TRANSFERABILITY IN *AB INITIO* QUANTUM CHEMISTRY: CORRELATED ELECTRONIC STRUCTURE THEORY FOR LARGE MOLECULES

By

THOMAS FRANK HUGHES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2008

© 2008 Thomas Frank Hughes

To my wife, Kelly C. Hughes, who I love for being an amazing human being with an infectious personality. And to my parents, Thomas James and Rose Ann Hughes; and my two older sisters, Erica Christine and Kelly Marie Hughes, for their inspiration.

ACKNOWLEDGMENTS

Special recognition needs to be given to my advisor, Rodney J. Bartlett, for his great mentorship, as well as to my committee for their supervision and assistance. Special thanks is given to the Bartlett group for their help with the ACES suite of programs and for very many rich and stimulating discussions. Additionally, I also owe great thanks to many students, postdocs, staff, and faculty at QTP for their friendship and to my Alma mater, the University of North Florida, for instilling intellectual curiosity. Lastly, on behalf of the Bartlett group I would also like to thank the National Science Foundation, Air Force Office of Scientific Research, and the Army Research Office for funding.

TABLE OF CONTENTS

page

ACKNOWLEDGMENTS 4

LIST OF TABLES 7

LIST OF FIGURES 8

ABSTRACT 10

CHAPTER

1 QUANTUM CHEMISTRY 12

1.1 *Ab Initio* Electronic Structure Methods 12
1.1.1 Hartree-Fock 12
1.1.2 Coupled-Cluster 16
1.2 Approximate Correlated Independent Particle Methods 18

2 LOCAL APPROXIMATIONS FOR THE GROUND STATE 20

2.1 Natural Localized Molecular Orbitals 21
2.2 Fragment Based Approaches 23
2.2.1 Fragment Molecular Orbitals 23
2.2.2 Hartree-Fock with Fragment Molecular Orbitals 24
2.2.3 Electron Correlation with Fragment Molecular Orbitals 27
2.3 Parallelization in the ACES Programs 28
2.3.1 Super Instruction Assembly Language 28
2.3.2 Super Instruction Architecture 29
2.3.3 Integral Direct Computation 30
2.3.4 Automated Dual-Layer Parallelization Strategy 30
2.4 Localized Coupled-Cluster 31
2.4.1 Localized T_3 Contributions 34
2.4.2 Avoiding T_3 Storage 36
2.4.3 Non-Local Interactions in Hybrid Methods 39
2.5 Implementation 42
2.6 Applications 44
2.6.1 Motivation 44
2.6.2 Poly-Glycine 46
2.6.3 Met-Enkephalin 48

3 LOCAL APPROXIMATIONS FOR IONIZED STATES 65

3.1 Methods for Calculating Excited States 65
3.2 Approximate Methods for Calculating Excited States 66
3.2.1 Localized Methods 66
3.2.2 Fragmented Methods 67

3.3 Chromophores . . . 67
3.3.1 Compact Representations . . . 67
3.3.2 Locating Chromophores with Lower Level Wavefunctions . . . 68
3.3.3 Interacting Chromophores and Charge-Transfer . . . 69
3.4 Methods for Calculating Ionized States . . . 71
3.4.1 IP-EOM-CC Wavefunction . . . 72
3.4.2 Eigenvalue Equation in $\bar{H}$ . . . 72
3.4.3 Left-Hand Eigenstate of $\bar{H}$ . . . 73
3.5 Localized Methods for Calculating Ionized States . . . 74
3.5.1 Locating Chromophores: Bond Ionization Energies . . . 75
3.5.2 Locating Chromophores: ΔE Bond Ionization Energies . . . 76
3.5.3 Locating Chromophores: IP-EOM-MP2 Density Matrices . . . 77
3.5.4 Electrostatic Potentials . . . 78
3.6 Implementation . . . 79
3.7 Applications . . . 79
3.7.1 Motivation . . . 79
3.7.2 Example Using Bond Ionization Energies . . . 82
3.7.3 Example Using IP-EOM-MP2 Density Matrix . . . 83
3.7.4 $(H_2O)_{13}$ . . . 85
3.7.5 Gly_5 α-Helix . . . 86
3.7.6 $(H_2O)_{28}$ . . . 86

4 LOCAL APPROXIMATIONS FOR MOLECULAR PROPERTIES . . . 107

4.1 Effective Hamiltonian Dynamic Polarizabilities . . . 108
4.1.1 Size-Extensivity . . . 109
4.1.2 Linear Response . . . 110
4.2 Localized Effective Hamiltonian Dynamic Polarizabilities . . . 111
4.2.1 Localized De-Excitation Operators . . . 111
4.2.2 Effective Bond Dynamic Polarizabilities . . . 113
4.2.3 Effective Bond Response Matrices via Löwdin Partioning . . . 115
4.3 Dispersion Interactions . . . 116
4.3.1 Casimir-Polder Equation . . . 117
4.3.2 Corrective Dispersion Potential . . . 117
4.3.3 Anisotropic Contributions . . . 118
4.4 Implementation . . . 118
4.5 Applications . . . 119
4.5.1 Motivation . . . 119
4.5.2 Tryptophan Dynamic Polarizabilities . . . 122
4.5.3 Poly-Glycine Dynamic Polarizabilities . . . 123
4.5.4 Alkane Dispersion Coefficients . . . 124
4.5.5 Di-Glycine Dispersion Coefficients . . . 124

REFERENCES . . . 139

BIOGRAPHICAL SKETCH . . . 145

LIST OF TABLES

Table page

2-1 Transferability of bond energies from different molecules. 58

2-2 NLSCC correlation energies for poly-glycine. 59

2-3 CCSD bond correlation energies for non-glycine residues. 63

2-4 Summary of NLSCCSD calculations on met-enkephalin. 64

3-1 Partial atomic charges for electronic states of H_2O. 89

3-2 Partial atomic charges for electronic states of $(H_2O)_2$. 90

3-3 IP-EOM-CCSD valence bond ionization energies. 93

3-4 The largest bond ionization energies for $(H_2O)_8$. 98

3-5 IP-EOM-CCSD valence ionization energies of $(H_2O)_{28}$. 106

4-1 Static bond polarizabilities for tryptophan. 132

4-2 Frequency dependent NLSCCSD polarizabilities for poly-glycine. 134

4-3 Dispersion coefficients per methylene unit cell. 136

4-4 Bond dispersion coefficients for di-glycine. 138

LIST OF FIGURES

Figure page

2-1 Increasing sparsity of the CCSDT-3 amplitudes with system size. 51

2-2 Sparsity of two-electron integrals in a localized virtual space. 52

2-3 Transferability of local amplitudes from size-(in)extensive methods. 53

2-4 Transferability of $|T_2|$ and $|T_3|$ from CCSDT-3 in NLMOs. 54

2-5 Transferability of CCSD and CCSDT-3 bond energies in NLMOs. 55

2-6 Effect of triple excitations on calculated bond energies. 56

2-7 Definition of the regions for the NLSCC calculation of poly-glycine. 57

2-8 NLSCC correlation energy per unit cell of poly-glycine. 60

2-9 Model of the three-dimensional penta-peptide met-enkephalin. 61

2-10 Model showing the side-chains of non-glycine residues. 62

3-1 Ball and stick models of water dimer geometries. 88

3-2 Transferability of the ionized state projection of $\bar{H}$. 91

3-3 Transferability of a valence ionized state wavefunction. 92

3-4 Transferability of IP-EOM-CCSD bond ionization energies. 94

3-5 Ball and stick model of quasi-linear $(H_2O)_8$. 95

3-6 Total core ionization energies of growing water clusters. 96

3-7 Bond ionization energy contributions of growing water clusters. 97

3-8 Bond ionization energies for $(H_2O)_8$. 99

3-9 IP-EOM-CCSD ionization energies from lower level density matrices. 100

3-10 Ionization chromophores in $(H_2O)_{13}$. 101

3-11 Bond ionization energies for the $(H_2O)_5$ chromophore. 102

3-12 Ionization chromophores in gly_5 α-helix. 103

3-13 Bond ionization energies for the gly_1 and gly_2 chromophores. 104

3-14 Ball and stick models of $(H_2O)_{28}$. 105

4-1 Transferability of the effective Hamiltonian. 126

4-2 Transferability of CCSD bond polarizabilities. 127
4-3 Frequency dependent bond polarizabilities and associated poles. 128
4-4 Frequency dependent bond polarizabilities for ethanol. 129
4-5 Frequency dependent bond polarizabilities for ethylamine. 130
4-6 Ball and stick models of tryptophan and di-glycine. 131
4-7 Frequency dependent bond polarizabilities for tryptophan. 133
4-8 Transferability of dispersion coefficients for alkanes. 135
4-9 Convergence of di-glycine dispersion coefficients with grid-points. 137

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TRANSFERABILITY IN *AB INITIO* QUANTUM CHEMISTRY: CORRELATED ELECTRONIC STRUCTURE THEORY FOR LARGE MOLECULES

By

Thomas Frank Hughes

December 2008

Chair: Rodney J. Bartlett
Major: Chemistry

The natural linear-scaled coupled-cluster [N. Flocke and R. J. Bartlett, J. Chem. Phys. **121**, 10935 (2004)] methodology is extended to more advanced electronic structure problems. This method exploits the extensivity of the coupled-cluster wavefunction to represent it in terms of transferable electronic structure regions, typically incorporating natural localized molecular orbitals, thereby providing a systematically improvable method for large molecules. Correlated accuracy is built in by allowing approximately fixed electronic structure regions to be perturbed through classifying interactions among regions in a molecule in terms of important inter-region excitations and de-excitations. Subject to its representation in terms of these localized orbitals the correlated effective Hamiltonian from equation-of-motion coupled-cluster theory, which in principle offers an exact solution of a chemical system, is transferable meaning that for a large target system the effective Hamiltonian is calculable from smaller effective Hamiltonians. It is then possible to extract bond dependent quantities, for example ground, excited, ionized, etc. energies, as well as response properties like dynamic polarizabilities and dispersion coefficients, and also transferable wavefunctions and density matrices. The extent to which bonds or functional groups, as opposed to atoms which are one order of interaction removed from covalent bonds, can be combined to provide the properties of a large target molecule is investigated. Given that the non-canonical localized orbitals are not energy eigenstates the theory lacks a Koopmans' theorem thereby complicating the treatment of

frequency-dependent theories requiring chromophoric regions to localize excited or ionized state wavefunctions. Less transferable long-range weak or charge-transfer interactions are represented using electrostatic potentials or other corrective long-range potentials, some built from a number of local regions as opposed to a single region. The natural linear-scaled methods reproduce conventional results for amenable examples. More realistic applications include water clusters, tryptophan, poly-glycine, and met-enkephalin.

CHAPTER 1
QUANTUM CHEMISTRY

Modeling chemical reactions poses a great challenge to theoretical methods because of the accuracy and time required in treating a large number of collective quantum mechanical (QM) and statistical mechanical degrees of freedom. Development of approximate electronic structure methods aim to take advantage of underlying simplicities via cancellations which are unfortunately difficult to track without resorting to many-body *ab initio* methods. These latter methods are based on first principles and thus have the rigor needed to fully describe equilibrium and non-equilibrium regions of the potential energy surfaces (PESs) of ground, excited, ionized, or attached electronic states and their spectra. With these advances it is possible to move beyond electronic structure theory and predict chemistry, for example reaction dynamics. Although gas phase results oftentimes approximate those of other phases, in general theories which bridge the gap between the microscopic properties of particles to the macroscopic properties of a large system are needed. Using density functional theory (DFT) with the B3LYP exchange-correlation- (XC-) functional remains the quantum chemical method of choice for treating large molecules because of its efficiency and its approximate inclusion of dynamic electron correlation. However, there are still many advances that need to be made in developing a density based theory which includes *a priori* the many-body cancellations needed to make an exact linear-scaled electronic structure method.

1.1 *Ab Initio* Electronic Structure Methods

1.1.1 Hartree-Fock

The focus of quantum chemistry is to solve the Schrödinger equation,

$$\mathcal{H}|\Psi\rangle = E|\Psi\rangle, \tag{1–1}$$

for the molecular Hamiltonian,

$$\mathcal{H} = -\sum_{\alpha} \frac{1}{2m_\alpha}\nabla_\alpha^2 - \sum_i \frac{1}{2}\nabla_i^2 - \sum_{i\alpha} \frac{Z_\alpha}{r_{ia}} + \sum_{i<j} \frac{1}{r_{ij}} + \sum_{\alpha<\beta} \frac{Z_\alpha Z_\beta}{r_{\alpha\beta}}, \tag{1–2}$$

where the first two terms are the kinetic energy operators of the nuclei, which are denoted by α, and electrons, which are denoted by i, respectively. The last three terms are nucleus-electron, electron-electron, and nucleus-nucleus interaction operators. The nuclear parameters m_α and Z_α denote nuclear masses and atomic numbers. Given the complexity of this Hamiltonian it is useful to make the following approximation, known as the Born-Oppenheimer approximation, which assumes that $m_\alpha >> 1$ for all α. This approximation reduces the Hamiltonian to an electronic form,

$$\mathcal{H}_{elec} = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_{i\alpha} \frac{Z_\alpha}{r_{ia}} + \sum_{i<j} \frac{1}{r_{ij}}, \tag{1–3}$$

where the kinetic energy of the nuclei is assumed to be negligible and where the nuclear-nuclear repulsion energy, E_{NN},

$$E_{NN} = \sum_{\alpha<\beta} \frac{Z_\alpha Z_\beta}{r_{\alpha\beta}}, \tag{1–4}$$

is simply a constant. The Schrödinger equation simply becomes an electronic equation,

$$\mathcal{H}_{elec}|\Psi_{elec}\rangle = E_{elec}|\Psi_{elec}\rangle, \tag{1–5}$$

where $|\Psi_{elec}\rangle$ and E_{elec} have parametric dependence on the nuclear coordinates. This defines a PES for the molecule where the electrons move according to $|\Psi_{elec}\rangle$ in the field of nuclei which are fixed in different user specified geometries. The total energy is $E = E_{elec} + E_{NN}$.

The Hartree-Fock (HF) approximation, which is central to the treatment of many-electron wavefunctions, assumes a trial $|\Psi_{elec}\rangle$ in the form of a single N-electron Slater-determinant,

$$|\Psi\rangle = \mathcal{A}|\psi_1\psi_2\cdots\psi_N\rangle = \frac{1}{(N!)^{1/2}}\begin{vmatrix} \psi_1(1) & \psi_2(1) & \cdots & \psi_N(1) \\ \psi_1(2) & \psi_2(2) & \cdots & \psi_N(2) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_1(N) & \psi_2(N) & \cdots & \psi_N(N) \end{vmatrix}, \tag{1–6}$$

with $\mathcal{A}$ an antisymmetrization operator necessary to introduce the proper antisymmetry for fermions and where the electronic subscript has been dropped for simplicity. The $\{\psi_i\}$ are a set of single-particle molecular orbitals which are variationally optimized to the HF molecular orbitals by minimizing the single-determinant energy expression,

$$E[\Psi^{HF}] = \langle\Psi^{HF}|\mathcal{H}|\Psi^{HF}\rangle = \min_{\Psi}\langle\Psi|\mathcal{H}|\Psi\rangle, \tag{1–7}$$

under the orthonormality constraint $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. The variation in the single-determinant energy expression, δE, with,

$$E[\Psi] = \langle\Psi|\mathcal{H}|\Psi\rangle = \sum_i\langle\psi_i|h|\psi_i\rangle + \frac{1}{2}\sum_{ij}\langle\psi_i\psi_j||\psi_i\psi_j\rangle, \tag{1–8}$$

is then minimized. Note that h denotes the one-particle core Hamiltonian which contains the first two terms of $\mathcal{H}_{elec}$ and the antisymmetrized two-electron integral in Dirac notation, $\langle\psi_i\psi_j||\psi_i\psi_j\rangle = \langle\psi_i\psi_j|\psi_i\psi_j\rangle - \langle\psi_i\psi_j|\psi_j\psi_i\rangle$. These integrals are defined as the classical coulomb,

$$\langle\psi_i\psi_j|\psi_i\psi_j\rangle = \langle\psi_i|J_j|\psi_i\rangle = \langle\psi_i|\int d\vec{r}_2\frac{\psi_j*(\vec{r}_2)\psi_j(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}|\psi_i\rangle, \tag{1–9}$$

and a purely quantum non-local exact exchange,

$$\langle\psi_i\psi_j|\psi_j\psi_i\rangle = \langle\psi_i|K_j|\psi_i\rangle = \langle\psi_i|\int d\vec{r}_2\frac{\psi_j*(\vec{r}_2)P_{12}\psi_j(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}|\psi_i\rangle, \tag{1–10}$$

contribution where P_{12} exchanges particles one and two. Note that these effective one-particle operators depend upon the orbitals themselves and thus must be solved iteratively thereby giving the self-consistent field method. Variational optimization gives the following equation for the effective one-particle Fock operator, f,

$$f|\psi_i\rangle = [h + \sum_j J_j - K_j]|\psi_i\rangle = \sum_{ij} \epsilon_{ij}|\psi_j\rangle, \tag{1–11}$$

where ϵ_{ij} is a matrix of Lagrange multipliers in the non-canonical HF basis. The unitary basis which diagonalizes the Hermitian matrix ϵ_{ij} is called the canonical HF basis and gives pseudo-eigenvalue equation for the Fock matrix,

$$f|\psi_i\rangle = \epsilon_i|\psi_i\rangle. \tag{1–12}$$

By expanding the single-particle molecular orbitals in an orthonormal basis of atom centered Gaussians, $\{|\mu\rangle\}$,

$$|\psi_i\rangle = \sum_\mu C_{\mu i}|\mu\rangle, \tag{1–13}$$

the following pseudo-eigenvalue equation for the Fock matrix is obtained,

$$\mathbf{FC} = \mathbf{C}\epsilon, \tag{1–14}$$

where,

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu)]. \tag{1–15}$$

The molecular orbital coefficients, $\mathbf{C}$, provide the density (bond-order) matrix, P,

$$\mathbf{P} = \mathbf{CnC}^\dagger, \tag{1–16}$$

where n is the occupation matrix. The self-consistent field energy is then found by tracing with the total effective one-particle operator,

$$E = \frac{1}{2}Tr[\mathbf{P}(\mathbf{H} + \mathbf{F})]. \tag{1–17}$$

The matrix of molecular orbital coefficients, **C**, has dimensions equivalent to the total number of basis functions. When this number is larger than the number of electrons, arbitrary virtual or unoccupied molecular orbitals, $|\psi_a\rangle$, are obtained in addition to the occupied molecular orbitals, $|\psi_i\rangle$.

1.1.2 Coupled-Cluster

The correlation energy of a system is defined as the difference between the exact and HF energies,

$$E^{corr} = E^{exact} - E^{HF}, \tag{1–18}$$

E^{exact} is determined using full configuration interaction (FCI). The CI method introduces excited determinants into the wavefunction by means of an excitation operator, C, which allows for the electrons to escape each others influence via one-, two-, etc. body electron interactions,

$$|\Psi^{CI}\rangle = C|\Psi^{HF}\rangle = (1 + C_1 + C_2 + C_3 + \cdots)|\Psi^{HF}\rangle, \tag{1–19}$$

with,

$$C_n = \frac{1}{(N!)^2} \sum_{ij\cdots ab\cdots} c_{ij\cdots}^{ab\cdots} a^\dagger b^\dagger \cdots ji, \tag{1–20}$$

where $\{c\}$ are the CI amplitudes. The second-quantized creation, $a^\dagger$, and annihilation, i, operators excite electrons,

$$|\Psi_i^a\rangle = a^\dagger i|\Psi^{HF}\rangle = a^\dagger i\mathcal{A}|\psi_1\psi_2\cdots\psi_i\cdots\rangle = \mathcal{A}|\psi_1\psi_2\cdots\psi_a\cdots\rangle. \tag{1–21}$$

Because of the extremely large number of permuted determinants in FCI, in practice it is only computationally feasible on systems a with small number of electrons spurring the need for truncated CI methods which include only up through single-, double-, triple-, etc. determinants. The CI amplitudes and energy are variationally determined as eigenvectors and eigenvalues by diagonalizing the Hamiltonian in the basis of excited determinants. In practice the diagonalization is done using iterative methods.

The coupled-cluster (CC) method offers a cumulant decomposition of the CI excitation operator by means of an exponential waveoperator ansatz,

$$|\Psi^{CC}\rangle = e^T|\Psi^{HF}\rangle = (1 + T + \frac{1}{2!}T^2 + \frac{1}{3!}T^3 + \cdots)|\Psi^{HF}\rangle, \tag{1–22}$$

where $T = T_1 + T_2 + T_3 + \cdots + T_N$ with N the number of electrons,

$$T_n = \frac{1}{(N!)^2} \sum_{ij\cdots ab\cdots} t_{ij\cdots}^{ab\cdots} a^\dagger b^\dagger \cdots ji. \tag{1–23}$$

The relationship between the CI and CC expansion coefficients,

$$\begin{aligned} C_1 &= T_1 \\ C_2 &= T_2 + T_1^2 \\ C_3 &= T_3 + T_2T_1 + T_1^3, \end{aligned} \tag{1–24}$$

shows that the CC is equivalent to FCI in the limit of N-particle excitations. By virtue of the exponential ansatz truncated CC theory (for example CCSD) approaches this limit more effectively than truncated CI theory (for example CISD) because CCSD theory contains not only all of the single and double excitation manifold but also some of the triple, disconnected T_1T_2, quadruple, disconnected T_2T_2, etc. manifolds. Because the CC amplitude equations,

$$\langle Q|\bar{\mathcal{H}}|\Psi^{HF}\rangle = 0, \tag{1–25}$$

are coupled and non-linear, and thus without an eigenvalue equation, they can't be solved using matrix diagonalization, instead they are solved iteratively. Here the Q space can be singly, doubly, triply, etc. excited determinants and $\bar{\mathcal{H}} = e^{-T}\mathcal{H}e^T = (\mathcal{H}e^T)_c$ where c means connectedness in the usual context of CC theory. The initial guess for T_1 and T_2 are oftentimes determined from second-order perturbation theory (MP2),

$$t_i^{a\ (1)} = \frac{f_{ia}}{\epsilon_i - \epsilon_a}, \tag{1–26}$$

and

$$t_{ij}^{ab\ (1)} = \frac{\langle ij||ab\rangle}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b}. \tag{1–27}$$

The energy is similarly given by,

$$\langle \Psi^{HF}|\bar{\mathcal{H}}|\Psi^{HF}\rangle = E = \sum_{ia} f_{ia} t_i^a + \frac{1}{4}\sum_{ijab} t_{ij}^{ab}\langle ij||ab\rangle + \frac{1}{2}\sum_{ijab} t_i^a t_j^b \langle ij||ab\rangle. \tag{1–28}$$

Obviously, other reference functions besides $|\Psi^{HF}\rangle$, can be used.

The highly accurate CC and CI methods are troubled by the task of including all possible permutations of one, two, three, etc. electrons among a large number of single-particle basis functions. As a result, in practice the excitation operator, T, is oftentimes truncated after the two-particle piece, $T = T_1 + T_2 + T_3 + \cdots \approx T_1 + T_2$, leading to singly (S) and doubly (D) excited determinants. Both the CCSD and CISD methods scale as $O(o^2v^4)$, where o and v are respectively the number of occupied and virtual orbitals, but as previously mentioned only the exponential ansatz, unique to CC theory, additionally includes disconnected parts of triple and quadruple excitations. These disconnected triple and quadruple excitations, like T_1T_2 and T_2T_2 respectively, make truncated CC converge to the full CI (FCI) much more quickly than truncated CI, thus making CC the method of choice for high-level applications to medium-sized molecules. The remaining pieces of the triple excitation manifold are said to be connected and may be required for $\approx 1\ kcal\, mol^{-1}$ accuracy in the energy and similar accuracy in other properties. The exponential ansatz used in CC theory makes it a size-extensive method which is a necessary condition for a system with a large number of electrons. For large molecules, a size-inextensive linear ansatz, for example as in truncated CI, would give a vanishing correlation energy per electron [1, 2].

1.2 Approximate Correlated Independent Particle Methods

Many approximate independent particle methods containing model correlation potentials have been proposed to circumvent the cost of many-body correlation methods,

as well as the underlying HF reference, in the treatment of large molecules. The semi-empirical methods add a single-particle correlation potential to Eqn. 1–15,

$$F_{\mu\nu} = H_{\mu\nu} + J_{\mu\nu} - K_{\mu\nu} + V_{c_{\mu\nu}}, \tag{1–29}$$

but represent it only effectively,

$$F_{\mu\nu} = H'_{\mu\nu} + J'_{\mu\nu} - K'_{\mu\nu}, \tag{1–30}$$

by means of empirical model forms for one- and two-electron integrals involving parameters fit to reproduce experiment or high-level theory [3]. Semi-empirical theory is routinely used for large molecule calculations and shown to give qualitative descriptions. Semi-empirical theory as it follows from a parameterization of HF theory is more of a means to an end than it is derivable from first principles. The latter marks the difference between semi-empirical methods and DFT which in principle targets Eqn. 1–29 from a density, as opposed to wavefunction, dependent perspective as given by the Hohenberg-Kohn theorems. Here $K_{\mu\nu}$ is replaced by an oftentimes local exchange potential, $V_{x\mu\nu}$, in the Kohn-Sham equation and many techniques have been studied to find $V_{xc} = V_x + V_c$. Despite being the most reliable electronic structure method for large molecules, DFT methods are not systematic with basis set in the same sense as wavefunction based methods. The oftentimes inaccurate forces, and representation of charge-transfer and weak interactions, obtained with these approximate Hamiltonians leads to incorrect conclusions about chemical reactivity. Additionally, DFT lacks in its description of the bond breaking regime. To remedy this a systematic isolation of the failures of DFT is needed perhaps by comparison with high-level, potentially multi-reference, CC methods. This provides the stimulus to develop faster and more accurate approximations which can systematically approach many-body methods.

