0 Hz, 2H), 7.49 – 7.40 (m, 2H), 7.35 – 7.29 (m, 2H), 7.13 – 6.43 (br s, 1H), 6.29 (d, J = 13.3 Hz, 3H), 3.68 (s, 6H), 1.71 (s, 12H). m/z = 399.4. Yield: 95%.

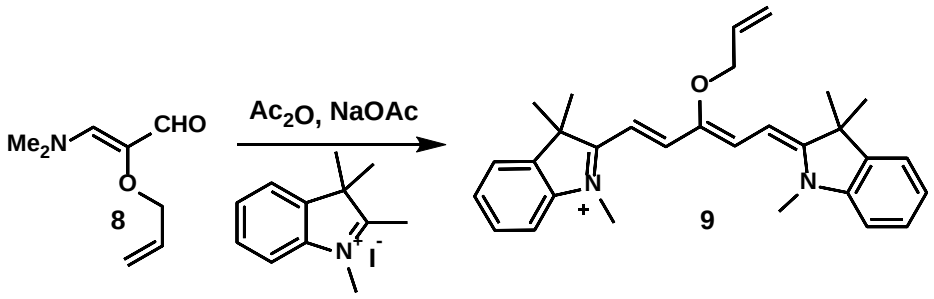


A mixture of **2** (1.076g, 8.02 mmol), powdered sodium hydroxide (0.35 g, 8 mmol), dry toluene, and allyl bromide (0.867 mL, 10 mmol) was heated under reflux and the progress of the reaction was followed by TLC. After 2 hours, the reaction was quenched with water and extracted 3x with toluene. The organic layers were collected and dried with magnesium sulfate. The solvent was evaporated and put under a high vacuum overnight. The resulting compound (**7**) was a pale yellow liquid. ^{1}H NMR (500 MHz, Chloroform-d) δ 5.96 – 5.82 (m, 1H), 5.28 (dq, J = 17.2, 1.6 Hz, 1H), 5.22 – 5.14 (m, 1H), 4.65 (t, J = 5.3 Hz, 1H), 4.56 (t, J = 5.4 Hz, 2H), 4.05 (dt, J = 5.7, 1.4 Hz, 2H), 3.73 (t, J = 2.4 Hz, 1H), 3.58 – 3.56 (m, 4H), 3.50 (d, J = 5.2 Hz, 2H), 1.84 (t, J = 6.6 Hz, 1H), 1.25 – 1.21 (m, 5H). Yield: 90%.

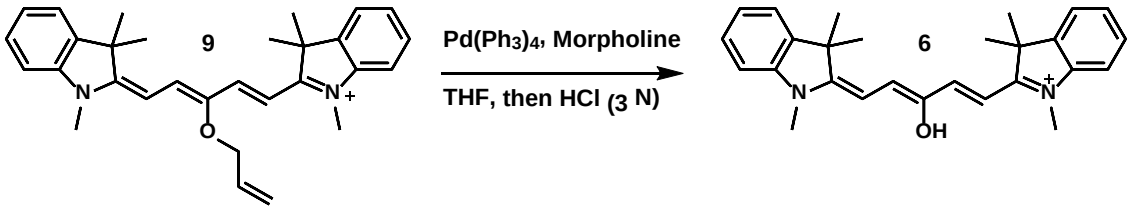


To 1.7 mL (22 mmol) DMF add 2.03 mL phosphorus pentachloride (22 mmol), maintaining the temperature at 25^{0}C by an external water bath. Upon completion of the addition, the resultant amber oil was warmed to 50^{0}C for 45 min and then cooled to 25^{0}C to form the Vilsmeir-Haack reagent. Chloroform (75 mL) was added and the resultant solution was brought to reflux. While reflux continued, **7** (1.26 g, 7.27 mmol) was added over a 30-min period. After a 3-h reflux following the addition, the cooled chloroform solution was cautiously added to a stirred mixture of 200g of K_2CO_3, 200 mL of ice water,

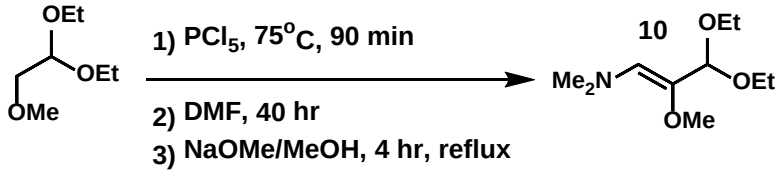
and 200 mL of a 9:l benzene:ethanol solution. The organic layer was separated, concentrated to a dark oil, and triturated with chloroform to remove inorganic salts. The chloroform was removed by rotary evaporator to give a dark red oil (**8**). The product was taken directly on to the dye forming reaction.