CHAPTER 2
LOCAL APPROXIMATIONS FOR THE GROUND STATE

> "By 1935 ... I felt that I had an essentially complete understanding of the nature of the chemical bond"
>
> Linus C. Pauling, Daedalus **99, 988 (1970).**

> "Attempts to regard a molecule as consisting of specific atomic or ionic units held together by discrete number of bonding electrons or electron-pairs are considered as more or less meaningless ..."
>
> Robert S. Mulliken, Physical Review **40, 55 (1932).**

Despite the advances in supercomputing technologies and highly efficient or parallel quantum chemistry programs which utilize them, there is great need for methods which scale favorably with molecule size but are sufficiently accurate to allow for critical assessment of molecular events. These approximate methods are a means to push the limits in applying theoretical chemistry to large molecules as well as to guide the development of both conventional quantum chemistry programs and theories outside of electronic structure programs. One particularly attractive route toward accomplishing this is based on performing rigorous *ab initio* calculations on small regions of a molecule and then processing these regions to provide the properties of large target molecules in the least ambiguous way possible.

Chemistry is built upon the transferability of functional groups. To a reasonable approximation such functional groups retain some essential characteristics in any molecule. Yet the changes that occur when they are combined provide the enormous variations in chemistry, and while often modest in a relative sense, are critical in others. *Ab initio* quantum chemistry, however, while providing quantitative results for the structure and spectra of molecules, does not offer a theory of the chemical bond or functional groups. Bonds are not among the observables of quantum mechanics. Hence, that kind of information can only be extracted *a posteriori*, from a computed wavefunction or density and then subject to a multitude of sometimes conflicting criteria. This failing of quantum mechanics prohibits the natural, *a priori* separation of a large molecule into smaller,

computational and conceptual groups, that would fit together to provide the whole. If this quantum chemistry dilemma were removed, then just as synthetic chemists build molecules from adding and removing functional groups, quantum chemists could do much the same for electronic structure.

In the limit of perfect transferability a given property would be completely characterized in terms of a simple set of universal functional groups. In practice, this is not possible because chemistry arises from the change in the charge distributions and subsequent molecular geometries, as the functional groups are bonded to each other. The next step is to extract the functional group from an environment that allows for such changes in the charge distribution. Representing the properties of molecules in terms of constituent atoms has been the focus of some work [4, 5, 6, 7, 8, 9, 10, 11, 12, 13] but for much of chemistry, functional groups would appear to be a superior choice because they include the effects of strong perturbations like covalent bonds.

One focus of this work is to maximally incorporate fixed electronic structure regions, usually via localized orbitals, in the description of correlated wavefunctions. The relaxation of these regions is accomplished via local correlation methods which are characterized by allowing these fixed electronic structure regions to interact via inter-region excitations. The reference function can be treated similarly, however given the long-ranged electrostatics, a different strategy is employed.

2.1 Natural Localized Molecular Orbitals

Natural localized molecular orbitals (NLMOs) [14, 15] are used to introduce a scale which offers certain advantages to local correlation methods. One benefit of using NLMOs is that the virtual space is composed of localized orthogonal orbitals which are obtained with little convergence difficulties even for large, diffuse basis sets. In the NLMO search for n-centered orbitals, where n is typically small $n < N$ with N the number of centers in the molecule, only small diagonalizations of the single-particle density matrix are required ensuring that the NLMO determination is quite fast [15]. In contrast to other

localization methods which use an initial set of orbitals, the NLMO procedure is based on a single-particle density matrix, $\gamma = \mathbf{C}^\dagger\mathbf{S}^\dagger\mathbf{PSC}$ with $\mathbf{C}$, $\mathbf{S}$, and $\mathbf{P}$ being orbitals, overlap, and single-particle density matrix in atomic orbitals, which can be correlated or uncorrelated. For example, localized correlated orbitals may have certain properties which are more desirable in treating quantum regions that would otherwise have been quite large. These localized correlated orbitals carry information about correlations only effectively due to the fact that they are still single-particle functions. This is to be contrasted with explicitly correlated orbitals, more appropriately called geminals, which are two-particle functions. Restricted as well as unrestricted NLMOs can be obtained from the corresponding restricted HF (RHF) or unrestricted (UHF) density matrices. The NLMOs derive from non-HF natural bond orbitals (NBOs) resulting in the formation of bond/anti-bond pairs. As previously mentioned, in the ACES II [16] NLMO program the bonds can be n-centered with $1 \leq n \leq N$ where N is the number of centers in the molecule and are classified as occupied cores ($n = 1$), lone-pairs ($n = 1$), and bonds ($n \geq 2$) and virtual anti-bonds ($n \geq 2$) and Rydbergs ($n = 1$). The $n > 2$-centered bonds are useful for describing delocalized moieties, for example aromatic molecules, without resorting to more artificial Kekulé structures.

The remarkable transferability of localized chemical bonds and/or functional groups, and their associated properties among the ground states of molecules is one of the most fundamentally important and useful concepts in chemistry. Exceptions to this rule include certain pathological cases, for example D_{3d} cyclohexane cation, where the unpaired hole remains delocalized, despite the fact that all the other electrons in the molecule are localized. For example, the concept of transferable groups in conjugated systems was recently investigated for substituted polyenes where it was found that effects due to substitution propagate to three or four methylene group distances [17]. Not surprisingly, the majority of applications of local methods have been on saturated molecules for which there is a large band-gap and therefore a more local electronic structure.

2.2 Fragment Based Approaches

2.2.1 Fragment Molecular Orbitals

Localized orbitals can also be determined by fragmenting a molecule into smaller regions and simply finding the associated canonical orbitals which by virtue of the fragmentation are localized to the specific fragments. In this way the electrons can be localized to particular fragmented regions of a molecule with higher-order terms given by fragment dimers, trimers, etc. These sets of orbitals are orthonormal within each fragment but the orbitals between fragments would have to be orthonormalized to determine the reference wavefunction of the whole molecule. If the fragments are chosen to be functional groups then these orbitals and the NLMOs should be similar.

The accuracy of this fragment molecular orbital (FMO) method [18, 19, 20] depends on how the molecule is fragmented and as such depends on how covalent bonds at the boundary are treated. The fragmentation is most commonly done by using the chemical intuition of the user, however, more robust means of fragmenting a molecule are needed and under development, hopefully making use of singular value decomposition (SVD). In the conventional implementation of the FMO method the bonds are fragmented electrostatically meaning that the bond is kept intact and incorporated into one of the adjacent fragments with no bond for the other fragment. In the current implementation it is assumed that the fragments are closed shells, however, changes can be incorporated as needed to allow for open shells as in certain charged molecules or radicals. Likewise slight changes are needed to account for diffuse or delocalized character. The FMOs can be used to determine reference, for example HF, wavefunctions, more specifically densities and energies, of large molecules.

The Hamiltonian for a fragment is given by,

$$H_I = \sum_i^{n_I} -\frac{1}{2}\nabla_i^2 - \sum_i^{n_I}\sum_\alpha \frac{Z_\alpha}{r_{i\alpha}} + \sum_i^{n_I}\sum_{J\neq I}^{N}\int dr_j \frac{\rho_J(r_j)}{r_{ij}} + \sum_{i<j}^{n_I}\frac{1}{r_{ij}}, \tag{2–1}$$

where $I, J, \ldots$ denote fragments of which there are N and where n_I specifies the number of electrons in fragment I. Note that α runs over all nuclei in the molecule. The electronic distribution of a given fragment, $\rho_I(r)$, is determined self-consistently in the electronic and nuclear electrostatic potential (ESP) due to the other $(N-1)$ fragments. The same is true for all other fragments, resulting in self-consistently determined ESPs which allow for respective polarization. Note that this ESP accounts for some of the N-body interactions in the molecule with the remainder to be recovered using higher-order fragment dimer, trimer, etc. [21] corrections. The computation of this ESP is the most time consuming step in the FMO calculation. The Hamiltonian for a pair of fragments is given by,

$$H_{IJ} = \sum_{i}^{n_I+n_J} -\frac{1}{2}\nabla_i^2 - \sum_{i}^{n_I+n_J}\sum_{\alpha}\frac{Z_\alpha}{r_{i\alpha}} + \sum_{i}^{n_I+n_J}\sum_{K\neq I,J}^{N}\int dr_k \frac{\rho_K(r_k)}{r_{ik}} + \sum_{i<j}^{n_I+n_J}\frac{1}{r_{ij}}. \tag{2–2}$$

This fragment dimer Hamiltonian includes the ESP from the other $(N-2)$ fragments which were determined in the fragment monomer calculations and are fixed during the dimer calculations. Typically the block diagonal density matrix from the monomer calculation is used as an initial guess for the dimer calculations. If the distance between the fragment monomers in the dimer is large the term can either be neglected or treated using some approximate integral scheme. These Hamiltonians give the following Schrödinger equations,

$$H_I|\Psi_I\rangle = E_I|\Psi_I\rangle \tag{2–3}$$

$$H_{IJ}|\Psi_{IJ}\rangle = E_{IJ}|\Psi_{IJ}\rangle \tag{2–4}$$

which are solved at some level of theory, for example HF.

2.2.2 Hartree-Fock with Fragment Molecular Orbitals

At the level of the Fock matrix exchange is included in the intra-monomer calculations,

$$F^I_{\mu\nu} = H^I_{\mu\nu} + \sum_{\lambda\sigma} P^I_{\lambda\sigma}[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu)] + V^I_{\mu\nu}, \tag{2–5}$$

where μ and ν, as well as λ and σ, belong to fragment I and where V^I is the ESP due to the other $(N-1)$ fragments,

$$V^I_{\mu\nu} = \sum_{J\neq I}^{N} \sum_{\alpha} \langle \mu | \frac{Z_\alpha}{r_\alpha} | \nu \rangle + \sum_{J\neq I}^{N} \sum_{\lambda\sigma} P^J_{\lambda\sigma} (\mu\nu|\lambda\sigma), \tag{2–6}$$

where μ and ν belong to fragment I while α, λ, and σ belong to fragment J. This ESP shows the absence of exchange for the inter-monomer calculations. Exchange is handled similarly for the dimer calculations,

$$F^{IJ}_{\mu\nu} = H^{IJ}_{\mu\nu} + \sum_{\lambda\sigma} P^{IJ}_{\lambda\sigma} [(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\sigma|\lambda\nu)] + V^{IJ}_{\mu\nu}, \tag{2–7}$$

where μ and ν, as well as λ and σ, belong to dimer $I \bigcup J$ and where V^{IJ} is the ESP due to the other $(N-2)$ fragments,

$$V^{IJ}_{\mu\nu} = \sum_{K\neq I,J}^{N} \sum_{\alpha} \langle \mu | \frac{Z_\alpha}{r_\alpha} | \nu \rangle + \sum_{K\neq I,J}^{N} \sum_{\lambda\sigma} P^K_{\lambda\sigma} (\mu\nu|\lambda\sigma), \tag{2–8}$$

where μ and ν belong to $I \bigcup J$ while α, λ, and σ belong to fragment K. These equations give the following pseudo-eigenvalue equations for the Fock matrices in terms of FMOs and orbital energies,

$$\mathbf{F_I C_I} = \mathbf{C_I} \epsilon_\mathbf{I} \tag{2–9}$$

$$\mathbf{F_{IJ} C_{IJ}} = \mathbf{C_{IJ}} \epsilon_\mathbf{IJ}. \tag{2–10}$$

The electronic energy of the molecule is given by the following telescoping series in monomer, dimer, etc. energies,

$$E = \sum_{I}^{N} E_I + \sum_{I<J}^{N} E_{IJ} - E_I - E_J = (2-N) \sum_{I}^{N} E_I + \sum_{I<J}^{N} E_{IJ}, \tag{2–11}$$

while the total energy is given by,

$$E^{tot} = E + \sum_{\alpha<\beta} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}}. \tag{2–12}$$

The total electron density of the molecule is computed as,

$$\rho(\vec{r}) = \sum_{I}^{N} \rho_I(\vec{r}) + \sum_{I<J}^{N} \rho_{IJ}(\vec{r}) - \rho_I(\vec{r}) - \rho_J(\vec{r}) = (2-N)\sum_{I}^{N} \rho_I(\vec{r}) + \sum_{I<J}^{N} \rho_{IJ}(\vec{r}), \qquad (2\text{–}13)$$

and similarly for its matrix representation. Note that FMO, correct through connected N-body interactions, is exact if the target molecule is divided into N fragments and thus larger errors result if the molecule is fragmented into too small of fragments. It is thus a systemically improvable method by virtue of including not only higher-order terms in the telescoping expansion but also by increasing fragment size.

An interesting feature of the FMO method is that because it is based on a theory of interacting fragment densities, it is possible to develop hybrid methods in which the fragment densities are calculated at different levels of theory corresponding to potentially different sized fragments. Such hybrid methods may be particularly useful for studying localized chemical reactions which might be surrounded by otherwise fixed electronic structure regions.

Additionally, by virtue of the fragmentation information about intra-molecular interactions is available *a priori* as opposed to having to extract this information from the single number obtained with a conventional calculation. Given the success of this method in reproducing conventional results, it is important to recognize the internal error cancellations necessary to do this from a more rigorous perspective. Then the systematic error cancellation used here can be exploited in other problem areas of electronic structure. However, not all of the errors are cancelled, resulting in their accumulation in certain cases manifest in the total energies of large molecules but to a lesser extent in energy difference calculations.

In the FMO method, only diagonalizations and integral evaluations over small matrices at a time are needed. This significantly decreases the calculation time necessary to get reference wavefunctions for large molecules to linear or possibly sub-linear, as well as removing difficulties, such as numerical instabilities, etc., when the conventional

methods are applied to large molecules. The FMO method is trivially parallelizable [22] because once the ESP is determined and the monomer calculations are complete, the subsequent independent dimer, trimer, etc. calculations can be done simultaneously. Note that each of the monomer, dimer, trimer, etc. calculations can be further parallelized meaning that the final method can use a dual-layer strategy to the overall parallelization.

The FMO approach has great potential [23] for routine, practical, accurate, and systematically improvable applications to large molecules, for example in protein structure prediction, assessing conformer stability, protein-ligand binding affinity, explicit solvent effects, molecular dynamics simulations, etc. Although large basis sets have been used, most applications have been limited to small, for example minimal, basis sets. Despite this, the results are usually in very good agreement to the conventional methods including cases for which there are multiple thousands of basis functions.

2.2.3 Electron Correlation with Fragment Molecular Orbitals

Correlation energies, for example MP2 or CC, can be computed using a telescoping series in the correlation energy [24, 25]. Unlike many local correlation methods which rely on large QM regions to obtain weak interactions, the FMO procedure accounts for them provided the parent method has sufficient physics to represent them. When treating the correlation problem it is important to note that the conventional FMO approach is not based on using the FMO reference density matrix, but rather focuses on a telescoping series in the correlation energy where monomers, dimers, trimers, etc. are calculated at the correlated level of theory. This gives an FMO correlation energy to be added to the FMO reference energy. There are a significantly smaller number of dimers which need to be calculated in the correlation problem than in the electrostatic one because the correlation contribution is much more local in space. Also in FMO the Fock pseudo-eigenvalue equations are represented in canonical form, as opposed to non-canonical or localized form, and therefore equations based on them are greatly simplified, for example as in using iterative versus non-iterative MP2. This is in contrast to using non-canonical

or localized orbitals in high-level methods. Within the FMO procedure the reference energy can be computed up to three-body terms while the more expensive correlation energy can be computed up to two-body terms. It is important when adding correlation corrections to ensure that the error in the reference function is sufficiently small to make adding correlation worthwhile.

2.3 Parallelization in the ACES Programs

2.3.1 Super Instruction Assembly Language

An option to calculate the FMOs was recently built into the ACES III program [26] which is an extension of the ACES II program [16] that allows for highly efficient parallel computation. The ACES III program is portable and designed to attain excellent performance and scalability on up to 1000 or more processors. The design strategy used for ACES III exploits the efficiency of modern microprocessors to perform floating point operations on a large number of numbers, a super number or block of numbers, as opposed to a single number. It is the microprocessors deep hierarchy of processing speeds versus data sizes which makes this possible. Thus the parallel extension of ACES II requires completely redesigning the way in which it is written. This effort was facilitated by introducing a new language called super instruction assembly language (SIAL) which allows the rather complex algorithms of quantum chemistry to be easily written in a compact, object oriented form which is then processed by a super instruction processor (SIP) responsible for interpreting in terms of super instructions on blocks of numbers, typically 10,000 of them, much like the message passing interface (MPI). Routines from the MPI library are used as needed, especially in sending and receiving data. Array indices can then be grouped into segments imposing a breakdown of the data into blocks which need to be sufficiently large so that computation as opposed to communication dominates at runtime. For example, consider the parallel do construct in SIAL which is used to run a piece of code locally on a block of data from a distributed array. In addition to using distributed arrays for large data structures, like two-electron integrals in which

case the SIP organizes all required communication, there are local arrays for smaller data structures, like one-electron integrals in which each node has a local copy obtained using MPI send.

The SIAL language is a means to perform the large array operations of quantum chemistry in parallel in such a way that all details pertaining to the parallelization are hidden to the programmer. All declarations, including a number of block stacks of fixed sizes, are made at the beginning of the SIAL code and upon compilation a dry-run phase is executed in which the number of block stacks needed for the calculation is determined. These block stacks are allocated from random access memory (RAM) as needed in runtime. Programs such as HF, MP2, CC, etc. are written in SIAL calling super instructions or objects which are written in C and Fortran 77, one such example is the integral computation [27] in ACES III. The ACES III executable runs one or more of these SIAL programs sequentially to achieve its desired computational task, passing data to the next SIAL code using scratch files created at the end of the current SIAL code.

2.3.2 Super Instruction Architecture

Execution is optimized by means of a super instruction architecture (SIA) where processing is performed asynchronously in an attempt to hide latent operations behind other operations, for example in accessing memory or in communication. The SIA divides the requested processors into a master process, responsible for initialization and clean up of SIAL programs, and a number of worker and server processes, responsible for computations, like tensor contraction, matrix diagonalization, etc., and data storage in scratch files, respectively. Efficiency is maximized when accessing memory by carefully scheduling the read from memory to start long before the data is needed. Similarly, other operations are scheduled to start while the results from a previous operation are being written to memory. Some tuning of the problem to the specific architecture available is automated or can be provided by the user as SIP parameters in a ZMAT. The ACES

III program is flexible enough to scale with a small about of RAM for each worker node, however, increasing the RAM per node can increase the overall speed of the calculation.

2.3.3 Integral Direct Computation

Generation of the FMOs in the ACES III program is done efficiently by means of direct computation of one- and two-electron integrals using a new object oriented parallel integral package [27]. The integral package uses state-of-the-art algorithms where great attention has been paid to modern microprocessors, for example avoiding cache misses, using block data structures, etc. Objects have the advantage of being completely independent of other programs from which they can be referenced. Only fragment super instructions, which allow for the calculation of fragment specific integrals, along with the appropriate SIAL code, from which these instructions and some others are referenced, needed to be written. Since the generation of FMOs follows from simply dropping certain atomic orbitals in the integral evaluations, a drop atomic orbital (DROPAO) namelist was created in the ZMAT specifying the mapping between atomic centers and fragments.

2.3.4 Automated Dual-Layer Parallelization Strategy

Automated calculation of HF reference wavefunctions, more specifically densities and energies, for large molecules using these FMOs was accomplished using a series of Perl scripts responsible for the programs initialization, management, and collection. The geometry section of a standard ACES III ZMAT is fragmented according to the user's specification by inserting lines containing Frag1, Frag2, ..., FragN, where N is the total number of fragments used, each followed by lines containing the usual atomic centers and Cartesian coordinates belonging to the given fragment. The initialization script creates all monomer, dimer, trimer, etc. ZMATs from the original ZMAT each with the appropriate DROPAO namelist. Given that the dimer, trimer, etc. ACES III calculations are all independent, they can be simultaneously computed in batches containing as little as one dimer, trimer, etc. or as many as all of them. A dual-layer parallelization strategy is used,

where a user specifies the number of simultaneous dimer, trimer, etc. calculations as well as the number of nodes in the conventional ACES III parallelization.

The first layer of parallelization is trivially accomplished using process forks from the Perl management script. Fork creates an additional processes, another thread of execution, by cloning the current process, called the parent process, and creating child processes which share the same executable code. To distinguish between parent and child processes the call returns non-zero and zero, respectively, where the non-zero value returned to the parent is the process identification (PID) of the child process. Each child process is responsible for managing an entire ACES III calculation. Batch cycles are managed using the waitpid command in Perl where the program waits on the specified number of child PIDs. Fork provides only course-grained parallelization, for example in its current implementation it assumes that for the chosen batch sizes the calculations will take approximately the same amount of time and possibly compete for resources, thus it is not rigorously load-balanced. This dual-layer parallelization strategy is much like how serial ACES II parallelizes finite difference derivatives, however, unlike this approach where each child is responsible for an intense serial ACES II calculation, the implementation in ACES III allows each child to further parallelize. Parsing of monomer, dimer, trimer, etc. output files to produce the final HF results for the large target molecule is performed by the collection script. This script also provides a summary of all of the interaction energies in the large target molecule. Similar parsing of results from correlated calculations provides the conventional way of doing correlated calculations in the FMO methods. The management script is sufficiently general so that it could be used for computing finite difference derivatives in ACES III using the dual-layer parallelization strategy with either FMO based HF or even localized CC.

2.4 Localized Coupled-Cluster

Since pair correlation is well described by MP2, which is cheaper than the well known CC methods, it is the computational chemists method of choice for treating correlation

from first principles. The localized correlation methods were first implemented in MP2 methods with great success and later extended to CC methods. The CC methods expand upon MP2 by allowing electron pairs to influence one another via a cumulant expansion which results in higher, for example three-fold, four-fold, etc., electron excitations. It is possible to avoid the increase in cost for CC methods by using the NLMOs, to simplify the electronic structure by imposing local excitations.

This natural linear-scaled CC (NLSCC) method [28, 29] has been implemented, using the Fortran and Perl programming languages, into the serial ACES II [16] quantum chemistry software package. The Molden [30] modeling package was used regularly for preprocessing input files. More recent work has focused on maximizing the performance of NLSCC by using the parallel ACES III [26] package which is written in SIAL, a new programming language, based on the MPI. Ultimately, the goal is to provide a blackbox, systematically improvable, affordable, and accurate CC method for large molecules. For size-extensive methods, like MP2 and CC in contrast to truncated CI, it is possible to obtain very accurate correlation energies with NLS methods. Obtaining total energies is more cumbersome due to long-ranged electrostatic terms.

Local correlation methods aim to circumvent the non-linear scaling by exploiting an underlying simplicity in both the electronic structure and computational procedure as in the NLSCC method [28]. The NLSCC method has been shown to give very accurate correlation energies for large molecules in which the rate limiting step is the size of the QM regions. This region can be determined self-consistently making these methods systematically improvable. For this region other reduced scaling methods, which are outside of the local approximation, can offer advantages [26, 31, 32, 33, 34, 35]. These methods become especially useful for three-dimensional systems where an accurate local correlation treatment can require large QM regions to represent weak interactions. In a localized basis the correlation space for a given occupied orbital does not grow with system size thereby reducing the number of free parameters and allowing unimportant

interactions to be eliminated from the amplitude equations [36]. Furthermore, the method insists upon transferability of electronic structure units and as such has the potential to provide transferable electronic structure, much the way as geometric structure depends upon transferable geometric units. Therefore, high-level local correlation methods have the potential to become a standard, systematically improvable tool for calculating energies of very large molecules. For systems with a band-gap it can be definitively shown that the computation of many-electron states and their associated density matrices have to scale linearly with respect to particle number [37], lending justification to a variety of linear-scaled approaches and in particular those which account for electron correlation [28, 36, 38, 25, 39, 40, 41, 42].

The ground state NLSCCSD [28] method follows by rewriting the wavefunction, without loss of generality, as,

$$|\Psi\rangle = e^{T_1+T_2}|0\rangle = \prod_i e^{T_1^i+T_2^i}|0\rangle, \tag{2–14}$$

where $|0\rangle$ is the single-determinant reference function currently assumed to encompass the whole molecule but could also be the FMO reference function or possibly some other reference function. In the non-interacting limit by using a size-extensive reference function our reference becomes $|0\rangle = \mathcal{A}(|i\rangle|j\rangle|k\rangle\cdots)$ with i, j, and k being occupied NLMOs or functional groups thereof and $\mathcal{A}$ is an antisymmetrizer. Additional terms like $|ij\rangle$ are needed as the groups are allowed to interact. The operator T_2^i, for an excitation specific to orbital i, is given as,

$$T_2 = \sum_i T_2^i = \sum_i \left(\frac{1}{4} \sum_{jab} t_{ij}^{ab} a^\dagger b^\dagger ji \right), \tag{2–15}$$

where $ij \in \{\text{occ}\}$, $ab \in \{\text{vir}\}$, and $pq \in \{\text{occ} \cup \text{vir}\}$. The assumption is that the operators T^i in the NLMO basis only need to be locally correlated due to the at worst R^{-3} decay of the "dipole-dipole" cluster amplitudes. T_1^i and T_2^i are approximated by only summing over a small region, QM2, composed of orbitals that are spatially close to i. This is shown

below for double excitations,

$$\sum_{jab} = \sum_{jab \in QM2} + \sum_{P(jab) \notin QM2} \approx \sum_{jab \in QM2}, \tag{2–16}$$

with $P(jab)$ a permutation operator which introduces local orbitals outside of QM2. Orbital i is defined to be in the first QM region, QM1, with $QM1 \subset QM2$ giving T_2^i as,

$$T_2^i \approx \frac{1}{4} \sum_{jab \in QM2} t_{ij}^{ab} a^\dagger b^\dagger ji \quad i \in QM1 \quad j, a, b \in QM2. \tag{2–17}$$

Note that the QM1 regions are disjoint from each other as opposed to the QM2 regions. This reduces the number of singles and doubles parameters in the amplitude equations leading to linear or even sub-linear scaling with respect to the number of basis functions. The method is exact in the limit of QM2 being the whole molecule. Unrestricted references can also be used provided that the corresponding α- and β- density matrices are localizable. The total correlation energy is then a sum of orbital correlation energies for each QM1 region,

$$E \approx \sum_i E_i, \tag{2–18}$$

which becomes exact in the limit of QM2 being the whole molecule.

2.4.1 Localized T_3 Contributions

Correlated methods such as MP2 and CC are a minimal requirement for predicting effects such as dispersion interactions which can play important collective roles in large, albeit conventionally unattainable, systems, such as peptides. There are local CC methods which have been designed for single and double excitations [43, 44, 36, 42, 28, 45, 46, 47, 48, 49, 50, 41, 51, 40, 52, 53], however, less work has been done on extending the local CC methods to higher excitations, for example triple excitations [39, 54, 25, 38, 55, 56]. Disconnected T_1T_2 or even T_1^3 obtained with CCSD accounts for some of the triples contribution to the correlation energy, but typically the connected T_3 contributions make significant contributions. It is expected that the triples amplitudes are longer-ranged than the doubles amplitudes because only two of the three occupied-virtual pairs need

to be spatially close to make a significant contribution. This could result in significant modification of local correlation methods which were intended for CCSD. Additionally, as the excitation level increases the non-linearity of the CC equations leads to complicated terms in which there are one, two, three, etc. excitation operators making it difficult to determine if the overall contribution is even local.

The necessity of connected triples is understood from an order-by-order comparison with perturbation theory in which it is found that the triply excited manifold is dominated by connected T_3 as opposed to T_1T_2 or T_1^3. This is in contrast to the dominant term in the quadruply excited manifold which is disconnected T_2^2 as opposed to connected T_4. These connected contributions are not obtained until explicit inclusion of T_3 as in CCSD(T), CCSDT-x (x = 1a,1b,2,3), or CCSDT. These methods scale as non-iterative $O(o^3v^4)$, iterative $O(o^3v^4)$, and iterative $O(o^3v^5)$ respectively, making these CC methods, without modification, prohibitive for large molecules.

In order to expedite the evaluation of triples within NLSCCSDT-3, it is useful to first consider the dominant linear term in the T_3^i equation,

$$\langle^{abc}_{ijk}|(W_N T_2)_c|0\rangle, \tag{2–19}$$

with,

$$H_N = \sum_{pq} F_{N\ pq}\{p^\dagger q\} + \frac{1}{4}\sum_{pqrs} W_{N\ pqrs}\{p^\dagger q^\dagger sr\}, \tag{2–20}$$

and,

$$H_N = H - \langle 0|H|0\rangle, \tag{2–21}$$

which is the cornerstone of CCSD(T) and CCSDT-1x (x = a,b) methods. Within the context of NLSCCSDT-3 non-linear contributions to T_3, for example $\langle^{abc}_{ijk}|(W_N T_2^2)_c|0\rangle$, allow the localized regions to indirectly influence one another, for example by certain strong inter-regional excitations, without destroying the locality or transferability of the

individual regions. These types of terms are incomplete in linear versions such as CCSDT-1x (x = a,b) or CCSD(T). Non-linear terms involving T_3 are limited to the doubles equation $\langle^{ab}_{ij}|(W_N T_1 T_3)_c|0\rangle$ within the NLSCCSDT-3 approximation meaning that all other terms involving T_3 are linear. Eqns. 2–22, 2–23, and 2–24 show the coupled non-linear T_1, T_2, and T_3 CCSDT equations, where in the latter those beneficial terms that were previously mentioned, namely those of CCSDT-3, are underbraced. The CCSDT-3 T_1 and T_2 equations are the same as they are for CCSDT. Prefactors have been dropped for simplicity.

$$\langle^{a}_{i}|(H_N(1 + T_2 + T_1 + T_1T_2 + T_1^2 + T_1^3 + T_3))_c|0\rangle = 0 \tag{2–22}$$

$$\langle^{ab}_{ij}|(H_N(1 + T_2 + T_2^2 + T_1 + T_1T_2 + T_1^2 + T_1^2T_2 + T_1^3 + T_1^4 + T_3 + T_1T_3))_c|0\rangle = 0 \tag{2–23}$$

$$\langle^{abc}_{ijk}|(H_N(\underbrace{T_2} + \underbrace{T_3} + \underbrace{T_2^2} + \underbrace{T_1T_2} + T_2T_3 + T_1T_3 + \underbrace{T_1^2T_2} + \underbrace{T_1T_2^2} + T_1^2T_3 + \underbrace{T_1^3T_2}))_c|0\rangle = 0 \tag{2–24}$$

It can be seen that CCSDT-3 is as close to CCSDT as possible, but still $O(N^7)$ and does not require storing the T_3 amplitudes in a canonical or semi-canonical basis because in those bases the T_3-into-T_3 contributions vanish. For example, the only T_3-into-T_3 contribution in CCSDT-3 is,

$$\langle^{abc}_{ijk}|(F_N T_3)_c|0\rangle, \tag{2–25}$$

which for canonical HF, or semi-canonical orbitals, $F_{pq} = \epsilon_p \delta_{pq}$ becomes diagonal establishing that the set of triples need not be stored. The CCSDT-3 method is correct through fourth-order in the energy and second-order in the wavefunction.

2.4.2 Avoiding T_3 Storage

As an example of locally correlated methods for connected triples consider both non-iterative CCSD(T) [38, 56] and iterative CCSDT-1b [55]. The easiest way to include connected triples contributions is via perturbation theory giving the well known CCSD(T)

method [57] and its non-HF generalization [58]. Perturbative triples corrections like CCSD(T) in a localized basis require both an iterative solution and storage of the triples amplitudes [58] which can destroy performance due to large memory requirements.
It is important to have approximations that eliminate the need to store the triples amplitudes. Note that for local perturbative triples approximations, like local CCSD(T), an approximate triples energy contribution is added to an already approximate singles and doubles energy. This is in contrast with approximate iterative triples methods in which the triples are coupled to the single and double equations, and vice versa. Due to the fact that CCSDT-3 contains many more non-linear triples contributions than CCSD(T) or CCSDT-1x (x = a,b) the triples excitation regions may need to be larger than the double excitation regions.

As previously mentioned the CCSD(T) method provides the easiest connected triples contribution, however, as in the context of second-order with non-canonical HF orbitals, the Fock matrix is block-diagonal, meaning that perturbative methods like CCSD(T) must be solved iteratively thereby requiring storage of the corresponding triples amplitudes [58]. One disadvantage of localized orbitals is that they are not eigenfunctions of any common energy operators,

$$h|p\rangle = \sum_q \epsilon_{pq}|q\rangle. \tag{2–26}$$

Where h is some single-particle energy operator, p, q are localized occupied or virtual orbitals, and ϵ is an energy matrix. The orbitals might diagonalize some other energy operator,

$$h^{eff}|p\rangle = \epsilon_p^{eff}|p\rangle, \tag{2–27}$$

or they might diagonalize some other operator, for example the NLMOs block-diagonalize the density matrix. This results in perturbative methods needing to be solved iteratively. This iterative solution negates one of the most useful properties of perturbative methods in that they are non-iterative for canonical orbitals. If the equations have to be iterated

it is worthwhile to include more terms per iteration amounting to a CC calculation. Since the overhead is in the storage of triples as opposed to the inclusion of more terms, CCSDT-3 is used which is as close to CCSDT as possible while still being $O(N^7)$. Additionally, for CCSDT-3 note that with canonical or semi-canonical orbitals the triples amplitudes need not be stored in contrast to CCSDT.

In the case of NLMOs there are off-diagonal occupied-occupied and virtual-virtual terms in F_N which require that the amplitudes be stored in the conventional implementation. The semi-canonical transformation that eliminates these terms in the case of a non-HF reference function [58], thereby preventing storage, cannot be used here because it would simply return the delocalized canonical HF orbitals. In NLSCCSDT-3 storage of the triples amplitudes in the localized basis is avoided by first solving the CC equations for our small QM regions in the canonical basis and then subsequently unitarily transforming amplitudes and integrals into NLMOs, shown below using the summation convention.

$$t_{i'j'\cdots}^{a'b'\cdots} = U_{aa'}U_{bb'}U_{ii'}U_{jj'}\cdots t_{ij\cdots}^{ab\cdots} \tag{2–28}$$

$$w_{p'q'r's'} = U_{rr'}U_{ss'}U_{pp'}U_{qq'}w_{pqrs} \tag{2–29}$$

That is the localization matrix is applied to the canonical CC wavefunction from a QM region instead of to the reference function. This *a posteriori* transformation does not affect our choice of QM regions because the transferability of the NLMOs necessitates that they can be determined from either a global or fragmented reference function. In Eqn. 2–28 and 2–29 U is the localization transformation obtained in the NLMO procedure from $\mathbf{C}' = \mathbf{CU}$ where $\mathbf{C}'$ are the NLMOs and $\mathbf{C}$ are the canonical HF orbitals. The distinction here in comparison to many other local CC methods is that by virtue of the fact that CC is an infinite-order method, in comparison to non-iterative perturbative methods, there is freedom to apply the localization to the final correlated wavefunction instead of to the

reference function. Therefore, within NLSCCSDT-3 the triples amplitudes do not need to be stored.

2.4.3 Non-Local Interactions in Hybrid Methods

The NLSCCSDT-3 method follows straightforwardly from the NLSCCSD method [28]. The ground state NLSCCSDT-3 wavefunction is rewritten from the CCSDT-3 wavefunction without loss of generality as,

$$|\Psi\rangle = e^{T_1+T_2+T_3}|0\rangle = \prod_i e^{T_1^i+T_2^i+T_3^i}|0\rangle, \tag{2–30}$$

where $|0\rangle$ is the single-determinant reference function. The operator T_3^i for an excitation specific to orbital i is given as,

$$T_3 = \sum_i T_3^i = \sum_i \left(\frac{1}{36} \sum_{jkabc} t_{ijk}^{abc} a^\dagger b^\dagger c^\dagger kji \right). \tag{2–31}$$

The T_1^i and T_2^i are approximated by only summing over a third QM region, QM3, and likewise for T_3^i with respect to a second QM region, QM2, both composed of orbitals that are spatially close and closer to i, respectively. This is shown below for triple excitations,

$$\sum_{jkabc} = \sum_{jkabc \in QM2} + \sum_{P(jkabc) \notin QM2} \approx \sum_{jkabc \in QM2}, \tag{2–32}$$

with $P(jkabc)$ a permutation operator which introduces local orbitals outside of QM2. Orbital i is defined to be in the first QM region, QM1, with $QM1 \subset QM2 \subset QM3$ giving T_3^i as,

$$T_3^i \approx \frac{1}{36} \sum_{jkabc \in QM2} t_{ijk}^{abc} a^\dagger b^\dagger c^\dagger kji \quad i \in QM1 \quad j,k,a,b,c \in QM2, \tag{2–33}$$

and T_2^i as,

$$T_2^i \approx \frac{1}{4} \sum_{jab \in QM3} t_{ij}^{ab} a^\dagger b^\dagger ji \quad i \in QM1 \quad j,a,b \in QM3. \tag{2–34}$$

The single excitation operator, T_1^i, is given as,

$$T_1^i \approx \sum_{a \in QM3} t_i^a a^\dagger i \quad i \in QM1 \quad a \in QM3. \tag{2–35}$$

Note that the QM1 regions are disjoint from one another as opposed to the QM2 and QM3 regions. This reduces the number of singles and doubles parameters and significantly reduces the number of triples parameters leading to linear or even sub-linear scaling with respect to the number of basis functions. By adopting a hybrid approach to NLSCC where NLSCCSDT-3 is performed in a small region, QM2, while NLSCCSD is performed in a larger region, QM3, it is possible to take advantage of simplifications which manifest in other hybrid methods that mix small CCSD regions with large MP2 regions. These latter hybrid schemes are based on the assumption that at large enough distances the localized electron pairs are only weakly interacting and are therefore well approximated by MP2, which does not allow these pairs to influence one another to the extent that they do within CC theory. A hybrid approach to treating triples within the NLSCC framework, as opposed to a full NLSCCSDT-3 treatment, should benefit from similar simplifications where the focus is on a small triples region in which the triples influence one another more strongly than those which are scattered over a larger region. For those systems which are well described by single reference, as opposed to a multi-reference system, it is expected that the hybrid NLSCC methods for triples would compare well with a full NLSCCSDT-3 method.

Amplitudes in QM2-QM1 have an indirect influence on those amplitudes T^i. As an example consider a QM2 region with three occupied orbitals, $i, j,$ and k and an arbitrary number of virtuals and let $i \in QM1$ and $j, k \in QM2 - QM1$. The T_2 equation will depend on $t_{ij}, t_{ik},$ and t_{jk}, where the virtual orbital labels have been dropped for convenience. It is tempting to think about dropping the t_{jk} term and representing the NLSCC wavefunction entirely in terms of $QM1 \rightarrow QM1$ and $QM1 \rightarrow QM2$ excitations, however, the $QM2 \rightarrow QM2$ excitations have an important, albeit indirect, effect on t_{ij} and t_{ik}. Obviously those outside of the largest QM region defined do not make a contribution and thus are neglected within NLSCC. This choice of QM regions should be done self-consistently to ensure that any important longer-range physics is included

and an automated QM determination is currently underway. The regions are picked to minimize tail contributions. In our current implementation these regions are picked using Molden [30] based on results from calculations performed on substituted alkanes of varying lengths.

In the limit of perfect localization where $|0\rangle = |i\rangle|j\rangle|k\rangle\cdots$, the NLSCCSDT-3 wavefunction becomes a product, by virtue of its size-extensive exponential ansatz,

$$|\Psi\rangle = |\Psi_i\rangle|\Psi_j\rangle|\Psi_k\rangle\cdots . \tag{2–36}$$

Due to the orthogonality tails in the orthonormal NLMO basis there will be higher-order terms shown in the following many-body expansion,

$$\begin{aligned} |\Psi\rangle &= |\Psi_i\rangle|\Psi_j\rangle|\Psi_k\rangle\cdots + |\Psi_{ij}\rangle|\Psi_k\rangle\cdots + |\Psi_{ik}\rangle|\Psi_j\rangle\cdots + |\Psi_{jk}\rangle|\Psi_i\rangle\cdots \\ &+ |\Psi_{ijk}\rangle\cdots . \end{aligned} \tag{2–37}$$

The total correlation energy, $E = \langle 0|(H_N e^T)_c|0\rangle = \langle 0|\bar{H}|0\rangle$,

$$E = \sum_{ia} f_i^a t_i^a + \frac{1}{4}\sum_{ijab} w_{ij}^{ab}\left(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a\right), \tag{2–38}$$

is then a sum of orbital correlation energies,

$$E = \sum_i E_i. \tag{2–39}$$

Using our truncation of excitation operators gives $E_i = E_i(QM1) + E_i(\notin QM1)$ and so i is picked to minimize $E_i(\notin QM1)$ obtaining $E_i \approx E_i(QM1)$ to a good approximation. This obviously becomes exact in the limit of QM2 being the entire molecule. If E^T is designated as the total CCSDT-3 correlation energy then Eqn. 2–39 gives,

$$E^T = \sum_i E_i^T, \tag{2–40}$$

in terms of triply correlated NLMO and/or functional group energies. The following hierarchy in terms of the QM regions is considered within the language of NLSCC,

$QM1 \subset QM2 \subset QM3$, where ideally the triples contributions to QM1 would be from as large a region as possible given by QM3. If E_i^T is written in terms of QM2 and QM3 contributions as,

$$E_i^T = E_i^T(QM3_i) + E_i^T(QM2_i), \tag{2–41}$$

an interesting hybrid approach can be developed by assuming that $E_i^T(QM3_i) \approx E_i^D(QM3_i)$ thereby avoiding longer-range contributions to the triples equation resulting in,

$$E_i^T \approx E_i^D(QM3_i) + E_i^T(QM2_i), \tag{2–42}$$

where E_i^D is a doubly correlated bond energy. Hybrid approaches are labeled as NLS(CCSDT-3/CCSD). This approximation is justified in the case of NLS(CCSD/MP2) by considering that the effects of coupled excitations that CC offers in comparison to MP2 would be less important at large distances. At large distances the localized electron pairs are only weakly interacting and so a theory in which the pairs are not allowed to influence one another, such as MP2 in contrast to CC, is appropriate.

Although this provides the energy, calculating gradients of the PES has proven more difficult within the context of local correlation methods. One question is simply if the quality of the one- and two- particle density matrices is sufficient. Although it is possible to define the gradients, there is a very expensive rectangular back transformation that becomes rate limiting, thereby significantly degrading the performance. This back transformation is necessary to avoid the derivative integral transformation that would be needed for each new geometric perturbation. In the event that this transformation is performed the subsequent forces are discontinuous because the correlation regions are geometry dependent. To remedy this it is necessary to ensure that the same orbitals are dropped in the same way for each force evaluation.

2.5 Implementation

All calculations were performed using a modified version of the serial ACES II [16], quantum chemistry software package. With the exception of those for met-enkephalin,

results were obtained using a 375 MHz power3 processor using a maximum of 4 of the available 8 GB of shared memory over four processors. The calculations on met-enkephalin were performed on a 8 dual-core ia64 processor SGI Altix machine with 256 GB of shared memory using multithreaded libraries. Such calculations made use of some recent advances in the ACES II [16] code made to take advantage of large shared memory machines. These advances included updating some old memory limitations to allow for large reference functions (739 basis functions for the case of cc-pVDZ met-enkephalin) and in-core contractions in the context of many-body methods.

Geometries for poly-glycine were taken from reference [36]. The geometry for met-enkephalin was taken from the Protein Data Bank (1PLW.pdb, model 1) [59]. Geometries of test systems are available upon request. Given that the NLSCC method is implemented as a pilot code in ACES II [16] a manual determination of the QM regions using Molden [30] is used.

We first consider a quasi-linear translationally periodic system because for these types of systems the arguments in favor of locality are most simply understood. Poly-glycine is meant to represent the simplest peptide but still containing a variety of functional groups. For example, it has a 3-center peptide-bond. Met-enkephalin is meant to represent a more realistic application because it is three-dimensional and has non-trivial side-chains. It is a conformationally flexible neuropeptide with a high affinity for the opiate receptor [59] for which there has been recent interest from the theoretical chemistry community [19, 60, 61, 62]. For poly-glycine, 2-center bonds are used, although 3-center bonds certainly could have been used, while for met-enkephalin up to 6-centered bonds are used. A cc-pVDZ basis is used unless specified otherwise, however, for more practical applications a triple-zeta or better basis should be used.

There are a number of figures which establish transferability by plotting a quantity for a functional group, for example a methyl group or a σ-bond, belonging to one side of a quasi-linear molecule as a function of its distance from a perturbation on

the opposite side of the molecule. This distance is taken to be the number of methylene groups between the two. Perturbations are taken as simple functional groups, $R = -CH_3,\ -OH,\ -NH_2,\ -F,\ -H$.

2.6 Applications

2.6.1 Motivation

It is demonstrated in Fig. 2-1 for C_1 water clusters that the NLMOs can localize the triples contribution to the CCSDT-3 wavefunction. These water clusters were chosen because they are examples of small systems which have C_1 symmetry, however, other C_1 molecules such as distorted alkanes, etc. could be used. It can be shown that the number of negligible amplitudes resulting from spatial locality is greater than the number due to symmetry. For a large symmetric molecule it is necessary that the molecule be sufficiently large so as to pass the crossover point for which spatial locality, as opposed to symmetry, dominates the number of negligible amplitudes. In the canonical basis $\approx 90\%$ of the $|T_2|$ amplitudes on average are above the 5.0×10^{-8} threshold while for the $|T_3|$ amplitudes this number reduces to $\approx 50\%$. These averages ignore the contribution from a single water molecule which because of the C_{2v} symmetry gives a crossover in the number of amplitudes above the threshold for canonical versus local bases. It is not surprising that the size of $|T_3|$ is smaller than $|T_2|$ because these clusters have relatively simple, non-degenerate ground states thereby lacking non-dynamical correlation effects usually associated with higher excitation operators. In the localized NLMO basis both $|T_2|$ and $|T_3|$ are sparse giving for example only $\approx 35\%$ and $\approx 10\%$ amplitudes above threshold, respectively for the water heptamer. The decay of localized amplitudes is monotonic for $|T_3|$ and nearly monotonic for $|T_2|$, however, due to the lack of symmetry in the clusters it is not necessarily expected that the behavior be monotonic. By virtue of the localization scale it is possible to get a handle on the dominant amplitudes.

Fig. 2-2 examines the locality of the virtual NLMOs, specifically the absolute value of diagonal two-electron integrals, $\langle ii|aa\rangle$, for a $(H_2O)_5$ calculation. The profile of the

integrals with respect to different virtual bases is shown where the key is understood as follows with the number giving the degree of localization: local 0 (canonical occupied and canonical virtual), local 1 (localized occupied and canonical virtual), and local 2 (localized occupied and localized virtual). For the localized occupied calculations $i = \sigma_{OH}$ and for the canonical occupied calculations i is delocalized. The local 0 curve shows that all the virtual orbitals give significant interactions with the chosen occupied. The local 1 curve shows that localizing the occupied orbitals does little to improve on local 0, for example a given σ_{OH} has significant interactions with all of the canonical virtual space, supporting the well known conclusion that a localized virtual space is imperative. The local 2 curve shows significant improvements on local 0 and local 1 because both the occupied and virtual spaces are localized by virtue of our NLMO implementation where the peaks represent intra-molecule interactions. The first set of peaks correspond to $\sigma_{OH} - \sigma^*_{OH}$ pair interactions, while the second and third correspond to interactions with the Rydberg spaces of oxygen and hydrogen, respectively.

The transferability of the CC wavefunction in terms of NLMOs is shown in Fig. 2-3 and Fig. 2-4 for a series of alkanes and substituted alkanes, respectively. Fig. 2-3 shows diagonal $|T_2|$ (CCSD) and $|C_2|$ (CISD) amplitudes for a $\sigma_{CH} - \sigma^*_{CH}$ pair confined to a methyl group as a function of alkane size. It is seen that because the CC wavefunction is size-extensive the CC amplitudes are transferable while the CI amplitudes are not. In Fig. 2-4 the transferability of $|T_2|$ and $|T_3|$ is shown for the CCSDT-3 level of theory. The amplitudes are confined to be on the methyl cap on the opposite side of the substituent, for example $t^{\sigma^*_{CH}\sigma^*_{CH}}_{\sigma_{CH}\sigma_{CH}}$ and $t^{\sigma^*_{CH}\sigma^*_{CH}\sigma^{*\prime}_{CH}}_{\sigma_{CH}\sigma_{CH}\sigma^{\prime}_{CH}}$ where the prime represents another σ-bond different from the first also on the methyl cap. The first point in Fig. 2-4 represents the amplitude and substituent as nearest neighbors and thus the amplitudes are very different among substituents. T_2 requires only one region of screening and therefore rapidly approaches a constant transferable value, while T_3 from a CCSDT-3 calculation requires two or more regions of screening before becoming transferable. Note that these results are similar over

all functional groups studied, including the very electronegative fluorine atom, because of the robust transferability of the CC amplitude equations as represented in terms of NLMOs.

The CCSD (left) and CCSDT-3 (right) σ_{CH}-bond energies computed from these transferable amplitudes using Eqn. 2–18 are themselves transferable as shown in Fig. 2-5 for the same series of substituted alkanes. The difference between CCSD and CCSDT-3 bond energies is, not surprisingly, small, $\approx 1\ mH$, given the simplicity of the electronic structure. Both CCSD and CCSDT-3 bond energies require two or more regions of screening before reaching a constant value and the rate at which they become transferable are nearly identical.

2.6.2 Poly-Glycine

Differences between CCSD and CCSDT-3 bond energies are shown in Fig. 2-6 for two molecules used in the NLSCC calculation of poly-glycine. The two molecules are designated by the fact that they are either the N-terminus or C-terminus and are shown in Fig. 2-7 on the lower-left and lower-right respectively. There are negligible triples contributions from the core orbitals and small $\approx 1.5\ mH$ on average contributions for the σ-bonds. The lone-pairs in these two molecules have larger triples contributions $\approx 2\ mH$, on average, and they are over a longer-range than the other bonds, $\approx 1\ mH - 3\ mH$. This is probably because they are more dependent on local environment due to their greater diffusivity. The occupied indices within each of the core, lone-pair, σ-bond, and π-bond sections are not in a specific order. The triples contribution from the π-bonds is largest, $\approx 4\ mH$, suggesting an interesting active space triples method based on NLSCC in which different bonds are correlated at different levels of theory.

Fig. 2-7 shows the QM regions defining the hybrid NLSCC calculation of poly-glycine. The functional groups of interest represented by the QM1 region, the QM label is to specify that this region is treated quantum mechanically as opposed to by more classical interactions, for example by an ESP, etc., which could be incorporated into regions outside

of QM3, but will not in the current formulation. There is one region for the methylene group, one for the peptide group, one for the N-terminus, and one for the C-terminus. This region is embedded in the QM2 region which will be treated at the CCSDT-3 level of theory, which in turn is embedded in the QM3 region to be treated at the CCSD level of theory. The reason for two embedding regions is to capture any strong correlations for those functional groups which are close to one another with CCSDT-3 and for weaker correlations between more distant functional groups CCSD can be used. The peptide-bond in poly-glycine is treated as a Kekulé structure composed of 2-center bonds.

The CCSD and CCSDT-3 functional group energies of the molecules used in the NLSCC calculations on poly-glycine are shown in Table 2-1. The diagonal elements are simply the energies taken from the corresponding QM1 regions of Fig. 2-7 while the upper-diagonal elements for peptide and methylene are the energies of those groups extracted from both the calculations for which the target was the N-terminus and C-terminus. The difference between CCSD and CCSDT-3 functional group energies is largest for the C-terminus, $\approx 35\ mH$, followed by the N-terminus and the average peptide group which are $\approx 24\ mH$ and $\approx 14\ mH$ respectively. The smallest difference is for the average methylene group which is $\approx 6\ mH$. This ordering follows from the total number of lone-pairs and π-bonds in each of the groups which from Fig. 2-6 are shown to give large triples corrections to the energy. The transferability of the peptide group among three different molecules is seen by considering the upper-diagonal elements which are within $\approx 1\ mH$ of one another for both CCSD and CCSDT-3. The same is true for the methylene group from the CCSD calculation, however, for the CCSDT-3 calculation the differences are slightly larger.