1.5 g (10 mmol) 1,2,2,3-tetramethylindolenium and **8** (0.766 g, 4.98 mmol) were dissolved in acetic anhydride in a flame-dried 50 mL round bottom flask. To this mixture, 0.5 g (20 mmol) sodium acetate was added. The reaction was stirred at room temperature overnight. After completion, add brine and quench with dichloromethane three times. The organic fractions were collected, dried with magnesium sulfate, and dried under vacuum. The crude oil was purified by silica gel chromatography. The mobile phase is MeOH:DCM 2:98. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.47 (d, J = 13.9 Hz, 2H), 7.55 – 7.48 (m, 2H), 7.35 – 7.32 (m, 3H), 7.25 – 7.16 (m, 3H), 7.08 (t, J = 9.0 Hz, 4H), 6.06 (d, J = 13.9 Hz, 2H), 5.56 (dd, J = 17.1, 1.7 Hz, 0H), 5.35 (dd, J = 10.5, 1.5 Hz, 0H), 4.96 (s, 2H), 4.43 (d, J = 5.3 Hz, 1H), 3.56 (s, 6H), 1.81 (s, 12H). m/z = 439.4. Yield (two steps): 5%.

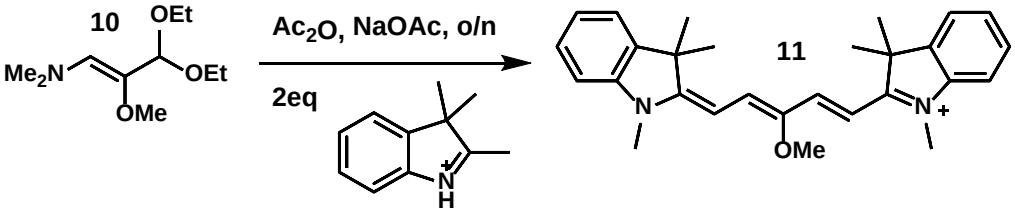


To a solution of **9** (102.9 mg, 0.234 mmol) in THF (5 mL) was added morpholine (50 µL, 0.3 mmol), sodium borohydride (13.28 mg, 3 mmol) and $Pd(PPh_3)_4$ (27 mg, 0.023 mmol) at 24 °C. The reaction was monitored by LC-MS. After 2 h, 3 N HCl (4 mL) was added very slowly to quench the reaction. The mixture was extracted with ethyl acetate, dried over Na_2SO_4, filtered, and concentrated under reduced pressure. **6** was purified by silica gel chromatography, with a mobile phase of 98:2 dichloromethane:methanol. ^{1}H NMR (400 MHz, DMSO-d6) δ 8.44 (d, J = 13.3 Hz, 2H), 7.68 (d, J = 7.6 Hz, 2H), 7.51 (d, J = 8.0 Hz, 2H), 7.49 – 7.40 (m, 2H), 7.35 – 7.29 (m, 2H), 7.13 – 6.43 (br s, 1H), 6.29 (d, J = 13.3 Hz, 3H), 3.68 (s, 6H), 1.71 (s, 12H). m/z = 399.4. Yield: 90%.



In a 3-neck flask (50 mL) equipped with an internal thermometer, add $POCl_3$ (0.84 mL, 8.61 mmol) in small portions to a cold (10^0C) solution of 1,1,2-trimethoxyethane (1.0 mL, 5.74 mmol) while keeping the internal temperature below 30^0C. Heat the mixture at 60^0C for 75 minutes and then place in an ice bath. Add DMF (1.33 mL, 17.22 mmol) via a dropping funnel while maintaining the internal temperature below a temperature of 10^0C. Stir the mixture at room temperature for 40 hours, and add MeOH (10 mL) dropwise via a dropping funnel while keeping the internal temperature below 10^0C. Transfer to an addition funnel and add dropwise to a 30% solution of sodium methoxide (40 mL) in MeOH while

maintaining the internal temperature below 20^0C. Reflux for 4hr then concentrate. Take the residue up in water (50 mL) and extract with dichloromethane (50 mL x3). The combined organics were dried, filtered and concentrated to yield a crude mixture of **10**. The reaction was taken directly on to the dye forming reaction. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.59 (s, 1H), 6.15 (s, 1H), 3.69 (s, 3H), 3.15 (s, 6H). Yield: 25%.