The total and unit cell NLSCCSD and NLS(CCSDT-3/CCSD) correlation energies for poly-glycine are shown in Table 2-2 along with some literature results obtained using local CC methods [43, 55]. The unit cell is glycine. The triples contribution acts to lower the correlation energy contribution per unit cell by $\approx 20\ mH$ for all poly-glycine examples including the infinite limit. The saturation of the correlation energy per unit

cell is clearer in Fig. 2-8. The energy difference for the infinite versus small molecule limit for the glycine unit cell is $\approx 30\ mH$ for both NLSCCSD and NLS(CCSDT-3/CCSD). For NLSCCSD over 99% of the correlation energy is obtained when compared to conventional results for the glycine monomer, dimer, and tetramer while for NLS(CCSDT-3/CCSD), the comparison with conventional methods was only possible with the glycine monomer for which 99.4% was recovered. In order to validate the NLSCC triples methods further comparison with conventional methods would be necessary. Despite the fact that a direct comparison between the NLSCC and local CC methods is not possible, because of the frozen core approximation in addition to the fact they are hybridized differently, the NLSCC method provides the literature results for extended systems. As an example of time savings consider $(gly)_1$ with 95 basis functions and $(gly)_{16}$ with 1160 basis functions. The time required for a CCSDT-3 calculation on $(gly)_1$ is 0.95 *days* which can be crudely extrapolated to 3.4×10^7 *days* for $(gly)_{16}$. The NLS(CCSDT-3/CCSD) serial calculation on $(gly)_{16}$ took 11.9 *days* to be compared to a parallel FMO CCSD(T) on four 3.2 GHz pentium4 processors which took 9.9 *days* [25]. By virtue of the fact that the NLSCC methods are built upon local QM regions the rate determining step of the calculations becomes that of the largest QM region. For a simple translationally periodic example once the energies of the QM1 regions shown in Fig. 2-7 are determined it is trivial to reuse these regions in symmetrically equivalent cases to determine the energies of larger and larger molecules. More timings are discussed in the section on met-enkephalin.

2.6.3 Met-Enkephalin

The three-dimensional penta-peptide (met-phe-gly-gly-tyr) met-enkephalin is shown in Fig. 2-9 along with a crude representation of the QM1 regions chosen for the NLSCC calculations. Side-chains of non-glycine residues methionine ($QM1_1$), phenylalanine ($QM1_4$), and tyrosine ($QM1_9$) are shown in Fig. 2-10 along with atom indices for use in Table 2-3.

Table 2-3 shows a decomposition of the CCSD correlation energy into functional group and NLMO contributions for non-glycine QM1 regions which are shown in Fig. 2-10. The NLMOs are ordered per column as core, lp, σ, and π orbitals as well as a 6-center phenolic-bond given by $1\pi_6 - 3\pi_6$. All NLMO core contributions are $\approx 1 - 2\ mH$ while the average correlation energy for an oxygen lp ($\approx 50\ mH$) is $\approx 5\ mH$ greater than a sulfur lp ($\approx 45\ mH$). The average correlation energy for a σ_{CH}-bond in methionine, $\approx 49\ mH$, is much greater than for phenylalanine and tyrosine which are both around $\approx 39\ mH$ because of the similarity in their side-chains. The average correlation energy for a σ_{CC}-bond in methionine is $\approx 56\ mH$ compared to $\approx 44\ mH$ for a ring σ_{CC} and $\approx 47\ mH$ for a non-ring σ_{CC} in phenylalanine and tyrosine. The average σ_{CS}-bond in methionine has $\approx 54\ mH$ in comparison to the average σ_{CC}-bond which is $\approx 56\ mH$. The σ_{OH}-bond in tyrosine has $\approx 53\ mH$. The delocalized 6-centered π-bonds contain the most correlation amounting to an average of $\approx 64\ mH$ for both phenylalanine and tyrosine. The similarity in the bond correlation energies for phenylalanine and tyrosine are attributed to the transferability of their electronic structure within NLSCC as represented in terms of NLMOs.

Table 2-4 summarizes minimal-basis and cc-pVDZ NLSCCSD calculations for met-enkephalin. On average each QM2 region contains 29 atoms and 53 occupied orbitals with 32 virtuals for STO-3G and 218 virtuals for cc-pVDZ. The largest QM calculations with 65 occupied and 272 virtuals are for $QM2_4$ and $QM2_9$ for which the phenylalanine-tyrosine interaction is calculated. The NLSCCSD result can be constructed out of effective functional groups as determined from the QM1 regions giving a correlation energy of $-2.4905\ H$ and $-5.7742\ H$ for STO-3G and cc-pVDZ bases respectively. For STO-3G conventional CCSD gives $-2.4987\ H$ and thus 99.7% of the correlation energy is recovered. Although this is the simple minimal-basis it is possible to get the result for a calculation with 87 virtual orbitals from calculations with no more than 42 virtuals thus saving a factor of 18 in the evaluation of $\langle ab|cd\rangle$ integrals. A similar comparison using

the cc-pVDZ basis, 587 virtuals, is currently under development using a new parallel version of ACES, ACES III [26], where similar savings should be expected given that the largest QM2 has only 272 virtuals. As a simple illustration of the time saved in the NLSCC method consider the STO-3G conventional CCSD calculation which took a total of 53.56 *hours* in comparison to the NLSCCSD calculation which took 0.98 *hours* for the rate limiting step. The other steps are shorter and can be done simultaneously. The cc-pVDZ NLSCCSD calculation on met-enkephalin had a rate limiting step of 137.64 *hours* in which the NLMO generation took only 69 *seconds*. Conventional HF provides reference energies although other reference functions for example DFT or FMO-HF [19] could be used. Collective three-body weak interactions from methionine, phenylalanine, and tyrosine were not necessary in the STO-3G basis but may be needed in larger and more diffuse basis sets. Accurate triples contributions for met-enkephalin are too intensive to calculate because the QM2 regions should contain at least amino acid dimers and these are already too expensive for a proper correlated treatment. Active space triples methods for aromatic bonds built upon conventional NLSCC would be quite useful in this situation.

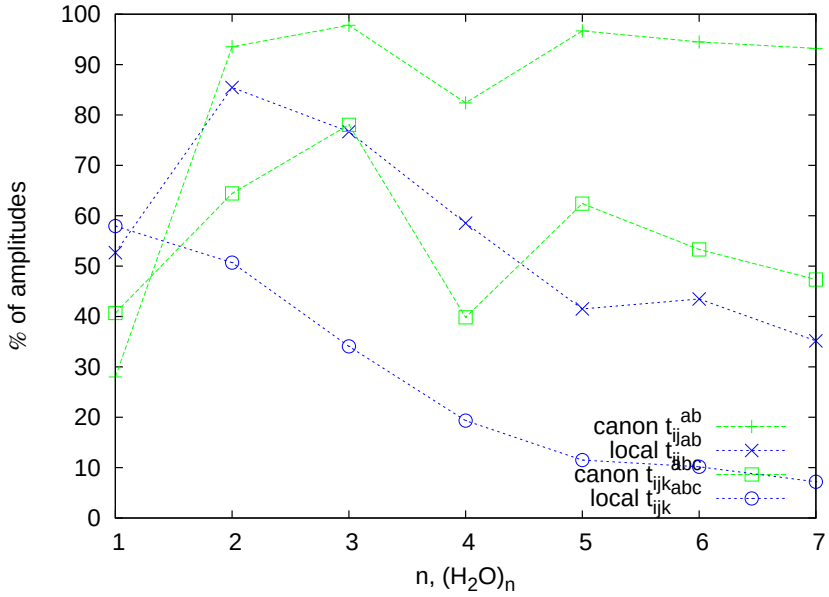

Figure 2-1: Increasing sparsity of the CCSDT-3 amplitudes with system size. The percentage of canonical (canon) and NLMO (local) based CCSDT-3 $|T_2|$ and $|T_3|$ amplitudes which are above the 5.0×10^{-8} threshold is shown as a function of C_1 water clusters of varying sizes.

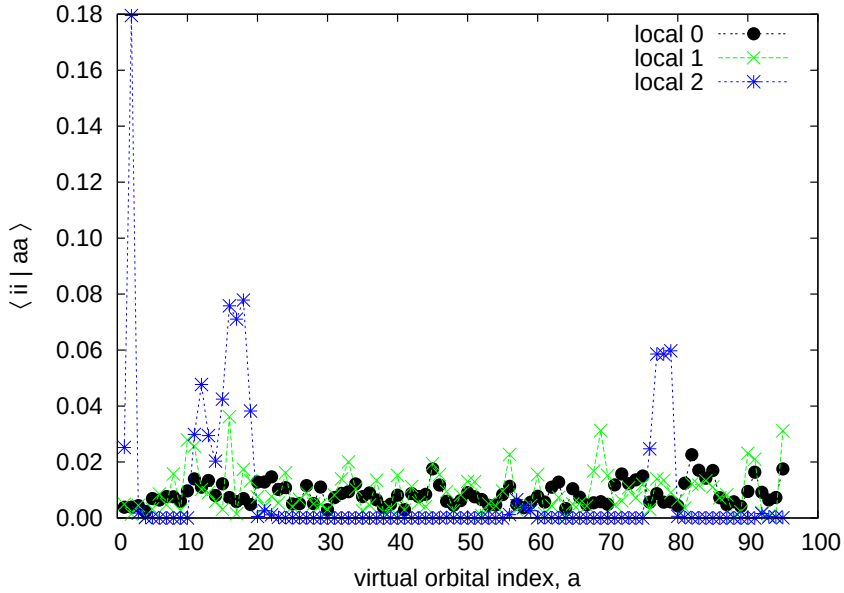

Figure 2-2: Sparsity of two-electron integrals in a localized virtual space. A profile of the absolute value of diagonal two-electron integrals, $\langle ii|aa\rangle$, is shown with respect to different virtual bases from a C_1 $(H_2O)_5$ calculation. The key is understood as follows with the number giving the degree of localization: local 0 (canonical occupied and canonical virtual), local 1 (localized occupied and canonical virtual), and local 2 (localized occupied and localized virtual).

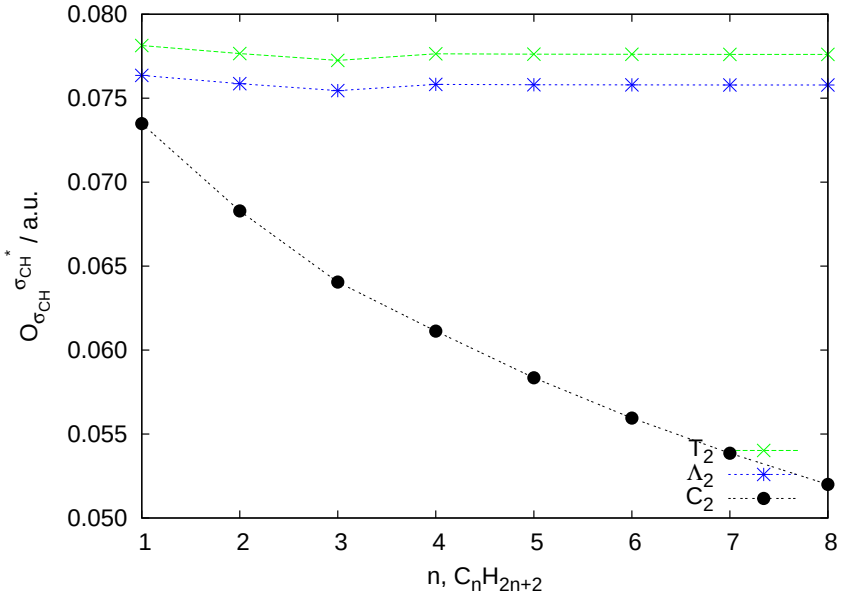

Figure 2-3: Transferability of local amplitudes from size-(in)extensive methods. Diagonal T_2 (ground state CC), Λ_2 (ground state CC), and C_2 (ground state CI) amplitudes for a σ_{CH}-bond belonging to a methyl group as a function of alkane size. The plot demonstrates that for the size-extensive CC wavefunction the amplitudes are transferable, while for the CI wavefunction, which is size-inextensive, they are not. Despite the fact that Λ_2 is a CI-like de-excitation operator it is approximately size-extensive within the current context.

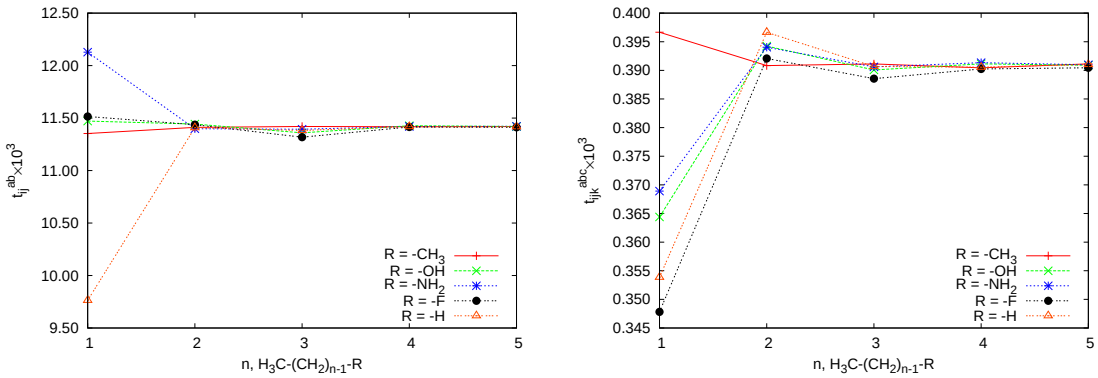

Figure 2-4: Transferability of $|T_2|$ and $|T_3|$ from CCSDT-3 in NLMOs. The amplitudes are confined to the methyl cap on the opposite side of the substituent and the x-axis represents increasing distance between the two in terms of the number of methylene groups belonging to the substituted alkane.

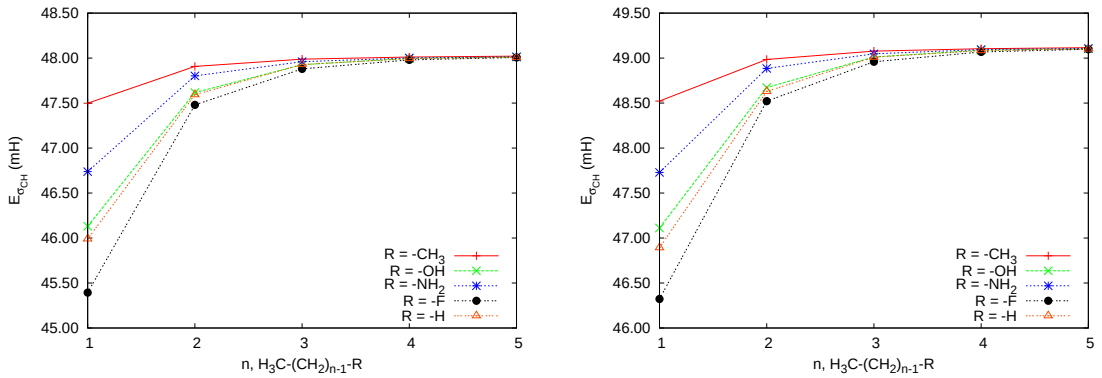

Figure 2-5: Transferability of CCSD and CCSDT-3 bond energies in NLMOs. The σ_{CH}-bond is on the opposite side of the substituent and the x-axis represents increasing distance between the two in terms of the number of methylene groups belonging to the substituted alkane. Absolute valued bond energies are shown.

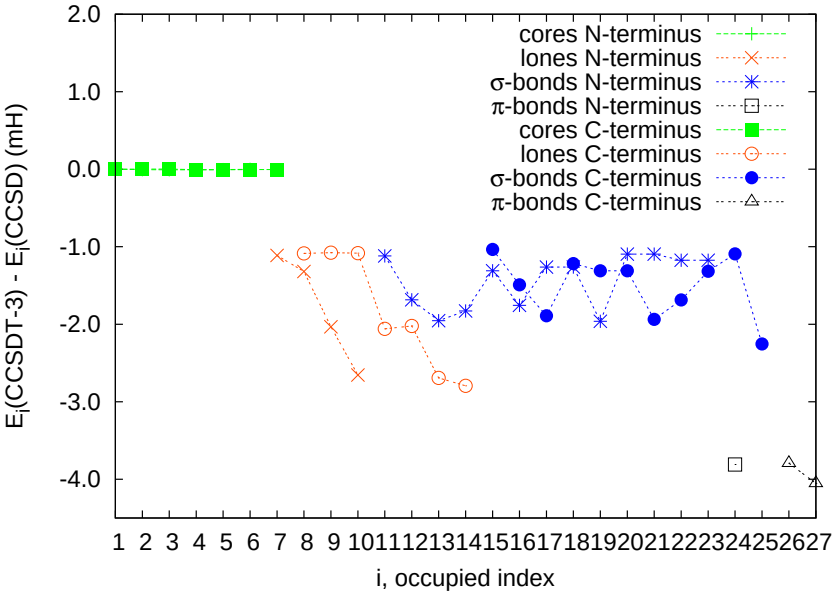

Figure 2-6: Effect of triple excitations on calculated bond energies. Difference between CCSD and CCSDT-3 bond energies for all of the occupied orbitals in two molecules used in the NLSCC calculation of poly-glycine. The two molecules are designated by the fact that they are either the N-terminus or C-terminus and are shown in Fig. 2-7 on the lower-left and lower-right respectively.

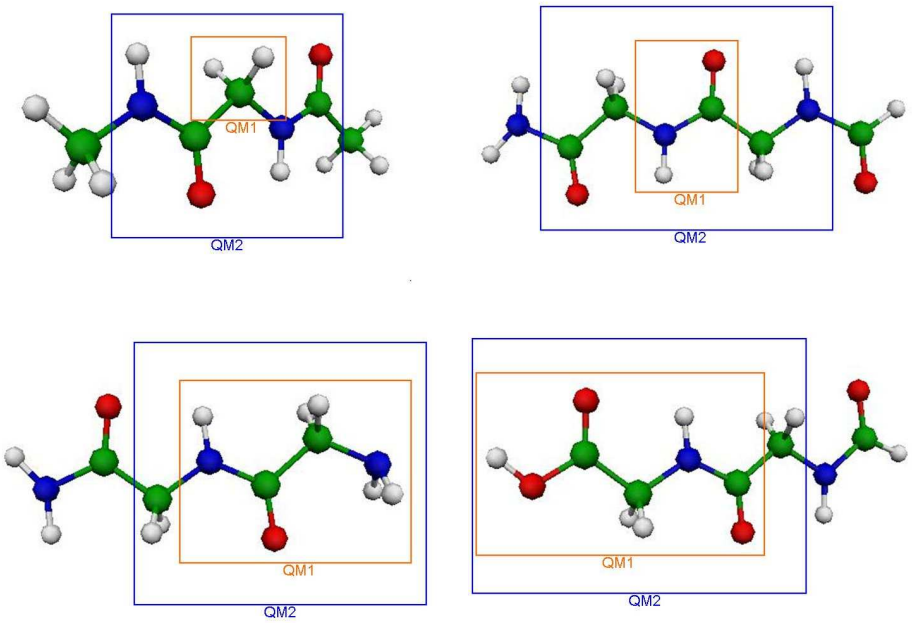

Figure 2-7: Definition of the regions for the NLSCC calculation of poly-glycine. The region QM1 simply represents the functional groups of interest, QM2 represents the CCSDT-3 region, and QM3, the whole fragment, represents the CCSD region. Due to the more complicated structure there are four calculations (methylene group, peptide group, N-terminus group, and C-terminus group) needed.

Table 2-1: Transferability of bond energies from different molecules. The diagonal elements are simply the energies taken from corresponding QM1 regions of Fig. 2-7 while the upper-diagonal elements for peptide and methylene are the energies of those groups extracted from both the N-terminus and C-terminus calculations. Functional group energies are reported in units of Hartree.

CCSD				
	C-terminus	N-terminus	peptide	methylene
C-terminus	-1.1039		-0.4184	-0.2129
N-terminus		-0.7914	-0.4171	-0.2123
peptide			-0.4173	
methylene				-0.2146
CCSDT-3				
	C-terminus	N-terminus	peptide	methylene
C-terminus	-1.1388		-0.4331	-0.2193
N-terminus		-0.8150	-0.4308	-0.2185
peptide			-0.4311	
methylene				-0.2212

Table 2-2: NLSCC correlation energies for poly-glycine. The unit cell is a glycine. Other values are taken from the local CC literature [43, 55]. The symbol H means that the energies are reported in units of Hartree.

$n_{glycine}$	NLSCCSD		LCCSD/ LMP2 FC	NLS(CCSDT-3/CCSD)		LCCSDT-1b/ LMP2 FC	LCCSD(T)/ LMP2 FC
n_{cell}	E_{corr} (H)	E_{corr}/n_{cell}	E_{corr} (H)	E_{corr} (H)	E_{corr}/n_{cell}	E_{corr} (H)	E_{corr} (H)
1	99.4%	-0.8466	99.1%	99.4%	-0.8703	98.9%	99.0%
2	99.9%	-0.7387	99.0%	-1.5212	-0.7606		
4	99.3%	-0.6854	-2.6783	-2.8262	-0.7066	-2.7552	-2.7570
7	-4.6376	-0.6625		-4.7838	-0.6834		
8	-5.2696	-0.6587	-5.1499	-5.4363	-0.6795	-5.3001	-5.3039
9	-5.9017	-0.6557		-6.0888	-0.6765		
10	-6.5338	-0.6534	-6.3857	-6.7413	-0.6741		
11	-7.1658	-0.6514		-7.3938	-0.6721		
14	-9.0620	-0.6473	-8.8574	-9.3513	-0.6680		
15	-9.6941	-0.6462		-10.0038	-0.6669		
25	-16.0147	-0.6405		-16.5289	-0.6611		
41	-26.1277	-0.6372		-26.9691	-0.6577		
61	-38.7690	-0.6355		-40.0193	-0.6560		
163	-103.2395	-0.6333		-106.5753	-0.6538		
265	-167.7100	-0.6328		-173.1313	-0.6533		
500	-316.2450	-0.6324		-326.4711	-0.6529		
1000	-632.2769	-0.6322		-652.7261	-0.6527		
∞		-0.6321			-0.6525		

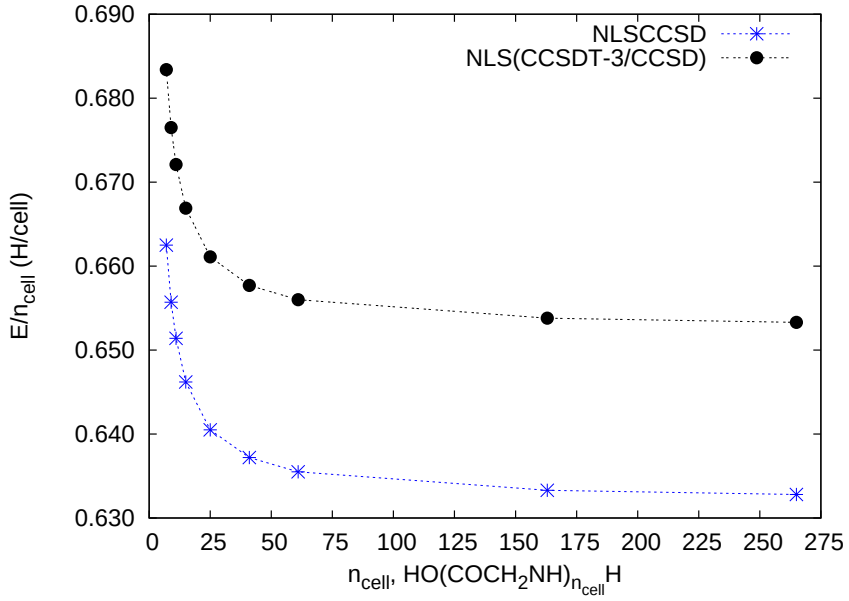

Figure 2-8: NLSCC correlation energy per unit cell of poly-glycine. The unit cell is glycine and the energies are shown absolute valued.

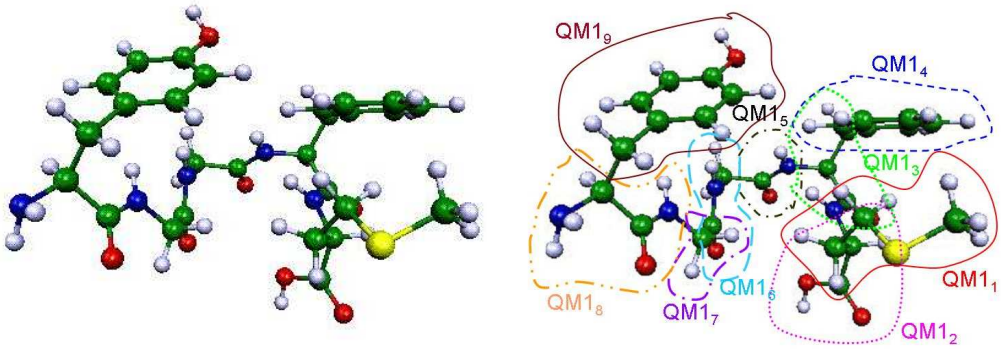

Figure 2-9: Model of the three-dimensional penta-peptide met-enkephalin. The coordinates (1PLW.pdb) were taken from reference [59]. Also shown is a crude representation of the QM1 regions chosen for the NLSCC calculations.

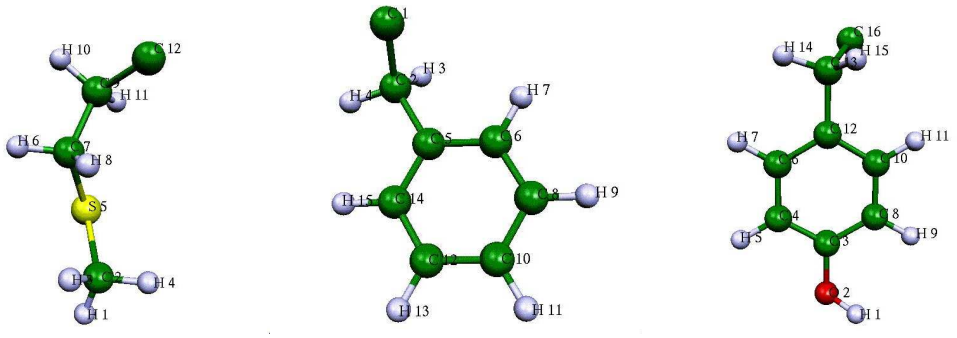

Figure 2-10: Model showing the side-chains of non-glycine residues. Atom indices are included for use in Table 2-3.