949 mg (3.15 mmol) 1,2,2,3-tetramethylindolenium and **8** (185 mg, 1.43 mmol) were dissolved in acetic anhydride in a flame-dried 50 mL round bottom flask. To this mixture, 47 mg (5.7 mmol) sodium acetate was added. The reaction was stirred at room temperature overnight. After completion, add brine and quench with dichloromethane three times. The organic fractions were collected, dried with magnesium sulfate, and dried under vacuum. The crude oil was purified by silica gel chromatography. The mobile phase is MeOH:DCM 2:98. ^{1}H NMR (300 MHz, Chloroform-d) δ 8.18 (d, J = 13.9 Hz, 2H), 7.40 – 7.31 (m, 4H), 7.21 (t, J = 7.5 Hz, 2H), 7.12 (d, J = 8.2 Hz, 2H), 6.15 (d, J = 13.9 Hz, 2H), 3.81 (d, J = 5.6 Hz, 4H), 3.71 (s, 6H), 1.81 (s, 12H). m/z = 413.4, Yield: 87%.

Photophysics:

All spectra were taken in Omnisolv purity spectroscopy-grade solvents and Millipore pure water. Absorbance spectra were taken using an Agilent™ Technologies Cary 60 UV-Vis with a pulsed xenon lamp as the excitation source. The error in reported wavelength values is at most 0.5 nm. Fluorescence and excitation spectra were recorded on a Jobin Yvon Spex Fluorolog-3-11. Sample excitation was achieved via a xenon arc

lamp with a monochromator providing wavelength selection. Right angle light emission was sorted using a monochromator and fed into a Hamamatsu R928P photomultiplier tube with photon counting. Short and long pass filters were used where appropriate. All emission and excitation spectra were taken as an average of two spectra subtracted from two blank spectra of the same excitation wavelength taken on the same day. Quantum yields of fluorescence were measured using Cy5 (φ_f=0.27) and *meso*-Chloro-Cy7 (φ_f=0.12) as references in acetonitrile. In addition to subtracting out background, anomalous features due to excitation and emission overtones were subtracted out of the spectra.

Kamlet-Taft analysis was performed by taking the absorbance and emission spectra of a dye at the same concentration (Absorbance of ~0.2) in a number of different solvents. The absorbance and emission maxima were recorded, and then the parameters were solved using the following equation:

(2) $\Delta\lambda = \alpha_s \times \alpha_D + \beta_s \times \beta_D + \pi^*{}_s \times \pi^*{}_D$

where $\Delta\lambda$ refers to the wavelength of etiher absorption or emission relative to that of the dye in cyclohexane, α is the hydrogen bond accepting parameter, β is the hydrogen bond donating parameter, and π^* is the parameter for the dipole moment. The subscript refers to if the parameter refers to the solvent (s) or dye (D).

A.4: References

1. Solntsev, K. M.; Huppert, D.; Agmon, N.; Tolbert, L. M., Photochemistry of "super" photoacids. 2. Excited-state proton transfer in methanol/water mixtures. *Journal of Physical Chemistry A* **2000,** *104* (19), 4658-4669.
2. Spry, D. B.; Goun, A.; Fayer, M. D., Deprotonation Dynamics and Stokes Shift of Pyranine (HPTS). *The Journal of Physical Chemistry A* **2006,** *111* (2), 230-237.
3. Agmon, N., Elementary Steps in Excited-State Proton Transfer. *The Journal of Physical Chemistry A* **2005,** *109* (1), 13-35.

4. Tolbert, L. M.; Solntsev, K. M., Excited-State Proton Transfer: From Constrained Systems to "Super" Photoacids to Superfast Proton Transfer. *Accounts of Chemical Research* **2002,** *35* (1), 19-27.
5. Pines, D.; Pines, E., Solvent Assisted Photoacidity. In *Hydrogen-Transfer Reactions*, Wiley-VCH Verlag GmbH & Co. KGaA: 2007; pp 377-415.
6. Solntsev, K. M.; Huppert, D.; Agmon, N., Solvatochromism of beta-naphthol. *Journal of Physical Chemistry A* **1998,** *102* (47), 9599-9606.
7. Catalan, J.; Lopez, V.; Perez, P.; Martinvillamil, R.; Rodriguez, J. G., PROGRESS TOWARDS A GENERALIZED SOLVENT POLARITY SCALE - THE SOLVATOCHROMISM OF 2-(DIMETHYLAMINO)-7-NITROFLUORENE AND ITS HOMOMORPH 2-FLUORO-7-NITROFLUORENE. *Liebigs Annalen* **1995,** (2), 241-252.
8. Kamlet, M. J.; Abboud, J. L.; Taft, R. W., SOLVATOCHROMIC COMPARISON METHOD .6. PI-STAR SCALE OF SOLVENT POLARITIES. *Journal of the American Chemical Society* **1977,** *99* (18), 6027-6038.
9. Kamlet, M. J.; Abboud, J. L. M.; Abraham, M. H.; Taft, R. W., LINEAR SOLVATION ENERGY RELATIONSHIPS .23. A COMPREHENSIVE COLLECTION OF THE SOLVATOCHROMIC PARAMETERS, PI-STAR, ALPHA AND BETA, AND SOME METHODS FOR SIMPLIFYING THE GENERALIZED SOLVATOCHROMIC EQUATION. *Journal of Organic Chemistry* **1983,** *48* (17), 2877-2887.
10. Kamlet, M. J.; Taft, R. W., SOLVATOCHROMIC COMPARISON METHOD .1. BETA-SCALE OF SOLVENT HYDROGEN-BOND ACCEPTOR (HBA) BASICITIES. *Journal of the American Chemical Society* **1976,** *98* (2), 377-383.
11. Taft, R. W.; Kamlet, M. J., SOLVATOCHROMIC COMPARISON METHOD .2. ALPHA-SCALE OF SOLVENT HYDROGEN-BOND DONOR (HBD) ACIDITIES. *Journal of the American Chemical Society* **1976,** *98* (10), 2886-2894.
12. Johannes, H. H.; Grahn, W.; Reisner, A.; Jones, P. G., Ethynylated, Vinylated, and Hetarylated Indodicarbocyanines by Palladium-Catalyzed Cross-Coupling Reactions. *Tetrahedron Lett* **1995,** *36* (40), 7225-7228.
13. Ismail, L. F. M.; Abdel-Mottaleb, M. S. A., Fluorescence studies of photoactivated acidity of 2,2′-Dihydroxy-1,1′-Binaphthalene. *International Journal of Photoenergy* **2003,** *5* (4).
14. Still, W. C.; Kahn, M.; Mitra, A., Rapid chromatographic technique for preparative separations with moderate resolution. *The Journal of Organic Chemistry* **1978,** *43* (14), 2923-2925.

Appendix B: Hartree-Fock Computation of Cation-π interactions

B.1: Introduction

The computational results in Chapters 2 and 3 were performed using DFT. As discussed in Chapter 1, DFT is not a variationally correct method. As a control, many of the same computations performed in Chapters 2 and 3 were also done using HF. Several were also performed using MP2, a Möller Plesset method that includes correlation terms and is more advanced than HF. What is important to note is that many of the results are qualitatively the same as DFT-based methods. However, the M06/6-31G(d,p) method does a better job of matching empirical results.

B.2: Computations of cation-π interactions between benzene and various cations

Table 1: Cation-π Binding Energies (kcal/mol)[a]

	Benzene	**F-Benzene**	**F_2-Benzene**		**F_3-Benzene**	
			1,3	1,4	1,2,3	1,3,5
Sodium	27.1	22.0	17.1	16.8	12.9	12.5
Ammonium	15.3	11.7	8.9	8.2	5.1	5.1
Tetramethylammonium (1)	4.7	3.4[b]	2.1[b]	-	0.8[b]	1.1
Tetramethylammonium (2)	5.6	4.1[b]	2.5[b]	2.5	1.0[b]	1.4
Tetramethylammonium (3)	6.8	5.0[b]	3.2[b]	3.2	1.5[b]	2.0
Guanidinium (T-Shaped)	10.8	8.0	5.6	5.3	3.5[b]	3.5
Guanidinium (Parallel)	4.5	3.2[b] /5.1[c]	1.8[b]	2.1	0.5[b]	1.0

[a] HF 6-31G** calculations; full geometry optimization unless otherwise noted. [b] Binding energy obtained via point energy method, and does not represent a geometry-optimized complex. [c] Structure is substantially offset from the center of the ring.

Table 2: Cation-π Binding Energies at high levels of theory (kcal/mol)[a]

	Benzene	**F_1Benzene**	**1,3-F_2Benzene**	**1,2,3-F_3Benzene**
Sodium	23.16[b]	18.35[b]	13.73[b]	9.42[b]
Ammonium	19.4	16.4	13.8	10.5
Guanidinium (Parallel)	10.5[c]	9.1[c]	7.8[c]	6.0[c]

[a] MP2/6-311+G** calculations; full geometry optimization unless otherwise noted. [b] HF 6-311+G** calculations, full geometry optimization. [c] Binding energy obtained via point energy method, and does not represent a geometry-optimized complex.