Table 2-3: CCSD bond correlation energies for non-glycine residues. These residues are shown in Fig. 2-10. NLMOs are ordered per column as core, lp, σ, and π orbitals as well as a 6-center phenolic-bond given by $1\pi_6 - 3\pi_6$. For each type of NLMO within each column the energy contributions are ordered by increasing magnitude. Energies are reported in units of Hartree.

methionine		*phenylalanine*		*tyrosine*	
$S5$	-0.0000	$C5$	-0.0017	$O2$	-0.0014
$S5$	-0.0013	$C8$	-0.0017	$C3$	-0.0016
$S5$	-0.0013	$C10$	-0.0017	$C6$	-0.0017
$S5$	-0.0016	$C12$	-0.0017	$C12$	-0.0017
$S5$	-0.0017	$C14$	-0.0018	$C4$	-0.0018
$C7$	-0.0018	$C6$	-0.0018	$C8$	-0.0018
$C2$	-0.0019	$C2$	-0.0019	$C10$	-0.0018
$C9$	-0.0020	$C10H11$	-0.0385	$C13$	-0.0019
$S5$	-0.0409	$C12H13$	-0.0386	$O2$	-0.0450
$S5$	-0.0480	$C8H9$	-0.0389	$O2$	-0.0548
$C2H1$	-0.0473	$C14H15$	-0.0393	$C6H7$	-0.0391
$C2H4$	-0.0474	$C6H7$	-0.0397	$C4H5$	-0.0391
$C2H3$	-0.0479	$C2H3$	-0.0401	$C8H9$	-0.0392
$C7H6$	-0.0492	$C2H4$	-0.0404	$C10H11$	-0.0393
$C9H10$	-0.0493	$C10C12$	-0.0431	$C13H14$	-0.0404
$C9H11$	-0.0495	$C8C10$	-0.0431	$C13H15$	-0.0404
$C7H8$	-0.0502	$C12C14$	-0.0434	$C3C4$	-0.0431
$C2S5$	-0.0539	$C6C8$	-0.0437	$C3C8$	-0.0431
$S5C7$	-0.0548	$C5C14$	-0.0441	$C4C6$	-0.0435
$C9C12$	-0.0553	$C5C6$	-0.0443	$C8C10$	-0.0437
$C7C9$	-0.0567	$C2C5$	-0.0473	$C6C12$	-0.0439
		$C1C2$	-0.0477	$C10C12$	-0.0442
		$1\pi_6$	-0.0536	$C12C13$	-0.0471
		$2\pi_6$	-0.0699	$C13C16$	-0.0476
		$3\pi_6$	-0.0699	$O2H1$	-0.0505
				$O2C3$	-0.0547
				$1\pi_6$	-0.0537
				$2\pi_6$	-0.0675
				$3\pi_6$	-0.0714
Total	-0.6619		-0.8380		-1.0049

Table 2-4: Summary of NLSCCSD calculations on met-enkephalin. The number of atoms, occupied orbitals, virtual orbitals, and group correlation energies for each QM1 and accompanying QM2 regions according to Fig. 2-9 are shown. The NLSCCSD results are shown to be a sum of the QM1 results and compared to conventional CCSD for the STO-3G basis. Energies are reported in units of Hartree. In this case conventional HF provides reference energies although other reference functions could be used. $E_{NLSCCSD}$ is the total energy, $HF + NLSCCSD$, within the NLS framework while E_{CCSD} is the total energy, $HF + CCSD$, using conventional CC.

i, $QM1$	$QM2$				$QM1$				E^i	
	n_{atom}	n_{occ}	n_{vir}^{STO-3G}	$n_{vir}^{cc-pVDZ}$	n_{atom}	n_{occ}	n_{vir}^{STO-3G}	$n_{vir}^{cc-pVDZ}$	STO-3G	cc-pVDZ
1	27	46	29	192	11	21			-0.2486	-0.6619
2	30	55	31	221	6	15			-0.2387	-0.6519
3	29	55	34	225	6	14			-0.2244	-0.5931
4	35	65	42	272	14	25			-0.4927	-0.8380
5	23	46	25	177	4	11			-0.1758	-0.4786
6	29	56	33	224	7	15			-0.2408	-0.6348
7	22	41	25	168	3	4			-0.0644	-0.1561
8	28	52	32	214	9	18			-0.2736	-0.7549
9	35	65	42	272	15	29			-0.5315	-1.0049
$NLSCCSD$					75	152			-2.4905	-5.7742
$CCSD$					75	152	87	587	-2.4987	
HF					75	152	87	587	-2208.9502	-2236.7860
$E_{NLSCCSD}$					75	152			-2211.4407	-2242.5602
E_{CCSD}					75	152	87	587	-2211.4489	

CHAPTER 3
LOCAL APPROXIMATIONS FOR IONIZED STATES

Years of research in *ab initio* quantum chemistry has shown the importance of both non-dynamic and dynamic correlation in chemical systems. Although state-of-the-art computing facilities and parallel computing offers an advantage to the calculation time necessary to go to such high excitation levels or multiple determinants which are needed in treating non-dynamical correlation effects, the steep scaling of these methods has given rise to a variety of approximate correlation methods. Of these alternatives the local correlation approximation [63, 64, 45, 65, 36, 66, 44, 48, 67, 28, 68, 69, 29] just discussed has been shown to be quite accurate for the ground states of large molecules. Given the success of local correlation methods in treating ground states it is interesting to consider the accuracy of such approximations for electronically excited, detached, or attached states of large molecules.

3.1 Methods for Calculating Excited States

In this work the focus is on applying ionization potential equation-of-motion CCSD (IP-EOM-CCSD) to large molecules, however, because developments in calculating ionization energies are less common than developments for excitation energies the excitation energies are reviewed first. The theoretical framework of *ab initio* methods for excited states is well established by methods such as CIS [70], CC2 [71], CIS(D) [72], P-EOM-CCSD [73], SAC-CI [74], EOM-CCSD [75], etc. Although CIS is oftentimes qualitatively correct for single excitations the absence of the doubles excitation component prevents the quantitative description of single excitations. The development of methods which treat the doubles block perturbatively, like CC2 [71], CIS(D) [72], and P-EOM-CCSD [73] in a general sense, offer successively better treatments than CIS. In cases where the doubles block of the effective Hamiltonian is diagonally dominant a Löwdin partioning offers a way to incorporate the effect of doubles into the single excitation manifold. Large scale CI calculations can be performed by reducng the dimension of the CI matrix by freezing CI

coefficients [76, 77]. The EOM-CC method [75] is the most accurate and most expensive method, oftentimes comparable to the FCI for single excitations. While ground state CC is built upon a single reference, EOM-CC is built upon a multi-reference wavefunction, i.e. the CC wavefunction, and thus has the potential to describe non-dynamical correlation effects. As is known the rate limiting step in the EOM methods is the contraction of the CC effective Hamiltonian with the trial wavefunction needed in the iterative Davidson diagonalization procedure. Many of the methods for increasing the accuracy and/or applicability of excited state calculations can also be used for ionized state calculations.

3.2 Approximate Methods for Calculating Excited States

3.2.1 Localized Methods

One approach to locally correlated excited states is simply to extend local CC methods developed for the ground state to EOM-CC excited states [78, 79, 80] where the most basic, but rather strict, assumption might be that the excited and ground state orbital domains are identical. Using a pilot version of a local EE-EOM-CCSD program with virtual projected atomic orbitals it was demonstrated for small organic molecules in diffuse basis sets that part of the excited state spectrum could be recovered using local approximations [78]. To complete the spectrum the excited state orbital domains must be relaxed so as to allow for non-local or charge-transfer events. A different pilot code shows that by using larger and more state specific excitation domains, for example determined using CIS, part of the high energy non-local excitation spectrum can be recovered [79]. Although pilot versions limited applications to small molecules a production level version ought to give meaningful timings for larger molecules and provide reasonable accuracy. A more complete version of a local EOM was recently used with the cheaper and more approximate EOM-CC2 method together with density-fitting of two-electron integrals [80, 81, 82].

3.2.2 Fragmented Methods

Other approaches aim to reproduce the excitation spectrum of a large molecule by a suitable fragmentation. As an example consider the multi-layered FMO methods for CIS [83] and CIS(D) [84] as well as the fast electron correlation methods for both CIS and EE-EOM-CCSD [85] all of which are based upon FMOs. In these methods a system is considered to be composed of a chromophore in the presence of an environment region responsible for inducing shifts in the chromophoric excitation spectrum. The fast EE-EOM-CCSD method [85], complete with ESPs in the usual context of FMO methods, has been shown to be quite accurate for weakly interacting systems. One other fragment based approach to calculating excited states of molecules is the *ab initio* fragment orbital theory (AFOT) [86].

3.3 Chromophores

3.3.1 Compact Representations

Excited, ionized, or attached wavefunctions which are represented in the canonical basis of energy eigenstates offer little in the way of systematically reducing their representation to a small set of only critical elements. Recently it was shown that a SVD of the CIS wavefunction offers a compact representation in which the number of excited state wavefunction amplitudes is very small [87]. Although this approach is attractive because it offers a less ambiguous prescription, after all it is built upon a SVD and associated excited state natural orbitals [88], it offers less in terms of interpretation, and possibly performance, than methods based on localized orbitals. Analyzing the excited, as well as ionized or attached, state wavefunctions in terms of chromophores defined using localized orbitals is appealing and could be used in methods more advanced than CIS. The excited, ionized, or attached state wavefunctions can then be written in terms of a small number of determinants which are localized to a given chromophore.

In HF theory the atoms which a given canonical orbital most largely spans represents a chromophore. In IP-EOM-CCSD all canonical orbitals contribute thereby potentially

delocalizing the ionization and requiring a larger chromophore. In HF theory localized ionizations depend on the existence of a localized Koopmans' theorem which in turn depends on the locality of the canonical orbitals. In contrast in some cases the chromophore is everywhere. As a simple example consider the cyclohexane molecule, where the local symmetry requires all orbitals of a certain type, for example core orbitals, axial σ_{CH}, equatorial σ_{CH}, and σ_{CC} orbitals, to make contributions to a given ionization. Given that the ionization energy is intensive, in contrast to the total energy, this means that the magnitude of the bond ionization energies must decrease to accommodate the increasing number of bonds when building larger cycloalkanes. This prevents the transferability needed to build the ionization energy of a large molecule in terms of smaller molecules. In other language, the orbitals of the D_{3d} cyclohexane cation can not be localized into 2-center bonds. Aside from this later pathological case a linear scaling method for IP-EOM-CCSD core and valence states can be developed using chromophores determined from a lower level of theory [89].

3.3.2 Locating Chromophores with Lower Level Wavefunctions

The location of the chromophores in large systems can be determined by means of a lower level wavefunction. As a simple example, consider which localized orbital on which water molecule dominates the kth ionized state wavefunction of a water cluster. The extent to which neighboring hydrogen-bonding waters shift the ionization energy must be taken into account because the chromophore is oftentimes larger than a single water molecule. Ionized states which involve interacting yet distant chromophores, as opposed to a single chromophore, can be treated similarly by using all of the chromophores with the rest of the molecule being vacuum or an ESP. The majority of results reported herein use bond ionization energies calculated at low levels of theory to determine the chromophoric regions which are then treated at more advanced levels of theory.

Another approach based on isolating the change in electron density when a molecule is ionized is also reported herein. As is known there are a number of tools which can be

used to do this [90]. The method used here involves comparing localized orbitals for the ground state with localized orbitals from an ionized state. This approach is similar to those based on the transition density, however here the final result is a transformation matrix which gives localized orbitals for the ionized state from localized orbitals for the ground state. Unlike other methods the transformation matrix here contains correlation as well as a change in particle number. Comparing these two sets of orbitals reveals new bonding motifs which are picked up in the ionized state and also shows bonds which are consistent with the ground state. Ideally, the calculation will pick the correct chromophoric region for the IP-EOM-CCSD calculation by comparing the ground state wavefunction with the ionized state wavefunction calculated at a lower level of theory, for example IP-EOM-MP2.

Bond ionization energies from a number of different chromophores combine to provide the final ionization energies for new, more complex chromophores. How they combine is a question in developing methods for calculating the ionized, additionally excited or attached, states of large molecules. The bond ionization energies must be appropriately weighted to ensure proper representation of the environment.

3.3.3 Interacting Chromophores and Charge-Transfer

To motivate the discussion of charge-transfer effects consider the CIS Hamiltonian for a system AB with subsystems A and B far apart given by

$$\langle {}^{a}_{i} | H | {}^{b}_{j} \rangle = \delta_{ij} f_{ab} - \delta_{ab} f_{ji} + w_{jabi}. \tag{3–1}$$

Taking a diagonal element of this matrix it can be seen that it is charge-transfer separable, despite the fact that CIS is size-inextensive by virtue of its linear ansatz for the excited state wavefunction,

$$\langle {}^{a}_{i} | H | {}^{a}_{i} \rangle = f_{aa} - f_{ii} + w_{iaai}, \tag{3–2}$$

Charge-transfer separability means that the excitation energy, ω^k, goes according to the Mulliken definition as $EA - IP + J$ where EA and IP are the electron affinity and

ionization potential respecively and J is the Coulomb interaction. In the case of CIS of course the Koopmans' approximation to the EA and IP could lead to large errors in the excitation energy but the structure for charge-transfer separability is present nonetheless. Note that as AB is separating if $a \in A$ and $i \in B$ then the CIS Hamiltonian as in Eqn. 3–2 is nonvanishing even though the Coulomb piece vanishes. This means that when the CIS Hamiltonian is diagonalized for a system with a charge-transfer state that the excited state wavefunction will also contain this non-locality.

Another manifestation of this is,

$$(HC)_i^a = \omega c_i^a = \sum_{jb} H_{ia}^{jb} c_j^b = \sum_b f_{ab} c_i^b - \sum_j f_{ji} c_j^a, \tag{3–3}$$

where $(HC)_i^a = \omega \langle |C| \rangle = \langle {}_i^a | HC | 0 \rangle$. Rewritting this equation as

$$\omega c_i^a = \sum_{b \in A} f_{ab} c_i^b + \sum_{b \in B} f_{ab} c_i^b - \sum_{j \in A} f_{ji} c_j^a - \sum_{j \in B} f_{ji} c_j^a \tag{3–4}$$

and considering $a \in A$ and $i \in B$ the second and third terms vanish because of the locality of the Fock matrix. This means that excited state wavefunction amplitudes which are long-ranged are non-zero thereby complicating evaluation via local approximations. Using the same analysis for the EE-EOM methods gives,

$$\langle {}_i^a | \bar{H} | {}_i^a \rangle = \chi_{aa} - \chi_{ii} + \chi_{iaai}, \tag{3–5}$$

which doesn't provide the same charge-transfer separability because χ_{aa} and χ_{ii} are poor approximations to the correlated electron attachments and ionization potentials. To ensure charge-transfer separability, Fock space methods, which introduce a size-extensive excitation operator, such as STEOM [91] need to be used. In this case the matrix elements go as

$$\langle {}_i^a | \bar{H} | {}_i^a \rangle = EA + IP + J. \tag{3–6}$$

To the extent that charge-transfer states are accessible using EE-EOM they can be obtained within a local approximation by surrounding each occupied with its region of

excitation as is done in the conventional approximations but now adding a de-excitation region for the virtual orbital and filling the rest of the molecule with vacuum. The charge-transfer excited state wavefunction amplitudes obtained in this reduced space will be transferable to the target system by virtue of the transferability of the corresponding pieces of the effective Hamiltonian [92].

For excited states in which it might be expected that the charge-transfer determinants make negligible contributions, the pair contribution, $\bar{H}_{ia}$, might be approximated as,

$$\sum_{ia} \bar{H}_{ia} \approx \sum_{i} \sum_{a \in QM2_i} \bar{H}_{ia}. \tag{3–7}$$

This form assumes that $\mathrm{QM2}_i = \mathrm{QM2}_a$ and is particularly interesting because it systematically allows for more non-local excited states by gradually relaxing the $\mathrm{QM2}_i = \mathrm{QM2}_a$ constraint. This form simultaneously avoids reducing to a conventional calculation by defining an appropriate $\mathrm{QM2}_a$ which follows a regardless of its distance to $\mathrm{QM2}_i$.

3.4 Methods for Calculating Ionized States

In many respects the theoretical development of electron excited, ionized, and attached states is quite similar, however, there are distinct differences as a result of the change in the total number of electrons in the final wavefunction. One obvious connection is that all ionized states are preceded by a number of Rydberg excited states. The excited state wavefunctions and energies from EE-EOM-CCSD approach those obtained from IP-EOM-CCSD as the Rydberg states approach the ionization continuum. In this work true ionizations are studied as opposed to modifying the basis set of an EE-EOM to represent a continuum excitation. Additionally, ionized states, as well as attached states, are used in the Fock space STEOM [91] method, which is an EOM-CCSD method for excited states which, unlike conventional EE-EOM-CCSD, has the correct charge-transfer separability meaning that the excitation energy goes as the difference of electron attachements (EA-EOM) and detachments (IP-EOM). The IP-EOM-CCSD wavefunction should be sufficiently robust to account for orbital relaxation effects accompanying the

loss of an electron which are present in energy difference calculations. These orbital relaxation effects in IP-EOM-CCSD should approximate those obtained with a correlated calculation on the cation. Also, since the final state can be a free radical, necessary static correlation effects must also be included via building the EOM wavefunction on top of a CC wavefunction.

3.4.1 IP-EOM-CC Wavefunction

The kth ionized state wavefunction in IP-EOM-CCSD, $|\Psi^k\rangle$, is written as [93],

$$|\Psi^k\rangle = R^k|\Psi\rangle, \tag{3–8}$$

where R^k is the ionization operator given by,

$$R = R_1 + R_2 = \sum_i r_i i + \frac{1}{4}\sum_{ijb} r_{ij}^b b^\dagger ji, \tag{3–9}$$

where the k-dependence has been dropped for simplicity. $|\Psi\rangle$ is the ground state CCSD wavefunction given by,

$$|\Psi\rangle = e^T|\Psi_0\rangle, \tag{3–10}$$

where $|\Psi_0\rangle$ is the reference HF determinant and T is the excitation operator written as,

$$T = T_1 + T_2 = \sum_{ia} t_i^a a^\dagger i + \frac{1}{4}\sum_{ijab} t_{ij}^{ab} a^\dagger b^\dagger ji. \tag{3–11}$$

In the case of using a MP2 reference for the ionized state wavefunction as in IP-EOM-MP2, R_1 and R_2, are built upon first-order T_2, i.e. $T_2^{[1]}$, as opposed to both converged T_1 and T_2 as it is in the case of the CC wavefunction. As is known [94], if the EOM linear operator is not truncated then regardless of the reference function the FCI result is obtained.

3.4.2 Eigenvalue Equation in $\bar{H}$

The Schrödinger equation for the ionized state wavefunction,

$$H|\Psi^k\rangle = E^k|\Psi^k\rangle, \tag{3–12}$$

can be rewritten using the CC effective Hamiltonian, $\bar{H} = e^{-T}He^{T} = (He^{T})_c$ where c means connectedness in the usual context of many-body methods, as,

$$\bar{H}R^k|\Psi_0\rangle = \omega^k R^k|\Psi_0\rangle, \tag{3–13}$$

where ω^k is the kth ionization energy and,

$$\bar{H} = \sum_{pq} \chi_{pq}\{p^\dagger q\} + \frac{1}{4}\sum_{pqrs} \chi_{pqrs}\{p^\dagger q^\dagger sr\} + \cdots , \tag{3–14}$$

where χ_{pq}, χ_{pqrs}, etc. are the one-, two-, and higher-body intermediates. The IP-EOM-CCSD amplitude equations are found by contracting the effective Hamiltonian with the trial ionized state wavefunction as in,

$$(\bar{H}R)_i = \langle_i|\bar{H}R|0\rangle = \omega\langle_i|R|0\rangle = \omega r_i, \tag{3–15}$$

for the one-hole ionization and,

$$(\bar{H}R)^b_{ij} = \langle^b_{ij}|\bar{H}R|0\rangle = \omega\langle^b_{ij}|R|0\rangle = \omega r^b_{ij}, \tag{3–16}$$

for the two-hole-one-particle ionization. Due to the fact that in IP-EOM-CC the ionized state wavefunction follows from an eigenvalue equation in the CC effective Hamiltonian projected into the ionized sector it is like a CI method built upon a CC reference function.

3.4.3 Left-Hand Eigenstate of $\bar{H}$

As is well known the effective Hamiltonian is non-Hermitian having both left-, L^k, and right-, R^k, hand eigenvectors with identical eigenvalues, ω^k, due to the similarity transformation of H as in $e^{-T}He^{T}$. In conventional IP-EOM-CCSD the kth right-hand eigenvector is solved using a non-Hermitian generalization of Davidson's iterative diagonalization procedure using Koopmans' guesses. This eigenvector gives the kth ionized wavefunction and its eigenvalue is the ionization energy which is determined directly as opposed to via energy differences. The left-hand de-excitation eigenvector of $\bar{H}$ is needed to compute the ionized state density matrices [95] and subsequent properties, for example

gradients, polarizabilities, etc. and is given by,

$$L = \sum_i l^i i^\dagger + \sum_{ijb} l_b^{ij} i^\dagger j^\dagger b, \tag{3–17}$$

where the k-dependence has been dropped for simplicity. As is well known the left-hand ground state eigenvector, L^0, of $\bar{H}$ is simply the conventional $(1+\Lambda)$ de-excitation operator used in ground state CC gradients and density matrices. In IP-EOM-CCSD the left- and right-hand eigenvectors form a bi-orthogonal set,

$$L^{k\,\dagger} R^{k'} = \delta_{kk'}. \tag{3–18}$$

The ionization energy can then be written as a generalized expectation value,

$$\langle \Psi_0 | L^k \bar{H} R^k | \Psi_0 \rangle = \omega^k, \tag{3–19}$$

which in expanded form is,

$$\sum_{ij} L^i \bar{H}_i^j R_j + \sum_{ijka} L^i \bar{H}_i^{jka} R_{jk}^a + \sum_{ijka} L_a^{ij} \bar{H}_{ija}^k R_k + \sum_{ijklab} L_a^{ij} \bar{H}_{ija}^{klb} R_{kl}^b = \omega, \tag{3–20}$$

again with k-dependence removed. Note that for canonical HF only the first term appears and the effective Hamiltonian is the Fock matrix which is Hermitian with $L = R = 1$, i.e. the Koopmans' guess. The ionization energy is simply the orbital energy given Koopmans' theorem,

$$-\epsilon_k = -F_{kk} = \omega^k. \tag{3–21}$$

The IP-EOM-CCSD methods scales as $O(N^6)$ with N being the number of basis functions, the rate determining step of which is the $(\bar{H}R^k)_i$ and $(\bar{H}R^k)_{ij}^b$ contractions. For singly ionized states IP-EOM-CCSD is comparable to FCI.

3.5 Localized Methods for Calculating Ionized States

The number of free parameters in the IP-EOM-CCSD equations can be greatly reduced by using localized approximations similar to those methods developed for EE-EOM-CCSD. It is important to note that using localized orbitals in IP-EOM-CCSD

without any approximations reproduces the same exact ionization energies as with the canonical HF orbitals, however the states are obtained in a different order unless the Koopmans' guess for the Davidson diagonalization procedure is changed to reflect the absence of a Koopmans' theorem in a localized basis. Oftentimes, many ionized states obtained with IP-EOM-CCSD are localized to a chromophoric region of the molecule with the rest of the molecule acting as an electrostatic [96] and, to a lesser extent, correlation perturbation. The perturbation due to correlation is less of an effect given the transferability of the effective Hamiltonian. This marks the difference of two approaches, one in which only the IP-EOM-CCSD part of the calculation is local to the chromophore with the HF and CCSD parts being global and one in which the CCSD part of the calculation is local to the chromophore.

3.5.1 Locating Chromophores: Bond Ionization Energies

The NLS method for ionized wavefunctions and energies is motivated by rewriting the ionization energy as a sum of contributions from localized chemical bonds,

$$\omega^k = \sum_i \omega_i^k. \tag{3–22}$$

These bond ionization energies are used to locate the chromophoric regions of the molecule in terms of transferable chemical bonds which are then calculated at the IP-EOM-CCSD level. For the HF bond ionization energies, the canonical energy eigenstates, $\mathbf{C}$, must be rewritten in terms of the unitary transformation matrix, $\mathbf{U}$, which rotates canonical orbitals into non-canonical localized orbitals, $\mathbf{C}'$,

$$\mathbf{C}' = \mathbf{C}\mathbf{U}, \tag{3–23}$$

giving the ionization energy as a sum of bond contributions,

$$\omega^k = -\epsilon_k = -F_{kk} = -\sum_{ij} U_{ki} F'_{ij} U^\dagger_{jk} = \sum_i \omega_i^k, \tag{3–24}$$

where F is the canonical Fock matrix and F' is the non-canonical Fock matrix. The Koopmans' theorem is lost in a localized basis [97] and therefore an ionization from a single canonical molecular orbital, as it is in HF theory, becomes an ionization from a linear combination of localized molecular orbitals. In the case of IP-EOM-MP2 or IP-EOM-CCSD the bond ionization energies follow from Eqn. 3–20 represented in terms of localized orbitals and using either the MP2 or CCSD effective Hamiltonian. In all calculations except those on the density matrices, $R^\dagger$ is used instead of L. In the correlated treatment of ionization energies, as in IP-EOM-CCSD, the size of chromophores necessary could be larger than those for HF because the problem is potentially more delocal because many canonical orbitals contribute to a given principle ionization and thus all of the localized orbitals associated with those canonical orbitals.

3.5.2 Locating Chromophores: ΔE Bond Ionization Energies

To compare the bond ionization energies calculated with HF and IP-EOM-CCSD with the corresponding total energy difference calculations the following decomposition of the total energy into bond energies for the case of UHF is used,

$$\begin{aligned} E_{corr}^{UHF} &= \frac{1}{2}\sum_{i_\alpha\, j_\alpha\, a_\alpha\, b_\alpha} \langle i_\alpha\, j_\alpha \,||\, a_\alpha\, b_\alpha\rangle [\frac{1}{2} t_{i_\alpha\, j_\alpha}^{a_\alpha\, b_\alpha} + t_{i_\alpha}^{a_\alpha} t_{j_\alpha}^{b_\alpha}] \\ &+ \frac{1}{2}\sum_{i_\beta\, j_\beta\, a_\beta\, b_\beta} \langle i_\beta\, j_\beta \,||\, a_\beta\, b_\beta\rangle [\frac{1}{2} t_{i_\beta\, j_\beta}^{a_\beta\, b_\beta} + t_{i_\beta}^{a_\beta} t_{j_\beta}^{b_\beta}] \\ &+ \sum_{i_\alpha\, j_\beta\, a_\alpha\, b_\beta} \langle i_\alpha\, j_\beta \,|\, a_\alpha\, b_\beta\rangle [t_{i_\alpha\, j_\beta}^{a_\alpha\, b_\beta} + t_{i_\alpha}^{a_\alpha} t_{j_\beta}^{b_\beta}]. \end{aligned} \tag{3–25}$$

Because of the mixed spin component it is not possible to extract pure spin bond energies and so this term is simply divided in two to give the following decomposition,

$$E_{corr}^{UHF} = \sum_{i_\alpha} E_{i_\alpha} + \sum_{i_\beta} E_{i_\beta}. \tag{3–26}$$

For CCSD energy difference ionization calculations the ionization energy is given by,

$$\omega = E_{CCSD}^{(0)} - E_{CCSD}^{(+1)} = E_{HF}^{(0)} + \Delta E_{corr}^{(0)} - E_{HF}^{(+1)} - \Delta E_{corr}^{(+1)}. \tag{3–27}$$

The first and third terms give ω_{HF} and the second and fourth terms give $\Delta\omega_{corr}$. By decomposing all terms into localized bond energies it is possible to form energy difference bond ionization energies, $\omega_i = \omega_{i\,HF} + \Delta\omega_{i\,corr}$,

$$\omega_{HF} = \sum_i \omega_{i\,HF} = \sum_i E^{(0)}_{i\,HF} -^{\alpha} E^{(+1)}_{i\,HF} -^{\beta} E^{(+1)}_{i\,HF}, \tag{3–28}$$

and,

$$\Delta\omega_{corr} = \sum_i \Delta\omega_{i\,corr} = \sum_i \Delta E^{(0)}_{i\,corr} - \Delta^{\alpha} E^{(+1)}_{i\,corr} - \Delta^{\beta} E^{(+1)}_{i\,corr}. \tag{3–29}$$

3.5.3 Locating Chromophores: IP-EOM-MP2 Density Matrices

Another approach to locate the chromophores is based on the density matrix from a lower level of theory and the localized orbitals which come from this ionized state density matrix. The kth IP-EOM-MP2 density matrix [98, 92] for the whole system,

$$\gamma^k_{qp} = \langle \Psi^k | p^\dagger q | \Psi^k \rangle = \langle \Psi_0 | L^k [p^\dagger q T_2]_c R^k | \Psi_0 \rangle, \tag{3–30}$$

is back transformed to the atomic orbital basis,

$$\mathbf{P^k} = \mathbf{SC}\gamma^{\mathbf{k}}\mathbf{C}^\dagger\mathbf{S}, \tag{3–31}$$

for use in the localized orbital procedure. The kth ionized state localized NBOs, $\mathbf{C'^k_{N-1}}$ where N is the total number of electrons in the original state, are then compared to the HF ground state localized NBOs, $\mathbf{C'_N}$, using the unitary matrix, $\mathbf{U^k}$, as in,

$$\mathbf{C'^k_{N-1}} = \mathbf{C'_N}\mathbf{U^k}, \tag{3–32}$$

where $\mathbf{U^k}$ is unitary in the whole ov space. Given the structure of $\mathbf{U^k}$ it is possible to order the ionized state orbitals according to their similarity with the corresponding ground state orbitals. This provides a procedure for dropping those orbitals which the two wavefunctions, ground and ionized, have in common becasue they do not play a significant role in the kth state ionization. The remainder of the ionized state orbitals defines a

chromophoric region in the molecule for use in a IP-EOM-CCSD calculation. This region represents the location in the molecule where the density changes the most during an ionization. These local ionized state orbitals sometimes have occupations which are far from integer resulting in convergence difficulties when used in the high-level IP-EOM-CCSD calculations. This approach needs to be studied using more advanced convergence accelerators.

3.5.4 Electrostatic Potentials

The local IP-EOM-CCSD problem is divided into two categories with respect to the treatment of the localized chromophore, one in which both the ground and ionized states are treated locally and one in which the ground and ionized states come from different sized regions. Given that the ground state calculation determines the effective Hamiltonian which is transferable, the former less computationally expensive approach is used herein. The ground state CC equations are solved using the localized orbitals from a chromophore which is determined in the ESP due to the rest of the molecule. The correlation energy for this region as well as the correlated wavefunction for this region are then well defined. Subsequently, the IP-EOM-CCSD equations are solved using this regionally correlated wavefunction giving ionized state energies and wavefunctions for this region. These ionized state energies and wavefunctions are in reasonable agreement with what would have been obtained by performing the CCSD and IP-EOM-CCSD on the whole molecule and then finding that particular localized ionized state from the spectrum obtained. Due to the dependence of the iterative Davidson diagonalization method on the initial guess, it is sometimes difficult to extract this particular localized ionized state because of the large number of ionized states determined for a large molecule calculation. Tailoring the initial guess wavefunction certainly helps to find the state but it is not a guarantee.

3.6 Implementation

All calculations were performed using a modified version of the serial ACES II [16], quantum chemistry software package. Considering the size of basis sets used these results are not meant to be definitive. For benchmark results larger basis sets with higher angular momentum functions should be used. Calculations were performed using a 375 MHz power3 machine with a total of 8 GB of shared memory as well as a 8 dual-core ia64 SGI Altix machine with a total of 256 GB of shared memory. Geometries are available upon request.

3.7 Applications

3.7.1 Motivation

Consider the localized orbitals for a single water molecule obtained using the HF, CCSD, EE-EOM-CCSD, IP-EOM-CCSD, and EA-EOM-CCSD density matrices. Although the focus is on the IP-EOM-CCSD methods, results for EE- and EA-EOM-CCSD are included for completeness. The ground state uncorrelated and correlated density matrices give the usual c_O, lp_O, and σ_{OH} NBOs with slightly different occupations, for example the HF gives an occupation of 1.99 for a σ_{OH}-bond while the CCSD gives an occupation of 1.95. A calculation on the cation shows that the electron is removed from a lp_O a result which can be reproduced by the IP-EOM-CCSD calculation. The IP-EOM-CCSD calculation can alternatively break the σ_{OH}-bonds thereby giving the oxygen another lone-pair, albeit with a lower occupation, potentially accompanied by empty pair NBOs. A calculation on the anion shows that the electron is added to a σ_{HH}-bond, a new bonding moiety, a result which can be reproduced by the EA-EOM-CCSD calculation. Other times the EA-EOM-CCSD calculation results in combinations of lone- and empty-pairs as well. Similar results are obtained for EE-EOM-CCSD. Empty pairs have occupations less than bonds but greater than anti-bonds. These results are consistent with water dimers.

Tables 3-1 and 3-2 show the partial atomic charges obtained from NAO population analysis of water monomer and ten different water dimers, shown in Fig. 3-1, using HF and CCSD ground states as well as the first valence state from EE-, IP-, and EA- EOM-CCSD states in that order. For each water monomer or dimer, partial charges for the HF and CCSD ground states as well as the EE-EOM-CCSD excited states sum to zero, while the partial charges for the IP- and EA- EOM-CCSD calculations sum to one and minus one, respectively. Note that if the molecule has symmetry then the charges reflect that symmetry. The difference in charges between the HF and CCSD ground states is small in all cases except for the first three water dimer geometries which happen to be those dimers with single hydrogen-bonds. It can be seen that the hydrogen being donated undergoes significant change in charge as the correlation contributions are added.

For water monomer the first excited state has only small charges assigned to the atoms, it is almost as if the atoms retain their density as atoms in the molecule because charge reorganization is quite small. Relative to the ground state the first excited state involves charge-transfer from the oxygen to the hydrogens essentially reversing the formation of bonds needed for the ground state in the first place. For dimers with local symmetry the electron is excited from both oxygens and is distributed appropriately among the hydrogens. If the dimers lack these symmetries than one of the waters dominates the excitation showing charge-transfer from oxygen to hydrogen while the other water undergoes small changes.

The first ionized state of water monomer shows that the electron is removed from the oxygen while the hydrogen charges change little from their ground state values. Water dimers with local symmetry have the electron removed equally from each of the two oxygens while dimers with no local symmetry have the electron removed from one of the oxygens, the hydrogen-bonding oxygen if one is present. The other water stays the same. For water monomer the electron attaches itself equally among the two hydrogens with little change in the oxygen. The results for attachements of water dimers are consistent

with the other studies. For example the first three geometries involve adding the electron to the pair of hydrogens which face away from the hydrogen-bond.

Of the locally excited dimers, the singly hydrogen-bonded dimers #1, #2, and #3 and the bifurcated hydrogen-bonded dimers #7, #9, and #10, the average excitation energies are 7.5 eV and 7.7 eV. Similar results are seen for the difference in ionization energies 11.7 eV and 11.9 eV, while the attachment energies are both around 0.3 eV. The tail effects due to the other water molecule are seen by comparing with the water monomer excitation, ionization, and attachement energies of 7.6 eV, 12.6 eV, and 0.6 eV and so it can be seen that the ionization and attachement energies are shifted by more than the excitation energy.

Transferability of the one-particle, $\bar{H}_{ij}$ and $\bar{H}^{i}_{a}$, and two-particle, $\bar{H}^{a}_{ijk}$, components of the CC effective Hamiltonian which are needed in IP-EOM-CCSD are shown in Fig. 3-2 for a series of symmetric alcohols. The $\bar{H}_{ij}$ contribution, shown on the top left, is slowly convergent because it contains pieces of the Fock matrix which by virtue of its incorporation of electrostatics is much longer-ranged than typical electron correlation quantities. Despite the long-ranged pieces introduced by the Fock matrix, the most troublesome example, $\bar{H}_{lpOlpO}$, varies only little as one moves from methanol to butanol, for example the variation is less than 0.004 a.u. If the Fock matrix contribution is removed, see top right of Fig. 3-2 and note the change of ordering in the key, the effective Hamiltonian has greater transferability. The $\bar{H}^{\sigma_{CH}}_{\sigma^*_{CH}}$, shown on the bottom left is slowly convergent requiring larger alcohols to establish its rate of transferability. The two-particle piece approaches a constant in every combination of localized orbitals studied.

A valence ionized state wavefunction from IP-EOM-CCSD for these same alcohols is shown in Fig. 3-3. Both the one-hole amplitudes (left) and two-hole-one-particle amplitudes (right) are transferable. The dominant singles contribution to the ionized state wavefunction is from the n_O while the dominant doubles contribution is from the n_O and the $\sigma_{OC} - \sigma^*_{OC}$ pair.

Bond ionization energies from IP-EOM-CCSD for this valence ionized state wavefunction are shown in Table 3-3 for the same series of alcohols. The largest bond ionization energy is for an n_O being 0.4003, 0.3960, 0.3935, and 0.3921 for methanol, ethanol, propanol, and butanol, respectively. Note the convergence of this value in Fig. 3-4 as well as for the σ_{CH} and σ_{OH} contributions. The next largest bond ionization contribution is the adjacent σ_{CH}-bond followed by the σ_{OH}-bond. The large contribution of σ_{CH}-bonds is due to its symmetry with respect to n_O. All core contributions are small. The total ionization energy is given by the sum of the bond ionization energies and also is transferable being 0.3719, 0.3665, 0.3640, and 0.3624 for methanol, ethanol, propanol, and butanol, respectively. Note that assuming that the ionization is completely local is space, i.e. from the n_O alone, only results in 0.8 eV (0.03 H) error in all cases.

3.7.2 Example Using Bond Ionization Energies

For a larger example consider the radius of a water cluster, a quasi-linear $(H_2O)_8$ cluster, shown in Fig. 3-5. Changes in the HF and IP-EOM-CCSD total core ionization energies upon building this larger system are shown in Fig. 3-6. The calculations are performed by adding additional waters and performing a complete CCSD and IP-EOM-CCSD using all orbitals. Fig. 3-7 shows the transferability of n_O and σ_{OH}-bond ionization contributions to this core ionization for the left most water molecule as a function of water cluster size. It is seen that the lone-pair contribution converges much more quickly than the bond contribution.

All HF and IP-EOM-CCSD bond ionization energies for the quasi-linear $(H_2O)_8$ cluster are shown in Table 3-4 and Fig. 3-8. Table 3-4 contains the largest values which are for the left most water in the eight water cluster. For HF the localized c_O orbital makes the largest contribution 20.3624 H with the n_O next with a value of 0.0841 H while these numbers for IP-EOM-CCSD are reduced to 17.1150 H and increased to 1.1219 H respectively by virtue of adding electron correlation contributions. This is a manifestation of the fact that for HF the Koopmans' theorem establishes that the ionization energy

is simply the energy of one occupied canonical molecular orbital, which in this case is local, while for IP-EOM-CCSD the ionization energy is given by including effects from all occupied and unoccupied orbitals. In this case building in larger contributions from other orbitals thereby slightly delocalizing the ionization. Note that the n_O and σ_{OH}-bond ionization energies for both HF and IP-EOM-CCSD, values 0.0841 H, 0.0343 H, 1.1219 H, and 0.7139 H, are also shown in Fig. 3-7 for the full $(H_2O)_8$. Fig. 3-8 contains the bond ionization energies for the second through eighth water molecules in the eight water cluster. The point which was made previously regarding the delocalized nature of the IP-EOM-CCSD wavefunction with respect to the HF one is much more apparent here where the HF result is delocalized only to the second water while the IP-EOM-CCSD spreads to the third water. The final result is that an accurate core ionization for this $(H_2O)_8$ cluster can be otained with only two waters for the case of HF and three for the case of IP-EOM-CCSD. That is the uncorrelated chromophore is the first two waters while the correlated chromophore extends to the third water as well.

3.7.3 Example Using IP-EOM-MP2 Density Matrix

Chromophores which correspond to certain ionized states can be located by using the bond ionization energies determined at some level of theory. There is a certain amount of arbitrariness used in the definition of these bond ionization energies because the ground state orbitals are still being used. It is questioned whether a more robust means of locating chromophores is possible via the ionized state density matrices. If these matrices are localizable then comparison of the ionized state localized orbitals with the corresponding ground state orbitals would provide a rigorous way to define the active space to be those orbitals which are present in the ionized state which are not present in the ground state. Again this is under the assumption that the density matrices from lower levels of theory sufficiently represent the target high-level calculation. In this way it is possible to determine the location of a chromophore in more complicated systems, for example bulk water clusters. Once the chromophore and the appropriate set of orbitals is

determined the high-level calculation can be performed in a small region with little loss in accuracy.

Fig. 3-9 shows the difference between conventional and reduced space valence IP-EOM-CCSD using NBOs for water dimer at three different geometries given in Fig. 3-1. The result using NBOs determined from the HF density matrix is on the left while the result using NBOs determined from the IP-EOM-MP2 density matrix is on the right. The x-axis represents the percentage of NBOs which were dropped for the IP-EOM-CCSD calculation, zero being the conventional result in the NBO basis which also serves as a reference. Ionized state NBOs are dropped according to their similarity with the NBOs determined from the ground state, for example 25%, 50%, and 75% mean that the given percentage of orbitals shared by the ground and ionized state are dropped. Similarity of the orbitals is determined from the structure of the unitary matrix, $\mathbf{U^k}$, from $\mathbf{C'^k_{N-1}} = \mathbf{C'_N U^k}$. Note that $\mathbf{U^k}$ is unitary in the full space. All orbitals are included in constructing $\mathbf{U^k}$ and so the given percentage of orbitals dropped applies to both the occupied space and the virtual space separately. Note that 20% is the conventional dropping of core orbitals only. The HF results were produced by simply dropping the same list of orbitals as was dropped in from the IP-EOM-MP2 calculations, those of which were determined using the above procedure. As can be seen from Fig. 3-9 using HF NBOs in the IP-EOM-CCSD calculation works very well for 20% and 25% giving significantly better results than those from the IP-EOM-MP2 density matrix, however when the amount of NBOs dropped increases to 50% and 75% the error with respect to the conventional calculation increases to approximately 14 eV and 8 eV, respectively. Using the IP-EOM-MP2 density matrix, as shown on the right, keeps this error below 3.5 eV for all percentages. The three curves are closer to one another for HF than they are for IP-EOM-MP2 because the NBOs determined from the HF density for the three systems are more alike than they are for the NBOs from the IP-EOM-MP2 density. Error with respect to the conventional IP-EOM-CCSD is not monotonic. For larger more complicated

examples the IP-EOM-MP2 calculation can be carried out and a corresponding set of localized ionized state NBOs can be determined, however there are problems in converging the CC equations using this set of orbitals. Advanced convergence algorithms can be employed in CC to help convergence.

3.7.4 $(H_2O)_{13}$

Core and valence ionization chromophores determined using HF bond ionization energies are shown in Fig. 3-10, respectively, for a $(H_2O)_{13}$ cluster. In both cases the IP-EOM-CCSD, as well as HF, calculations are for a $(H_2O)_5$ chromophore which in this case is done in vacuum. The effects of an ESP due to the rest of the molecule will be considered later. The figures also show the canonical orbitals, a core and highest occupied molecular orbital (HOMO) orbital respectively, which dominate the ionization. The resultant error in the HF ionization energies is 0.6 eV. Fig. 3-11 shows the bond ionization energies of waters which are increasingly distant from the dominant water shown in mesh. The bond ionization energies for the localized core orbital are the largest contributor and are thus shown in the key for clarity. The HF contributions from all localized orbitals, with the exception of the leading localized core, are small. For IP-EOM-CCSD the other contributions from the first water are much larger. Smaller contributions can be seen from waters two and three as well. Fig. 3-10 shows another chromophore in this case the HOMO ionization. The error in the calculation for HF is increased from the error for the core chromophore to 1.1 eV given that the HOMO is less local. Bond ionization energies for this chromophore are shown in Fig. 3-11 where again the leading bond ionization energy, in this case from a lone-pair orbital on the dominant water which is in the center of the circle in Fig. 3-10, is shown in the key. For HF it can be seen that the first, fourth, and fifth waters have important lone-pair contributions to the HOMO ionization energy despite the fact that the fourth and fifth waters are more distant to the first water than

the second and third waters. This is because there is no hydrogens screening the oxygen-oxygen interaction between waters one and five. The IP-EOM-CCSD results are similar with additional contributions from the second and third waters as well.

3.7.5 Gly_5 α-Helix

Core and valence ionization chromophores in gly_5 α-helix are shown in Fig. 3-12 to be localized to one and two glycine residues respectively. This system was chosen as a test of these methods to covalently bonded systems. The error in using the chromophore of one glycine residue for the core ionization is 0.4 eV while the error in using two glycines as the chromophore for the HOMO ionization is larger at 0.9 eV. As can be seen from Fig. 3-13 the core ionization is local having the obvious contribution from the localized oxygen core shown in the key with residual contributions from only those bonds belonging to the hydroxyl group and σ_{OC}. The IP-EOM-CCSD result is similar except that it includes contributions from the carbonyl group such that the overal core excitation is localized to the carboxyl group of the glycine molecule. In Fig. 3-13 it can be seen that the HOMO ionization is dominated by two peptide groups, with larger contributions from the first peptide group, for both HF and IP-EOM-CCSD. More specifically the n_N and n_O both from the first peptide group for HF and the σ_{CO}-bond from the first peptide group and the n_N from the second peptide group for IP-EOM-CCSD. The first of each of these contributions is largest and shown both in the figure and in the key. In this case both the HF and IP-EOM-CCSD HOMO ionizations have small contributions from a number of bonds along the di-glycine molecule.

3.7.6 $(H_2O)_{28}$

Table 3-5 shows low energy and high energy valence ionized states of $(H_2O)_{28}$ calculated using IP-EOM-CCSD. Chromophores for both bulk (eight waters) and surface (seven waters) ionizations were determined using HF bond ionization energies are shown in Fig. 3-14. The first column of the table is simply the ionization energy of the central water of each chromophore, water nine for the bulk chromophore and water four for

the surface chromophore. The same data is shown in this case for the bulk and surface because the calculations are simply a water molecule in vacuum. The ionized state amplitudes show that the low energy state is dominated by a lone-pair while the higer energy state is dominated by a sigma-bond. Column two shows the ionization energy of this water now calculated in the ESP of the other twenty-seven water molecules. The bulk low energy state is shifted the most 1.2 eV up from the vacuum ionization energy. The high energy bulk state is shifted up by 0.5 eV. Both the low and high energy state for the surface states are shifted down by 0.6 eV and 1.1 eV respectively. It requires less energy to ionize the surface water than the bulk water for both low and high energy states. If the entire chromophoric region is considered, as opposed to just the central water, then the shift for the high energy bulk state is up at 0.8 eV while the lower energy surface state is up shifted by 0.4 eV and the higher energy surface state is down shifted by 1.2 eV. In this case both of the states for the surface chromophore are high in energy, one being dominated by a sigma-bond on oxygen twenty-three and the other dominated by the sigma-bond on oxygen five. Note that no low energy state was found for the bulk chromophore and the high energy state is dominated by a sigma-bond involving oxygen twenty-two.

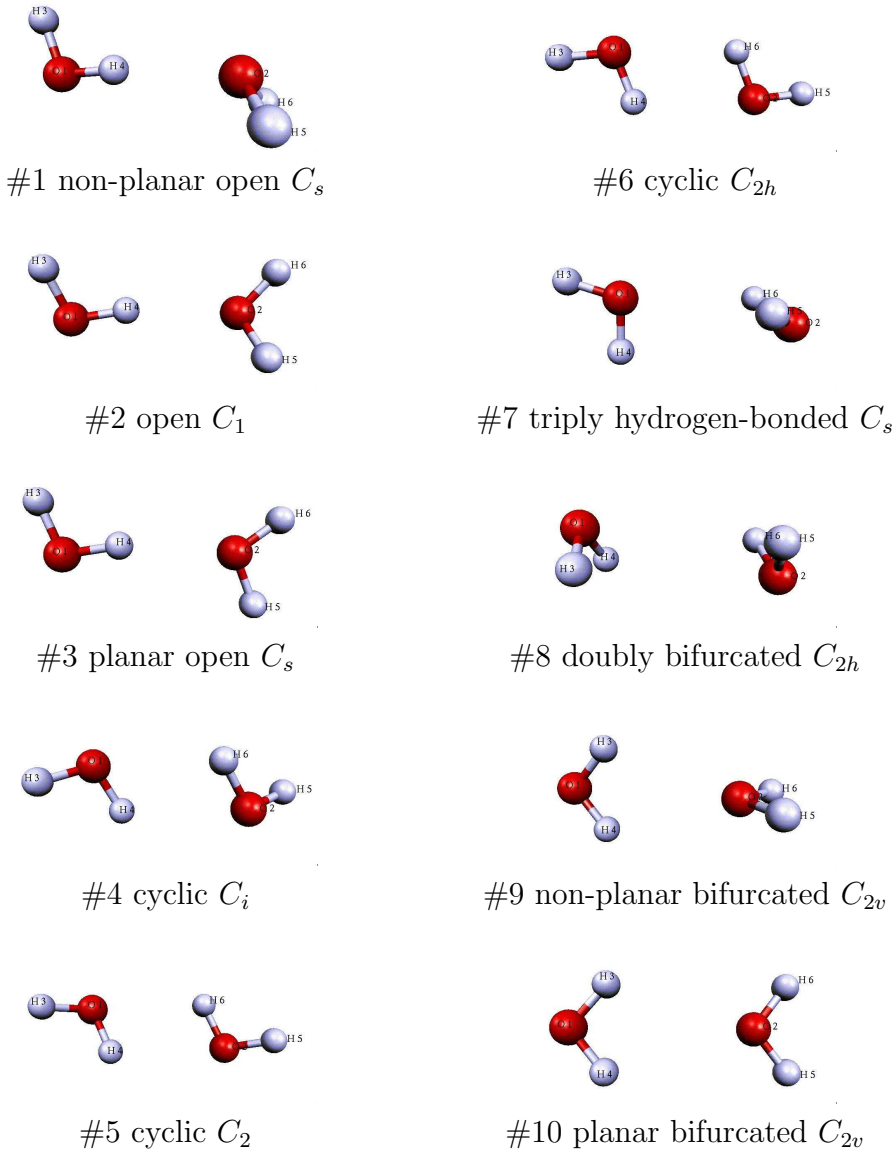

Figure 3-1: Ball and stick models of water dimer geometries. Ten different geometries are shown along with point group symmetries and other characteristics [99].

Table 3-1: Partial atomic charges for electronic states of H_2O. States include the HF and CCSD ground states as well as the first EE-, IP-, and EA- EOM-CCSD states in a aug-cc-pVTZ basis. For convenience EE-, IP-, and EA- EOM-CCSD state energies are shown in eV in place of the corresponding labels.

	HF	CCSD	7.613	12.608	0.626
O_1	-0.929	-0.904	-0.244	-0.034	-0.960
H_2	0.464	0.452	0.122	0.517	-0.019
H_3	0.464	0.452	0.122	0.517	-0.019

Table 3-2: Partial atomic charges for electronic states of $(H_2O)_2$. States include the HF and CCSD ground states as well as the first EE-, IP-, and EA- EOM-CCSD states in a aug-cc-pVTZ basis for geometries shown in Fig. 3-1. For convenience EE-, IP-, and EA-EOM-CCSD state energies are shown in eV in place of the corresponding labels.