Table 3: Edge-on Binding Energies (kcal/mol)[a]

	Benzene	**F-Benzene**	**1,3-F_2-Benzene**	**1,2,3-F_3-Benzene**
Sodium	5.4	22.3	19.2	26.7
Tetramethylammonium (1)	0.02	5.6	4.2	–
Tetramethylammonium (2)	0.02	6.5	–	–
Tetramethylammonium (3)	0.02	8.5	6.7	8.9
Guanidinium	-0.01	11.8	9.7	10.2

[a] HF 6-31G** calculations; full geometry optimization in all cases.

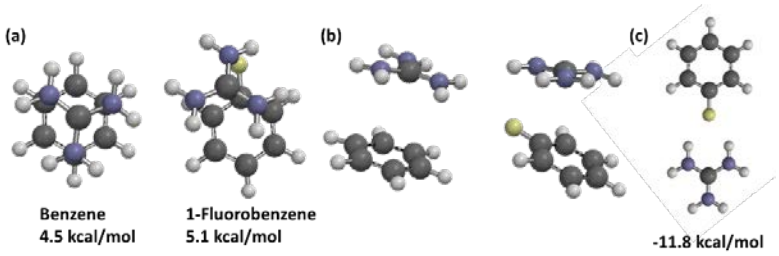


Figure 1: Geometry-minimized structures of guanidinium binding in a stacked/parallel fashion to (a) benzene and fluorobenzene with (b) alternate views of both structures. The guanidinium-fluorobenzene side-on interaction is also shown (c). Binding energies are computed at HF/6-31G** and are all geometry minimized structures.

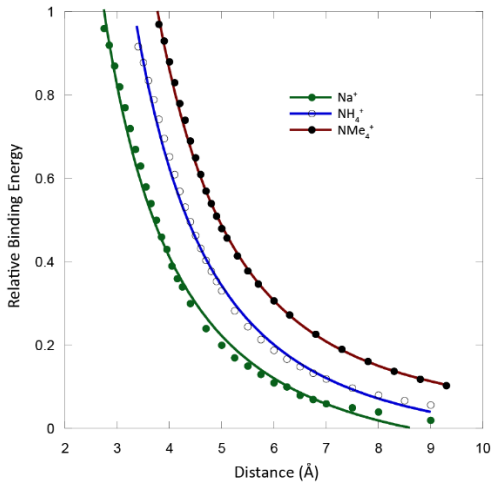


Figure 2: Distance dependence of cation-π binding for sodium, ammonium, and NMe4+ (one methyl down) ions. Binding energy is relative to the optimal binding, and distance is from the center of the benzene to Na^+ ; N atom of ammonium; or methyl C of tetramethylammonium. Complexes with very short distances deviate from this electrostatic plot and were not included in the fitting.

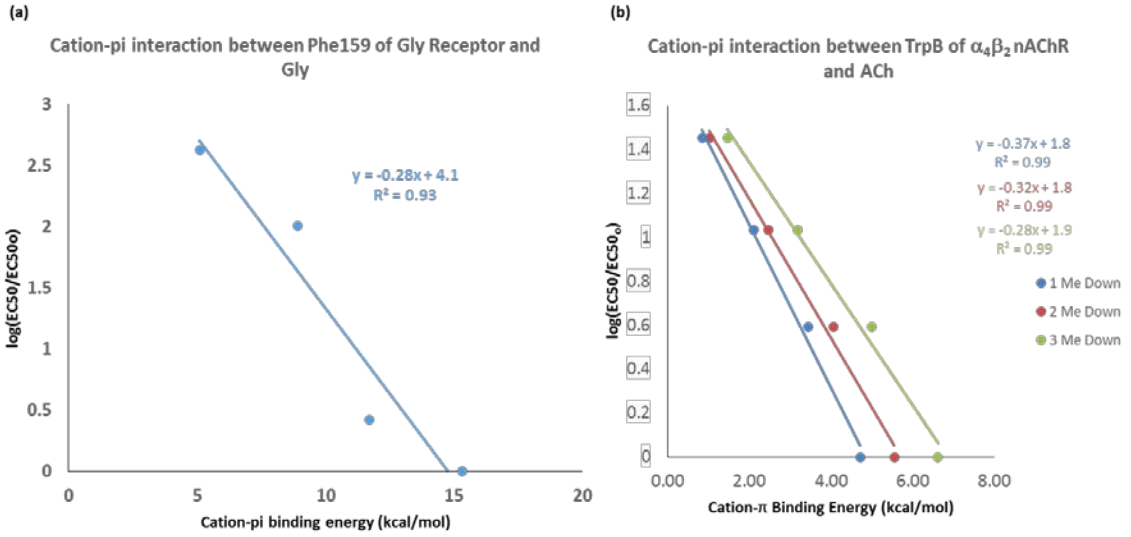


Figure 3: Cation-π binding plots for (a) Glycine at the glycine receptor using computed cation-π binding energies for ammonium ions and (b) acetylcholine at the α¬4β2 nicotinic acetylcholine receptor (nAChR) using cation-π energies for various conformations of tetramethylammonium ions.