			#1					#6		
	HF	CCSD	7.585	11.746	0.400	HF	CCSD	7.719	12.288	0.603
O_1	-0.963	-0.783	-0.230	-0.070	-0.815	-0.952	-0.927	-0.657	-0.495	-0.939
O_2	-0.944	-0.888	-0.949	-0.950	-0.931	-0.952	-0.927	-0.657	-0.495	-0.939
H_3	0.458	0.457	-0.043	0.505	0.436	0.468	0.458	0.417	0.510	0.125
H_4	0.495	0.279	0.262	0.533	0.282	0.484	0.469	0.239	0.484	0.313
H_5	0.476	0.467	0.480	0.490	0.014	0.468	0.458	0.417	0.510	0.125
H_6	0.476	0.467	0.480	0.490	0.014	0.484	0.469	0.239	0.484	0.313
			#2					#7		
	HF	CCSD	7.530	11.726	0.353	HF	CCSD	7.774	12.038	0.400
O_1	-0.958	-0.782	-0.229	-0.070	-0.814	-0.945	-0.921	-0.959	-0.958	-0.965
O_2	-0.945	-0.893	-0.943	-0.958	-0.944	-0.941	-0.914	-0.242	-0.029	-0.935
H_3	0.456	0.455	-0.064	0.505	0.437	0.471	0.460	0.477	0.498	-0.112
H_4	0.495	0.285	0.277	0.536	0.287	0.473	0.461	0.468	0.464	0.089
H_5	0.477	0.467	0.477	0.488	0.139	0.470	0.457	0.128	0.512	0.462
H_6	0.475	0.466	0.482	0.498	-0.105	0.470	0.457	0.128	0.512	0.462
			#3					#8		
	HF	CCSD	7.547	11.708	0.374	HF	CCSD	7.567	12.339	0.462
O_1	-0.958	-0.783	-0.243	-0.070	-0.812	-0.933	-0.909	-0.589	-0.470	-0.937
O_2	-0.948	-0.898	-0.936	-0.967	-0.945	-0.933	-0.909	-0.589	-0.470	-0.937
H_3	0.457	0.456	-0.050	0.506	0.436	0.466	0.454	0.294	0.485	0.218
H_4	0.494	0.289	0.272	0.538	0.293	0.466	0.454	0.294	0.485	0.218
H_5	0.478	0.468	0.474	0.490	0.152	0.466	0.454	0.294	0.485	0.218
H_6	0.477	0.467	0.481	0.502	-0.125	0.466	0.454	0.294	0.485	0.218
			#4					#9		
	HF	CCSD	7.849	12.429	0.584	HF	CCSD	7.722	11.841	0.336
O_1	-0.952	-0.926	-0.590	-0.487	-0.939	-0.942	-0.915	-0.220	-0.051	-0.942
O_2	-0.952	-0.926	-0.590	-0.487	-0.939	-0.944	-0.920	-0.959	-0.977	-0.965
H_3	0.468	0.456	0.259	0.507	0.136	0.470	0.457	0.116	0.523	0.463
H_4	0.484	0.469	0.331	0.480	0.302	0.470	0.457	0.116	0.523	0.463
H_5	0.468	0.456	0.259	0.507	0.136	0.472	0.460	0.473	0.490	-0.009
H_6	0.484	0.469	0.331	0.480	0.302	0.472	0.460	0.473	0.490	-0.009
			#5					#10		
	HF	CCSD	7.727	12.293	0.568	HF	CCSD	7.670	11.932	0.339
O_1	-0.952	-0.926	-0.594	-0.491	-0.939	-0.936	-0.910	-0.397	-0.049	-0.937
O_2	-0.952	-0.926	-0.594	-0.491	-0.939	-0.941	-0.917	-0.905	-0.965	-0.963
H_3	0.467	0.457	0.266	0.508	0.113	0.468	0.454	0.208	0.521	0.462
H_4	0.484	0.469	0.327	0.482	0.325	0.468	0.454	0.208	0.521	0.462
H_5	0.467	0.457	0.266	0.508	0.113	0.470	0.459	0.443	0.485	-0.011
H_6	0.484	0.469	0.327	0.482	0.325	0.470	0.459	0.443	0.485	-0.011

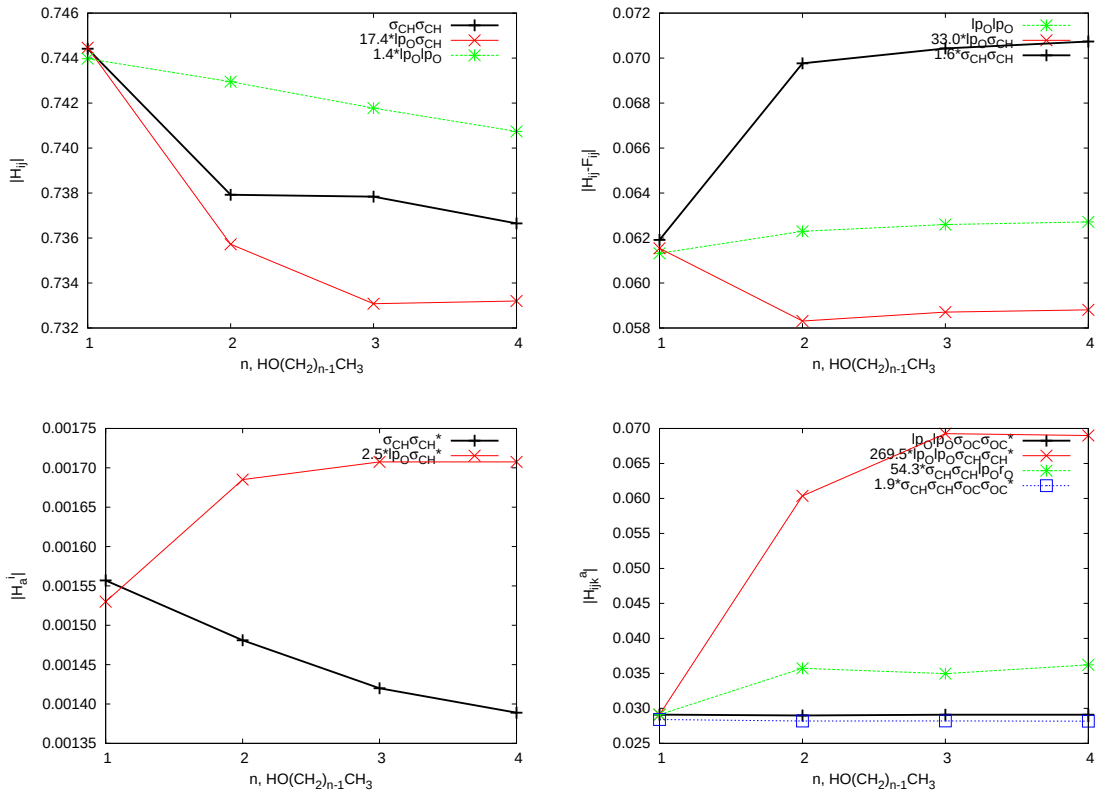

Figure 3-2: Transferability of the ionized state projection of $\bar{H}$. The CC effective Hamiltonian is shown for a series of cc-pVDZ symmetric alcohols. The one-particle elements, $\bar{H}_{ij}$ and $\bar{H}^i_a$, as well the two-particle element, $\bar{H}^a_{ijk}$, are shown where all localized orbitals, n, σ_{CH}, σ_{OC}, σ^*_{CH} and r_O, belong to the left-hand side of the molecule and are shown in the keys. Also shown is $\bar{H}_{ij} - F_{ij}$ where F is the Fock matrix. The elements are scaled by the values shown in the keys so that they are roughly the same for methanol.

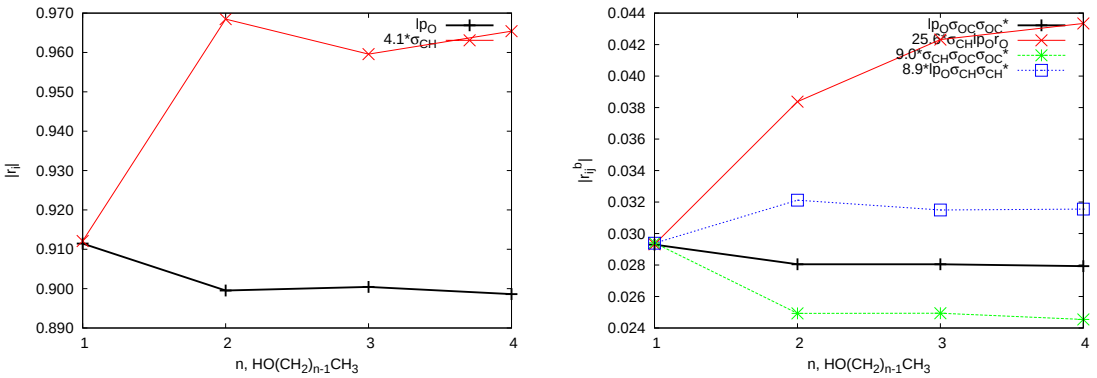

Figure 3-3: Transferability of a valence ionized state wavefunction. Calculations are for IP-EOM-CCSD on a series of cc-pVDZ symmetric alcohols. The one-hole amplitudes, r_i, and two-hole-one-particle amplitudes, r_{ij}^b, are shown on the left and right respectively, where all localized orbitals, n, σ_{CH}, σ_{OC}, σ_{CH}^*, σ_{OC}^*, and r_O, belong to the left-hand side of the molecule and are shown in the keys. The amplitudes are scaled by the values shown in the keys so that they are roughly the same for methanol.

Table 3-3: IP-EOM-CCSD valence bond ionization energies. Also shown is the corresponding total ionization energies for a series of cc-pVDZ symmetric alcohols. Moving down in the table corresponds with moving left to right in the molecule for each type of NLMO. All values are in Hartree. Note that an alternative definition of bond ionization energies is used here where $\bar{H}R^k$ elements are divided by R as seen from Eqns. 3–15 and 3–16, hence the possibility of negative values.

NLMO	$HO(CH_2)_0CH_3$	$HO(CH_2)_1CH_3$	$HO(CH_2)_2CH_3$	$HO(CH_2)_3CH_3$
c_O	0.0001	0.0001	0.0001	0.0001
c_C	0.0000	0.0000	0.0000	0.0000
c_C		0.0000	0.0000	0.0000
c_C			0.0000	0.0000
c_C				-0.0000
lp_O	-0.0086	-0.0085	-0.0085	-0.0084
lp_O	0.4003	0.3960	0.3935	0.3921
σ_{OH}	0.0014	0.0014	0.0014	0.0014
σ_{CO}	-0.0001	-0.0004	-0.0005	-0.0005
σ_{CH}	-0.0107	-0.0115	-0.0113	-0.0114
σ_{CH}	-0.0107	-0.0115	-0.0113	-0.0114
σ_{CC}	0.0002	0.0001	0.0001	0.0001
σ_{CH}		0.0002	0.0002	0.0002
σ_{CH}		0.0002	0.0002	00002
σ_{CC}		0.0001	0.0000	0.0000
σ_{CH}			-0.0000	-0.0000
σ_{CH}			-0.0000	-0.0000
σ_{CC}			0.0000	0.0000
σ_{CH}				0.0000
σ_{CH}				0.0000
σ_{CH}				0.0000
IP	0.3719	0.3665	0.3640	0.3624

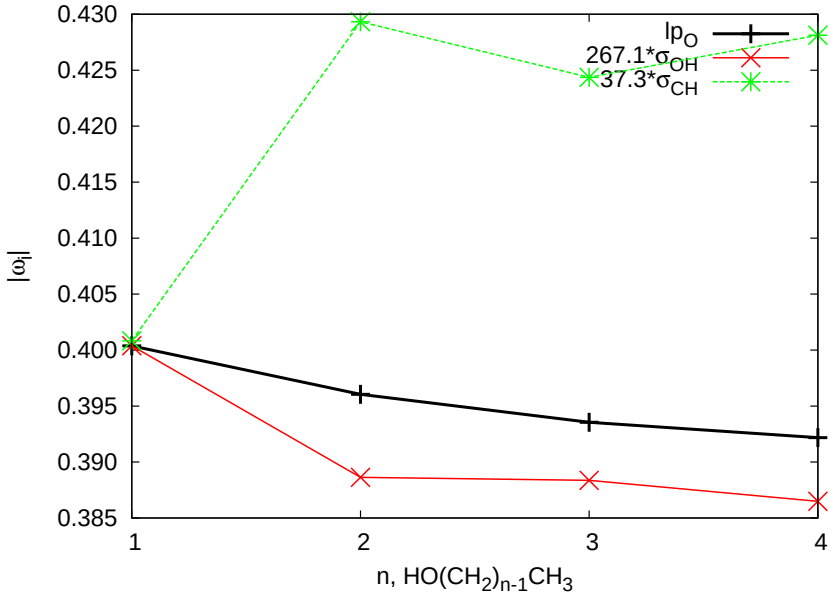

Figure 3-4: Transferability of IP-EOM-CCSD bond ionization energies. A valence ionized state wavefunction from IP-EOM-CCSD calculations was studied for a series of cc-pVDZ symmetric alcohols. All localized orbitals, n, σ_{CH}, and σ_{OH} belong to the left-hand side of the molecule and are shown in the keys. The energies are scaled by the values shown in the keys so that they are roughly the same for methanol. Hartree units are used throughout.

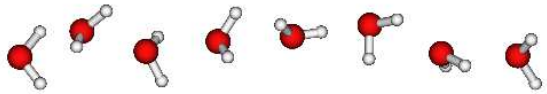

Figure 3-5: Ball and stick model of quasi-linear $(H_2O)_8$. The cluster represents the radius of a three-dimensional bulk cluster.

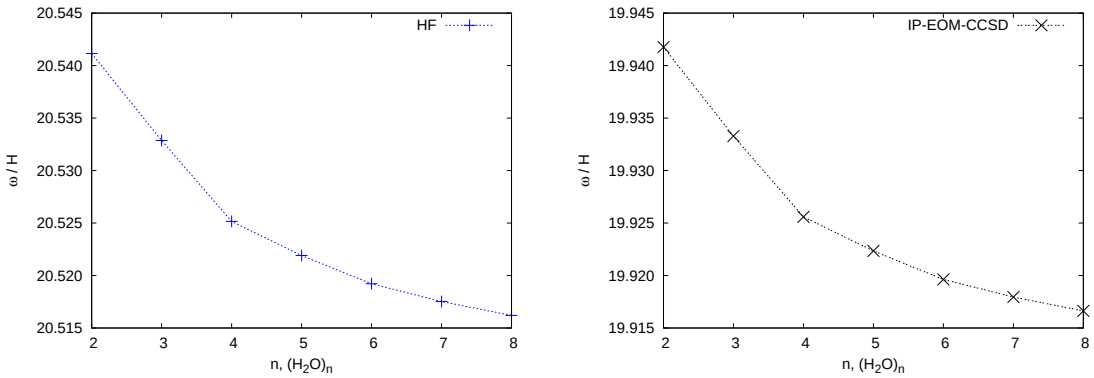

Figure 3-6: Total core ionization energies of growing water clusters. Results for HF (left) and IP-EOM-CCSD (right) are shown for clusters leading up to the quasi-linear $(H_2O)_8$ radius shown in Fig. 3-5. The calculations are performed by adding additional waters and performing a complete CCSD and IP-EOM-CCSD in the 6-31G basis set with no orbitals dropped.

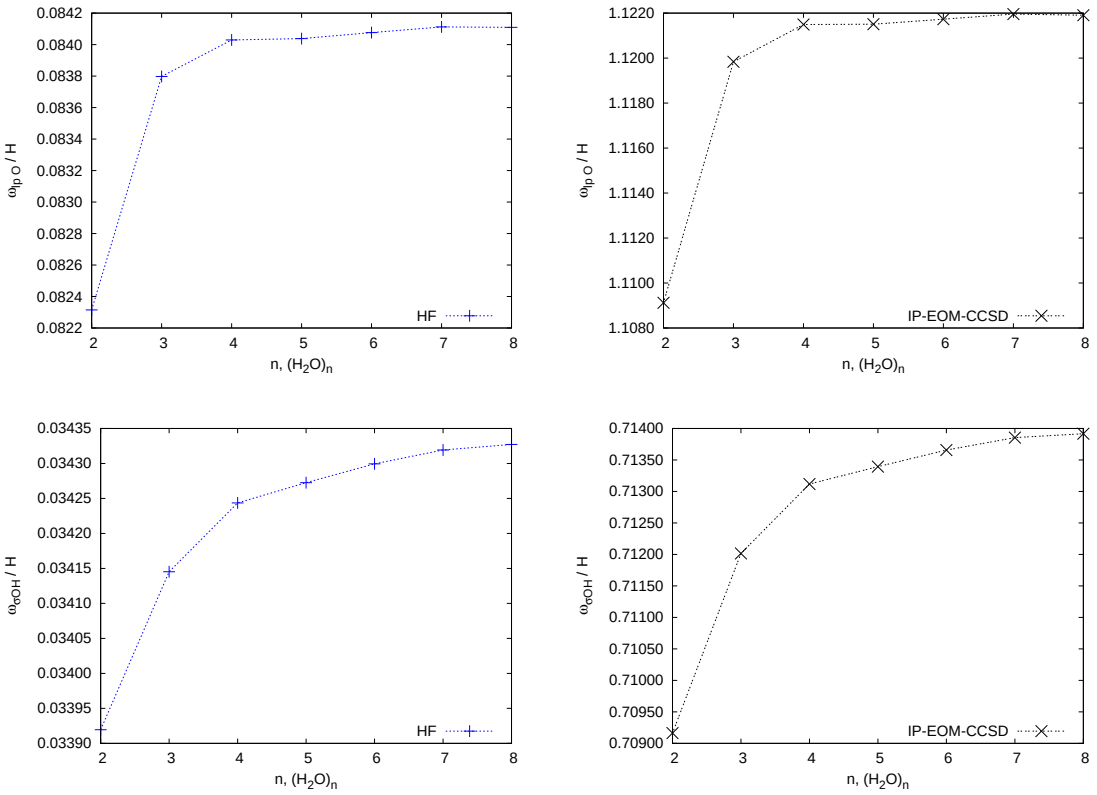

Figure 3-7: Bond ionization energy contributions of growing water clusters. Results correspond to the core ionization energies for HF (left) and IP-EOM-CCSD (right) leading up to the quasi-linear $(H_2O)_8$ cluster shown in Fig. 3-5. The calculations are performed by adding additional waters and performing a complete CCSD and IP-EOM-CCSD in the 6-31G basis set with no orbitals dropped. The transferability of the bond ionization energy contributions corresponding to the left most water are shown for n_O and σ_{OH}.

Table 3-4: The largest bond ionization energies for $(H_2O)_8$. Values are shown for both HF and IP-EOM-CCSD for the left most water contributions to the core ionization energy of the quasi-linear $(H_2O)_8$ cluster shown in Fig. 3-5. The rest of the values for the second through eighth waters are shown in Fig. 3-8. All values are in Hartree units and a 6-31G basis set was used with no orbitals dropped.

NLMO	HF	IP-EOM-CCSD
c_O	20.3624	17.1150
n_O	0.0841	1.1219
n_O	0.0048	0.2697
σ_{OH}	0.0304	0.6720
σ_{OH}	0.0343	0.7139

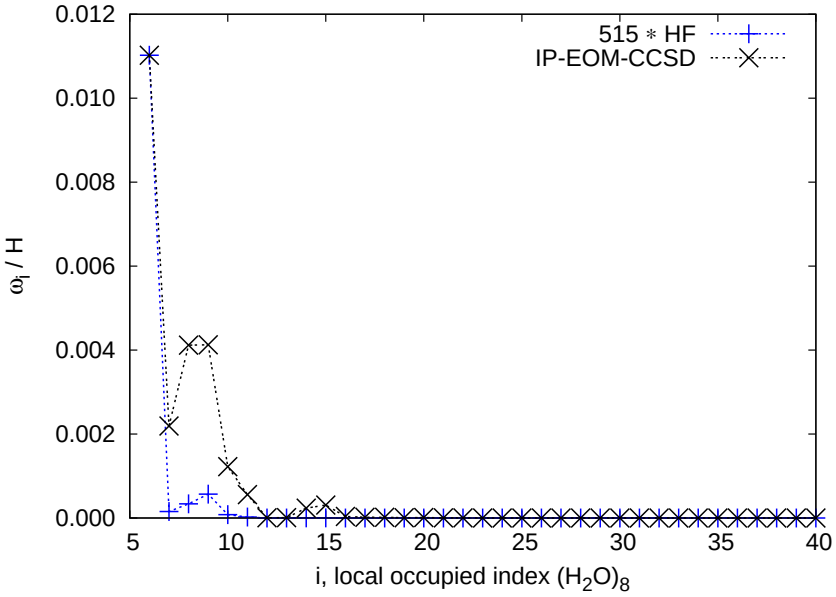

Figure 3-8: Bond ionization energies for $(H_2O)_8$. Values are shown for both HF and IP-EOM-CCSD for the second through eighth water contributions to the core ionization energy of the quasi-linear $(H_2O)_8$ cluster shown in Fig. 3-5. All values are in Hartree units and a 6-31G basis set was used with no orbitals dropped.

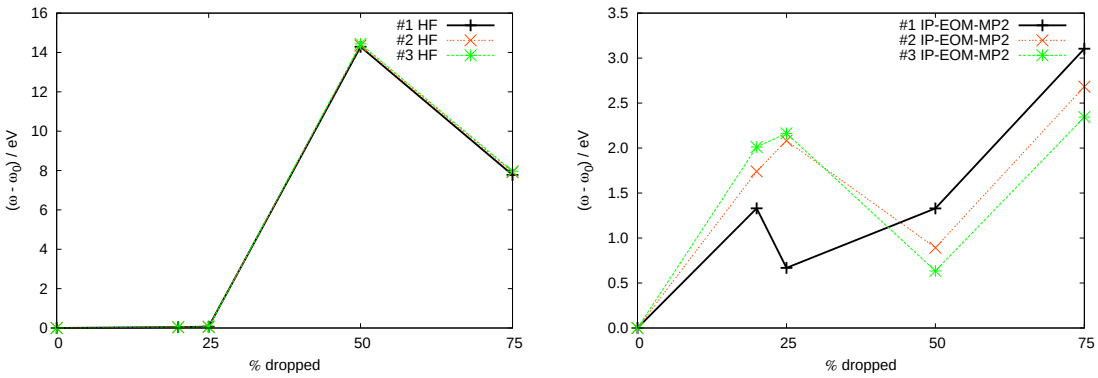

Figure 3-9: IP-EOM-CCSD ionization energies from lower level density matrices. $(H_2O)_2$ was studied at three different geometries given in Fig. 3-1 calculated using different percentages of the total number of NBOs determined from the HF (left) or IP-EOM-MP2 (right) density matrices. The x-axis indicates that the given percentage of NBOs are dropped from the occupied and separately from the virtual spaces, except for 20% which is the conventional dropping of only core occupied orbitals. The y-axis shows the difference in the ionization energy without dropping orbitals compared to dropping the given percentage. An aug-cc-pVTZ basis was used.

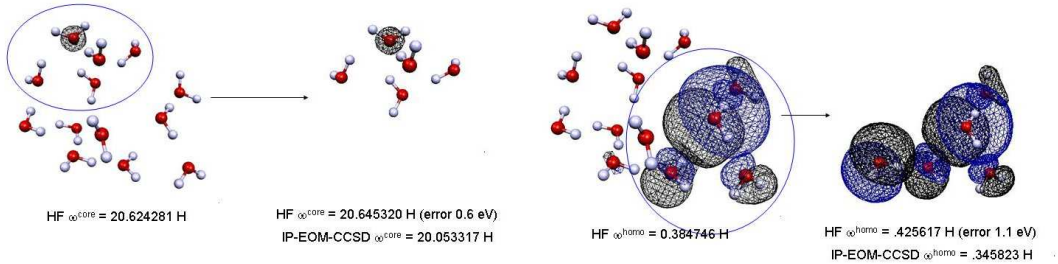

Figure 3-10: Ionization chromophores in $(H_2O)_{13}$. Core (left) and valence (right) chromophores are determined using 6-31G HF bond ionization energies. The corresponding canonical orbitals shown in circles. HF and IP-EOM-CCSD calculations on the $(H_2O)_5$ chromophore are done in vacuum as opposed to an ESP due to the other waters. The canonical core and HOMO orbitals which dominate the ionization are shown in mesh.

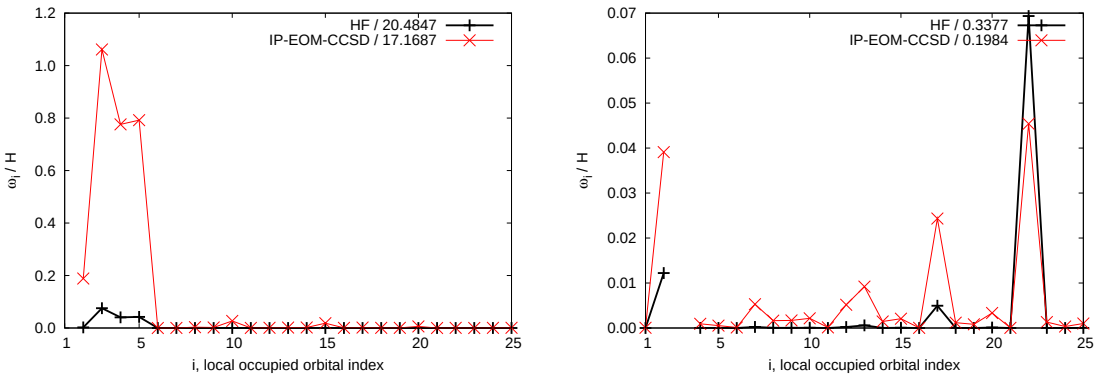

Figure 3-11: Bond ionization energies for the $(H_2O)_5$ chromophore. HF and IP-EOM-CCSD 6-31G bond ionization energies for core and HOMO ionizations are shown for the chromophore in Fig. 3-10. The x-axis represents occupied orbitals of waters which are an increasing distance from the dominant water shown in the center of the corresponding circles in Fig. 3-10. The dominant localized core and lone-pair ionization energies corresponding to orbital one and three, respectively are shown in the key for sake of clarity.