B.3: Calculations between cations and more than one aromatic group

Table 4. Binding energies (kcal/mol) to a multiple-benzene aromatic boxes[a]

Cation	Number of benzenes	Binding Energy	Theoretical[b]	Discrepancy
NH_4^+	3	38.5	45.9	7.4
	4	46.0	61.2	15.2
NMe_4^+	3	19.3	20.4	1.1
	4	24.8	27.2	2.3
	5	27.2	31.9	4.7

[a] all calculations at HF/6-31G(d,p), energies in kcal/mol [b] The predicted binding energy if every cation-π interaction in the complex was geometrically ideal. Mathematically, it's the number of benzenes multiplied by the relevant binding energy in Table 1 (above).

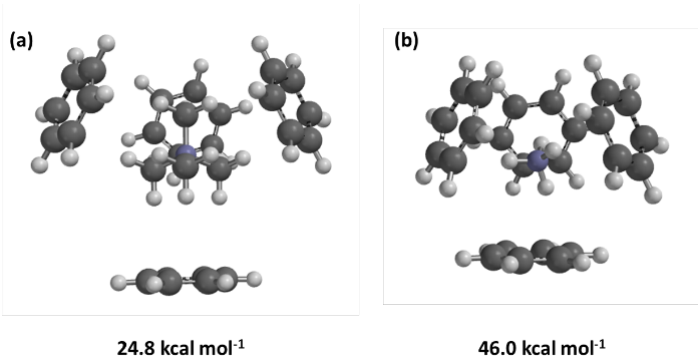


Figure 4: Geometry-optimized complexes of four benzenes to (a) one tetramethylammonium and (b) one ammonium ion. Binding energies computed at HF/6-31G**.

Table 4: Binding of tetramethylammonium to protein aromatic boxes. [a]

AChBP		Histone binding Protein	
Aromatic Box	17.7	Aromatic Box	15.1
TrpB	6.5	Tyr26	4.3
TrpD	4.8	Trp47	6.4
TyrA	1.1	Trp50	7.2
TyrC1	3.7		
TyrC2	4.6		
Sum	20.7	Sum	17.9

[a] HF 6-31G** calculations; single-point energies.

B.4: Methods

All calculations were performed using Spartan 14[1] unless otherwise stated.

Calculating Cation-π Energies:

Cation-π interactions to benzene and derivatives were evaluated with full geometry optimization at HF/6-31G(d,p) with energies calculated using the equation below:

$$BE = (E_{+} + E_{\pi}) - E_{TOT}$$

where E_{TOT} is the total complex energy, E_{+} is the energy of the cation alone, and E_{π} is the energy of the aromatic system, whether it is a single aromatic or an aromatic box bereft of a cation. In this form, a positive BE signifies a favorable interaction.

In most cases, full geometry optimization was performed for complexes of cations to fluorinated aromatics. In some cases, however, this resulted in geometries that would not be considered cation-π interactions. In such cases, fluorines were appended to the benzene ring with C-F bond distances set to values determined by optimization of the isolated aromatic system. This allowed determination of a cation-π binding energy with the cation at the position appropriate to the benzene-ion complex. We will refer to this as the single-point binding energy. The structures were then allowed to relax to the lowest-energy geometry in the gas phase. The binding energy itself was calculated using equation 1.

Several cation-π structures represented relatively shallow local minima, and were challenging to isolate. For the three separate orientations of the tetramethylammonium ion (one, two and three methyl groups facing the benzene), starting structures were the molecular mechanics-minimized structure. This enforced a symmetry on the structure that was maintained while optimizing at a higher level of theory. To achieve a stacked conformation for guanidinium-benzene, the distance from the center of the guanidinium

molecule to three meta carbons of benzene was constrained to 3.4 Å and the (constrained) equilibrium geometry was calculated at HF/6-31G(d,p). The final stacked geometry was obtained using this restricted geometry as a starting point for full geometry optimization.

Aromatic box calculations were performed as described in Chapter 3, using Gaussian 09 for the computations and Gaussview 5 for visualization.[2-3]

B.5: References

1. Wavefunction Inc, I. C., Spartan '14.
2. 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.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; 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, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.
3. Roy Dennington, T. K., and John Millam, Gaussview, Version 5. Semichem, Inc: Shawnee Mission, KS, 2009.

Appendix C: Orbital mixing of aromatic-acetylene cyanine dyes using Hartree-Fock

C.1: Introduction

As discussed in Chapter 1, while TDDFT is a valuable resource to study the excited state of extended chromophores, it (a) does not conform to the variational theorem and (b) the orbitals it is based off of (the Kohn-Sham orbitals) are not based in wavefunction theory. To that end, this appendix contains orbital mixing diagrams analogous to those presented at the end of chapter 5 computed entirely in Hartree Fock. Although the energies may change, the orbitals and conclusions are qualitatively the same.

C.2: Orbital Mixing Diagrams

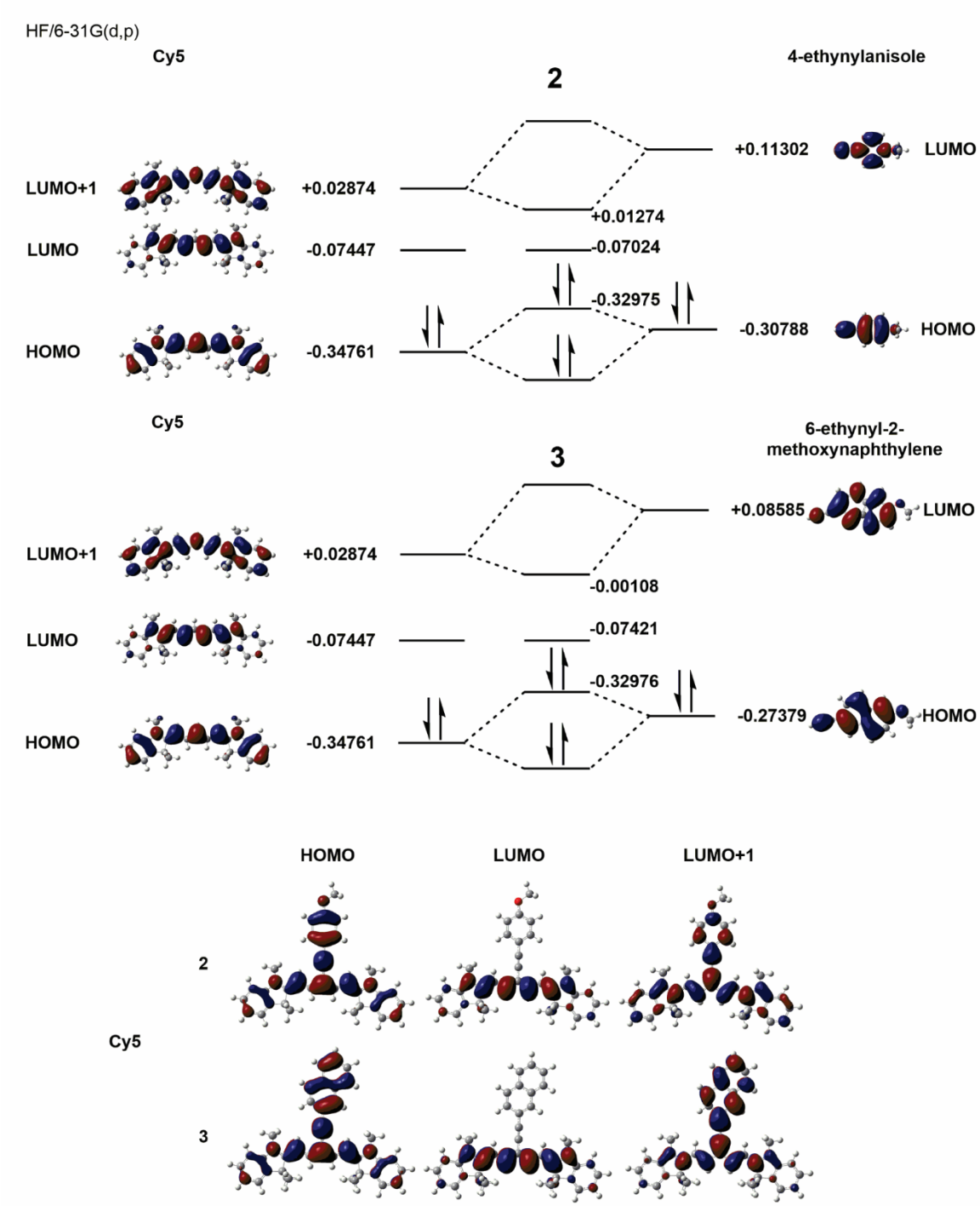


Figure 1: Orbital mixing diagram of (top) **2** and (middle) **3**, with the resulting orbitals of interest shown at the bottom. Compound numbering is consistent with Chapter 5. Energies in hartrees, all calculated by HF/6-31G(d,p).

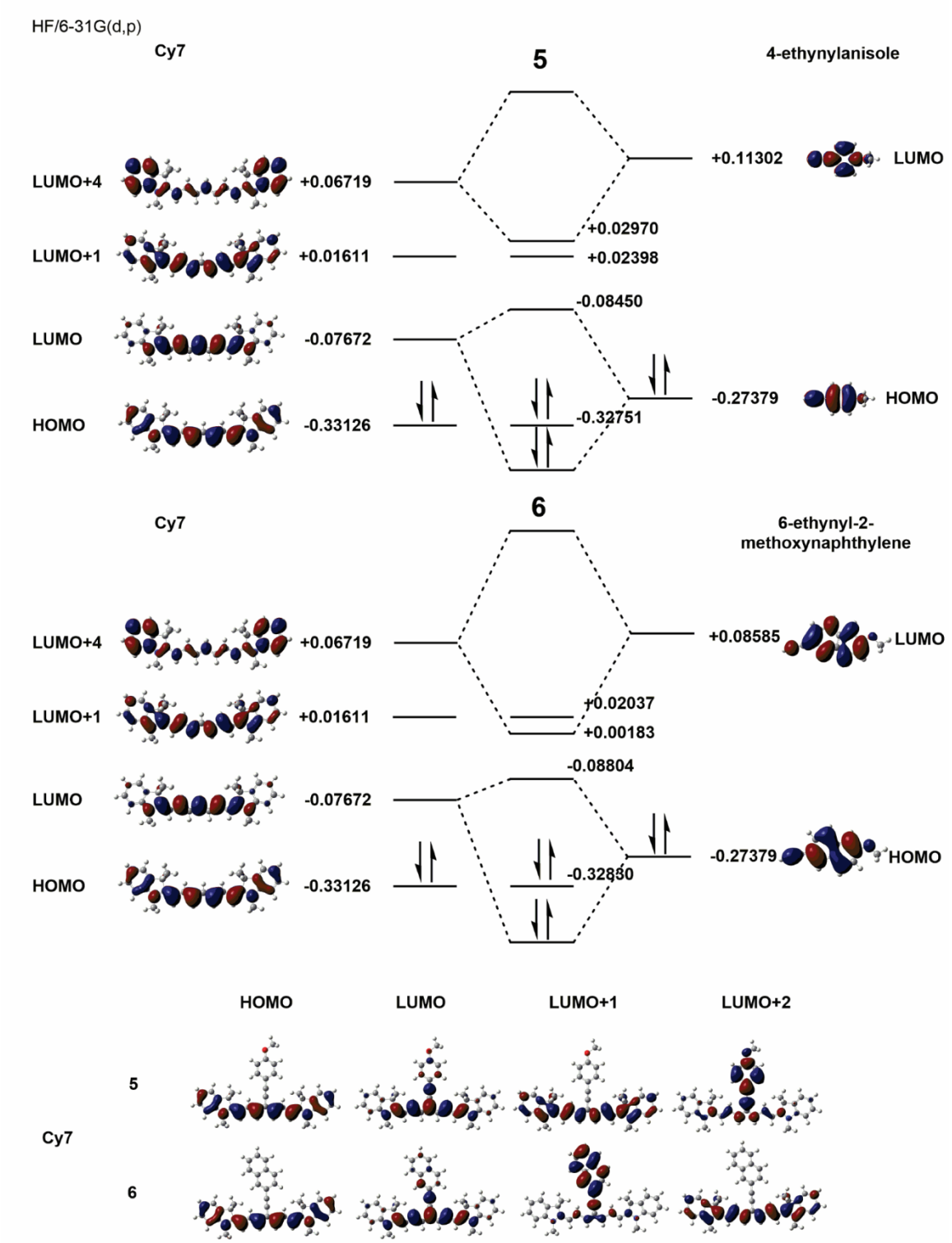


Figure 2: Orbital mixing diagram of (top) **5** and (middle) **6**, with the resulting orbitals of interest shown at the bottom. Compound numbering is consistent with Chapter 5. Energies in hartrees, all calculated by HF/6-31G(d,p).