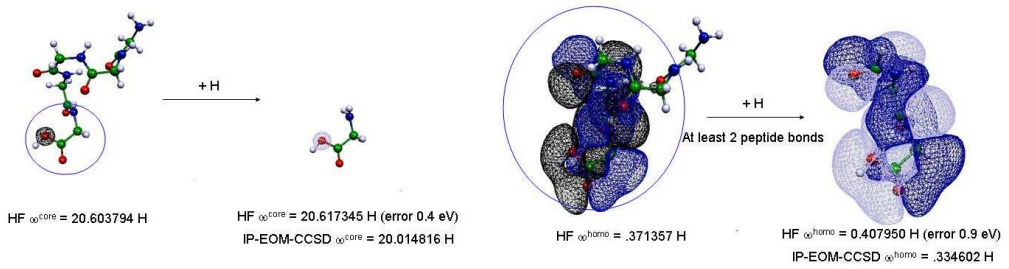

Figure 3-12: Ionization chromophores in gly_5 α-helix. Core (left) and valence (right) chromophores are determined using 6-31G HF bond ionization energies. The corresponding canonical orbitals shown in circles. HF and IP-EOM-CCSD calculations on the gly_1 and gly_2 chromophores are done in vacuum as opposed to an ESP due to the other residues. The canonical oxygen core and HOMO orbitals which dominate the ionizations are shown in mesh.

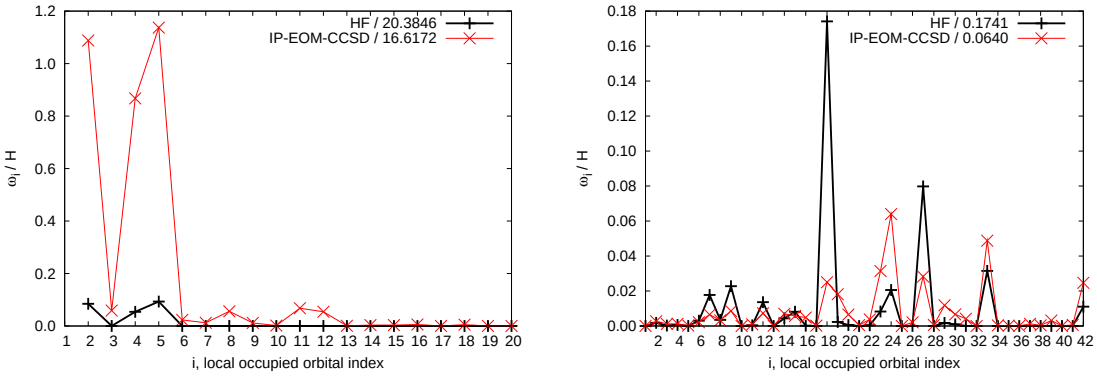

Figure 3-13: Bond ionization energies for the gly_1 and gly_2 chromophores. HF and IP-EOM-CCSD 6-31G bond ionization energies for the core and HOMO ionizations are shown in Fig. 3-12. The x-axis represents occupied orbitals which are along the peptide moving from C-terminus to the N-terminus. The dominant localized oxygen core ionization energies corresponding to orbital one are shown in the key on the left for sake of clarity. The dominant nitrogen lone-pair in the case of HF and σ_{CO} in the case of IP-EOM-CCSD ionization energies are shown in both the figure and key on the right.

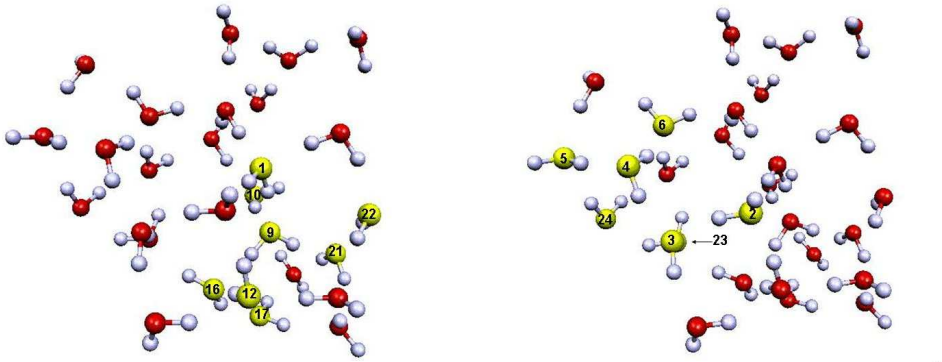

Figure 3-14: Ball and stick models of $(H_2O)_{28}$. One cluster with a region containing eight bulk water molecules is on the left while one with a region containing seven surface water molecules is on the right. Both regions are shown in yellow. The eight water molecules were chosen to be those within a distance of 4.0 $\mathring{A}$ from the center molecule given by oxygen 9 while the seven water molecules were chosen to be those within a distance of 4.5 $\mathring{A}$ from the center molecule given by oxygen 4.

Table 3-5: IP-EOM-CCSD valence ionization energies of $(H_2O)_{28}$. The chromophoric regions shown in Fig. 3-14 were used with a 6-31G* basis. There is one chromophoric region representative of the bulk and one for the surface both of which were determined from bond ionization energies at the HF level. The first two columns show the ionization energies in eV for the corresponding center water molecule in vacuum as well as in the ESP of the other twenty-seven waters. The last two columns show the ionization energies of the entire chromophoric region, eight water molecules for the bulk region and seven for the surface region, in vacuum as well as in the ESP of the other twenty and twenty-one water molecules, respectively. Also shown for all states are the leading ionized state wavefunction amplitudes. For each of the bulk and surface calculations a lower energy and higher energy state are shown, note the absence of a low energy state for the bulk $(H_2O)_8$ region.

bulk			
$(H_2O)_1$	$(H_2O)_1(H_2O)_{27}$	$(H_2O)_8$	$(H_2O)_8(H_2O)_{20}$
11.83 0.9733 $r_{9\,n_O}$	13.03 0.8708 $r_{9\,n_O}$		
18.58 0.6944 $r_{9\,\sigma_{OH}}$	19.08 0.7003 $r_{9\,\sigma_{OH}}$	19.77 0.4737 $r_{22\,\sigma_{OH}}$ 0.4238 $r_{16\,\sigma_{OH}}$	20.57 0.4475 $r_{22\,\sigma_{OH}}$ 0.3412 $r_{10\,n_O}$ 0.2505 $r_{10\,\sigma_{OH}}$
surface			
$(H_2O)_1$	$(H_2O)_1(H_2O)_{27}$	$(H_2O)_7$	$(H_2O)_7(H_2O)_{21}$
11.83 0.9733 $r_{4\,n_O}$	11.27 0.6972 $r_{4\,n_O}$	18.46 0.5436 $r_{23\,\sigma_{OH}}$ 0.4308 $r_{5\,\sigma_{OH}}$	18.84 0.5175 $r_{23\,\sigma_{OH}}$ 0.3947 $r_{24\,\sigma_{OH}}$
18.58 0.6944 $r_{4\,\sigma_{OH}}$	17.45 0.6906 $r_{4\,\sigma_{OH}}$	37.63 0.6191 $r_{5\,\sigma_{OH}}$ 0.2664 $r_{5\,n_O}$ 0.1548 $r_{23\,\sigma_{OH}}$	36.40 0.6029 $r_{5\,\sigma_{OH}}$ 0.2099 $r_{23\,\sigma_{OH}}$

CHAPTER 4
LOCAL APPROXIMATIONS FOR MOLECULAR PROPERTIES

When local correlation methods fail, the concern is the proper inclusion of non-locality in the calculations, preferably via intermolecular force theory. As the intent is to build all of the electronic structure of a large molecule from its largely transferable functional groups, besides simple extensive properties like energies and densities, it is necessary to address higher-order properties. Primary among these is the dynamic polarizability that leads to excited states and their associated density matrices [100]. The objective is to begin to assess the degree to which such properties can be constructed from localized, transferable regions, and to actually be able to observe when the locality approximation breaks down. Conceivably then, small interaction matrices involving the component groups could be built to allow for essential further interactions.

A first approximation would include London dispersion forces. The extent to which London dispersion coefficients are transferable has been a major theme of chemistry for a number of years, for example, consider the combination formulas due to Slater and Kirkwood [101] which can be used to construct heteronuclear coefficients from homonuclear coefficients. However, often times these atomic contributions are not very transferable. Some of these models have been shown to be even qualitatively incorrect [6]. By virtue of the n-center NLMOs some internal cancellations are incorporated *a priori* by defining functional groups in terms of multi-centered bonds.

The polarizability and higher-order response terms provide information about the optical properties of molecules via the response of a molecule to an external electric field. The higher-order terms provide the non-linear optical (NLO) properties [102] which are important in the design of optical materials [103, 104], while the poles of the frequency-dependent polarizability provide the ultra-violet/visible excitation spectrum of a molecule [105, 101].

4.1 Effective Hamiltonian Dynamic Polarizabilities

It is well known that these properties are given as external electric field derivatives of the energy or the quasi-energy in the time-dependent theories. The polarizability can be calculated using analytical derivative methods or perturbation theory [104, 106, 107, 108]. The latter method based upon the EOM-CC theory is used herein because of its accuracy in calculating excited states [75], dynamic polarizabilities [107], dispersion coefficients [109], and NMR coupling constants [110].

The CC method [1, 2] is well known to be the most accurate tool for determining the electronic structure of small to medium-sized molecules. Determining the correlated polarizabilities of molecules follows similarly from time-dependent HF (TDHF) [106, 104] except that the focus is on the correlated CC effective Hamiltonian, $\bar{H} = e^{-T}He^{T} = (He^{T})_c$, as opposed to the bare Hamiltonian. Here T is the CC excitation operator and c means connectedness as in the context of CC methods. The effective Hamiltonian approach to CC dynamic polarizabilities has been discussed by Stanton and Bartlett [107] and generalized by Rozyczko, et al. [108] and Sekino and Bartlett [111] to a fully extensive linear approximation used here. There has been great progress in the calculation of correlated optical properties like polarizabilities and dispersion coefficients, some of which have emphasized larger systems [107, 108, 109, 112, 113, 114, 33, 115, 116]. The effective Hamiltonian approach has been shown to reproduce FCI values for atomic polarizabilities and dispersion coefficients [107].

The induced electric dipole of a molecule is proportional to the electric field with proportionality constant given by the frequency-dependent polarizability which at time zero is given by,

$$\mu_q = \mu_q^0 + \alpha_{qq'}(\omega; -\omega)F_{q'} + \cdots , \tag{4–1}$$

with $\mu_q^0 = \langle 0|(qe^{T})_c|0\rangle$, where $|0\rangle$ is the single-determinant reference function, and with F_q being the field amplitude from the time-dependent perturbation $F_q cos(\omega t)$. This equation comes from the following Taylor series of the energy expanded about zero electric field

strength taken at time zero,

$$E = E^0 + \mu_q^0 F_q + \alpha_{qq'}(\omega; -\omega) F_q F_{q'} + \cdots , \tag{4–2}$$

with $E^0 = \langle 0|(He^T)_c|0\rangle$ and where the q-component of the electric dipole, μ_q^0, is given as the first derivative of the energy with respect to the q-component of the field. The (un)primed letter q designates Cartesian coordinates. Mixed second partial field derivatives of the energy give the polarizability tensor, $\alpha_{qq'}$, which is also given as the first dipole derivative as in Eqn. 4–1. Higher-order derivatives of the energy give the NLO properties of molecules.

4.1.1 Size-Extensivity

A method to determine dynamic electric dipole polarizabilities using the CC effective Hamiltonian has been previously described [107] and will only briefly be summarized. This approach is equivalent to a sum-over-states (SOS) approach [105] for the frequency-dependent polarizability as given by,

$$\alpha_{q\,q'}(\omega; -\omega) = \sum_k \left(\frac{\langle \Psi_0|q|\Psi_k\rangle\langle \Psi_k|q'|\Psi_0\rangle}{\omega^k + \omega} + \frac{\langle \Psi_0|q|\Psi_k\rangle\langle \Psi_k|q'|\Psi_0\rangle}{\omega^k - \omega} \right), \tag{4–3}$$

in which ω is the external field while ω^k and Ψ_k are the excited state energy gaps and wavefunctions, respectively. In practice the effective Hamiltonian approach avoids a diagonal representation of the effective Hamiltonian which would be necessary in the above SOS expression.

Using perturbation theory with a CC reference function [107] gives the following general form for the polarizability tensor,

$$\begin{aligned} \alpha_{q\,q'}(\omega; -\omega) &= \sum_{l=0}^{1} \langle 0|(1+\Lambda)\left[\bar{q} - \langle \bar{q}\rangle\right]|\mathbf{g}\rangle\langle \mathbf{g}|\left[\bar{H} - E + (-1)^l \omega\right]|\mathbf{g}\rangle^{-1} \times \\ &\quad \langle \mathbf{g}|\left[\bar{q}' - \langle \bar{q}'\rangle\right]|0\rangle \qquad (4\text{–}4) \\ &= \sum_{l=0}^{1} \langle 0|(1+\Lambda)\left[\bar{q}, X_{q'}^l\right]|0\rangle, \qquad (4\text{–}5) \end{aligned}$$

where X_q^l are the perturbed amplitudes. The total zero-field ground state CC energy is given by E and $|0\rangle$ is the reference determinant while $|\mathbf{g}\rangle$ are excited determinants which are complementary to the reference determinant. In this general case the zero-field Hamiltonian, $\bar{H}$, is non-Hermitian having left, $\langle\tilde{\Psi}_k| = \langle 0|\mathcal{L}_k e^{-T}$, and right, $|\Psi_k\rangle = e^T\mathcal{R}_k|0\rangle$, eigenvectors where $\mathcal{L}_0 = (1+\Lambda)$ the de-excitation operator from the theory of CC gradients and density matrices [2, 75]. The SOS Eqn. 4–3 is recovered if the Hamiltonian and complementary functions are represented in terms of excited state wavefunctions where the Hamiltonian is diagonal, as opposed to its representation in terms of determinants. Note that upon expansion the $\langle\bar{q}'\rangle\langle\mathbf{g}|0\rangle$ term vanishes. In the linear approximation the $[\bar{q} - \langle\bar{q}\rangle]$ term with $\langle\bar{q}\rangle = \langle 0|(qe^T)_c|0\rangle$ is a dipole operator which is represented with respect to the ground state CC wavefunction and is necessary because there is no requirement that Λ and $\bar{q}$ be connected. However, by imposing the condition that $[\bar{q} - \langle\bar{q}\rangle]\, X_{q'}^l = \left[\bar{q}, X_{q'}^l\right]$, the previously proposed linear approximation that is fully extensive [108, 111] is obtained.

4.1.2 Linear Response

In the effective Hamiltonian approach the linear response equations,

$$\langle\mathbf{g}|\left[\bar{H} - E + (-1)^l\omega\right]|\mathbf{g}\rangle\langle\mathbf{g}|X_q^l|0\rangle = \langle\mathbf{g}|\bar{q}|0\rangle, \tag{4–6}$$

are solved for the perturbed wavefunction amplitudes, with $l = 0$ or 1, using iterative methods. Once the amplitudes are determined they are contracted with the perturbation to give the polarizability,

$$\begin{aligned}\alpha_{q\,q'} &= \sum_{ia}\langle 0|(1+\Lambda)\left[(qe^T)_c, \left[X_{q'}^0 + X_{q'}^1\right]\right]|_i^a\rangle \\ &+ \sum_{ijab}\langle 0|(1+\Lambda)\left[(qe^T)_c, \left[X_{q'}^0 + X_{q'}^1\right]\right]|_{ij}^{ab}\rangle \end{aligned}\tag{4–7}$$

which has been written in a more complete form for clarification. The usual declaration of $i, j, \cdots \in \{occ\}$, $a, b, \cdots \in \{vir\}$, and $p, q, \cdots \in \{occ\} \cup \{vir\}$ is used. Oftentimes it is

useful to work with the isotropic polarizability, $\langle\alpha\rangle$, given by,

$$\langle\alpha\rangle = \frac{1}{3}\left(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}\right), \tag{4–8}$$

to ensure averaging over molecular orientation.

4.2 Localized Effective Hamiltonian Dynamic Polarizabilities

Calculations on larger systems are often prohibitive with high-level methods due to an exponential scaling wall. To correct these limitations it is necessary to have *ab initio* methods which scale favorably with system size but which hopefully are still systematically improvable. Considering the success of local correlation methods for large molecules it is interesting to consider whether the same advantages hold for more complicated properties. Generally these methods can reproduce conventional CC results, emphasizing the correlated energies, and have been applied to optical properties with some success [114, 115]. These methods can be supplemented with an ESP as has been done for the optical properties of large molecules using the FMO method [117]. Given that the optical properties are sensitive to the quality of the unperturbed wavefunction it is important that the local approximation not destroy the reference function. The local CC methods should allow CC theory to be applied to calculate the optical properties of larger molecules in a systematically improvable fashion. The major objective is to assess the potential transferability of response properties of chromophoric regions in a molecule. The same tools can be used to obtain London dispersion interactions by evaluating such polarizabilities at fictitious imaginary frequencies and using the Casimir-Polder formula [118].

4.2.1 Localized De-Excitation Operators

The NLSCCSD approach to calculating the polarizabilities of large molecules follows from the ground state NLSCCSD wavefunction by considering, in analogy with a local excitation operator, for example T_2^i as in the case of double excitations, a local de-excitation operator, Λ_2^i. This linear de-excitation operator is the left-hand eigenvector of

$\bar{H}$ and is a de-excitation specific to orbital i given by,

$$\Lambda_2 = \sum_i \Lambda_2^i = \sum_i \left(\frac{1}{4} \sum_{jab} \lambda_{ab}^{ij} i^\dagger j^\dagger ba \right). \tag{4–9}$$

Because the left-hand eigenstate of the effective Hamiltonian is parameterized by a linear CI-like operator, much like the excitation operator from EOM-CC [75], some higher-order properties are not always size-extensive in the usual CC terminology. To guarantee size-extensivity for such properties the linear approximation is used herein. Compared to CI, the deviations from size-extensivity in the Λ operator by virtue of its disconnectedness are small although formally still present. While C from CI, see Fig. 2-3, is not transferable, Λ from CC is transferable similarly to T the CC excitation operator. The excitation operators C_2 and T_2 from CISD and CCSD, respectively, are related by,

$$C_2 = T_2 + \frac{1}{2} T_1^2, \tag{4–10}$$

while for the de-excitation operators Λ_2 and Σ_2 from a linear and exponential left-hand ansatz, respectively,

$$\Lambda_2 = \Sigma_2 + \frac{1}{2} \Sigma_1^2, \tag{4–11}$$

is obtained. In the case of CISD the operator C_2 is size-inextensive and thus does not scale properly with system size. One manifestation of this scaling is understood by considering the non-cumulant piece of double excitations, T_1^2, which can be nonvanishing in cases where $t_i^a \in A$ and $t_j^b \in B$ with A and B two systems which are distant from one another. Although it is expected that similar behavior would be observed for Λ_2 by virtue of its linearity, as in the context of extensivity in EOM-CC [75], the size of Σ_1 is quite small requiring that Λ_2 be approximately extensive. For example, the Λ_2-equation contains two disconnected terms, one of which vanishes because the NLMO basis is only non-canonical HF as opposed to being non-HF. For HF reference functions, where T_1 and

Λ_1 are expected to be small, contributions from the other disconnected term,

$$\lambda_{ab}^{ij} \longleftarrow \sum_{kc} w_{ikac} t_k^c \lambda_b^j, \tag{4–12}$$

are present, albeit small, thus leading to the approximate extensivity of Λ_2. Note that for non-HF reference functions it is possible that Λ_2 may be less extensive. A qualitative explanation of the extensivity of Λ_2 is provided by recognizing that the CC effective Hamiltonian is almost Hermitian and thus CC theory, in the present context, is almost variational making the role of Λ_2 in restoring an expectation value only minor. This is in contrast to the major role of C_2 which must build in all two-electron correlations from a single-determinant. The de-excitation operator, Λ, is thus transferable in analogy with the excitation operator, T. Higher excitation levels of Λ, as in $\Lambda_3 = \Sigma_3 + \Sigma_2\Sigma_1 + \Sigma_1^3$, may be somewhat less transferable due to higher-order product contributions. The assumption that the operator Λ^i in the NLMO basis only needs to be locally correlated is built in by summing over a small region, QM2, composed of orbitals that are spatially close to i as in Eqn. 2–16 giving,

$$\Lambda_2^i \approx \frac{1}{4} \sum_{jab \in QM2} \lambda_{ab}^{ij} i^\dagger j^\dagger ba \quad i \in QM1 \quad j, a, b \in QM2. \tag{4–13}$$

Note that the same QM regions for both excitation and de-excitation operators are used. However, this is not necessary and it may be beneficial to relax this condition in certain unforeseen cases. The CCSD Λ-equations can thus be solved in linear or even sub-linear time.

4.2.2 Effective Bond Dynamic Polarizabilities

Expanding the right-hand side of Eqn. 4–7 shows that for both singles and doubles contributions to the polarizability there will be no terms with more than a given number, three in this case, of occupied summing indices, i, j, and k. This is in contrast to the CC correlation energy for which this number is two. Breaking down operators in terms of local contributions is useful in determining which functional groups have important

excitations and de-excitations into one another. This is similar in spirit to our recent work on triples contributions to the energy [29] where it is shown that triples contributions are most important for delocalized moieties suggesting an interesting active space method. This decomposition is seen by using the local T^i and Λ^i operators in Eqn. 4–7, along with the appropriate form for a one-particle operator, for example the dipole $q = \sum_p q_p = \sum_{pq} q_{pq}\{p^\dagger q\}$. As an example, consider a term from the singles equation,

$$\langle 0|\Lambda_2^i q^j T_2^k|_i^a\rangle \longleftarrow \sum_{bc} \lambda_{ac}^{ik} q_b^j t_{kj}^{cb}, \tag{4–14}$$

where the virtual orbital dependence has been summed out. When contracted with the perturbation amplitudes,

$$\alpha_{qq'_{ijk}} \longleftarrow \sum_a \langle 0|\Lambda_2^i q^j T_2^k|_i^a\rangle \left[X_{q'\,i}^{0\,a} + X_{q'\,i}^{1\,a}\right], \tag{4–15}$$

we get the following decomposition for the polarizability in terms of occupied triples,

$$\alpha_{qq'} \longleftarrow \sum_{ijk} \alpha_{qq'_{ijk}}. \tag{4–16}$$

Similarly consider the term from the doubles equation,

$$\langle 0|\Lambda_2^i q^j T_1^k|_{ij}^{ab}\rangle \longleftarrow \sum_c \lambda_{ab}^{ik} q_c^j t_k^c, \tag{4–17}$$

which when contracted again with the perturbation amplitudes,

$$\alpha_{qq'_{ijk}} \longleftarrow \sum_{ab} \langle 0|\Lambda_2^i q^j T_1^k|_{ij}^{ab}\rangle \left[X_{q'\,ij}^{0\,ab} + X_{q'\,ij}^{1\,ab}\right], \tag{4–18}$$

also results in a sum over occupied triples.

This breakdown of operators into local orbital or functional group contributions allows the representation of the polarizability in terms of single $\langle 0|\Lambda_2^i q^i T_2^i|_i^a\rangle$, double $\langle 0|\Lambda_2^i q^i T_2^j|_i^a\rangle$, and triple $\langle 0|\Lambda_2^i q^j T_2^k|_i^a\rangle$ contributions for the single excitation equations, ignoring permutations thereof. Taking the occupied singles piece and representing the polarizability in terms of QM1 regions alone is tempting, however, results indicate that

it is necessary to account for relaxation effects of orbital excitations and de-excitations into other groups. These would not be obtained with such a diagonal structure of the doubles excitation operators. Relaxing this constraint recovers occupied doubles terms but it is not until the dipole operator is allowed to be independent of T_2 or Λ_2 as it is for the occupied triples piece, that the conventional polarizability is recovered. Assuming that Λ scales properly with system size many of the $\alpha_{qq'_{ijk}}$ contributions are negligible thereby allowing useful linear-scaling algorithms. Rather than transferring these n_{occ}^3 dependent contributions, effective bond contributions are defined which act to incorporate longer-ranged contributions in a given larger QM region, QM2, into an occupied singles contribution, $\langle 0|\Lambda_2^i{}' q^i{}' T_2^i{}'|_i^a\rangle$, in the QM1 region. Note that for every term it is the de-excitation operator Λ_2 or q which carries the i dependence.

4.2.3 Effective Bond Response Matrices via Löwdin Partioning

The perturbation amplitudes, X_q^l, are determined from the linear response equation given by Eqn. 4–6. This equation can be rewritten in simplified notation as,

$$\langle \mathbf{g}|A|\mathbf{g}\rangle\langle \mathbf{g}|X|0\rangle = \langle \mathbf{g}|\bar{q}|0\rangle, \tag{4–19}$$

where $A = \bar{H} - E + (-1)^l\omega$ and where the sub- and super-scripts on X have been dropped. It is useful to consider the following Löwdin partitioning of the response into P and Q spaces for which P is specific to the occupied NLMO, i, given by $|_i^a\rangle \in P$ and Q is given by $|_j^b\rangle \in Q$. The singles block is chosen for simplicity. It is well known that the set of linear equations can be represented in the P space by redefining the linear response and perturbation matrices as follows,

$$A'_{PP}X_P = \bar{q}'_P, \tag{4–20}$$

where $A'_{PP} = A_{PP} - A_{PQ}A_{QQ}^{-1}A_{QP}$ and $\bar{q}'_P = \bar{q}_P - A_{PQ}A_{QQ}^{-1}\bar{q}_Q$. The single orbital response matrix, $\langle_i^a|A|_i^b\rangle$, alone, provides an oversimplification of the response. Although it contains the direct response of i to the field, it lacks the response of i to other orbitals which in turn are responding to the field. This piece is corrected by $A_{PQ}A_{QQ}^{-1}A_{QP}$ which

again builds in longer-ranged interactions from the QM2 region. By using the Löwdin partitioning it is possible to define an effective response matrix, $\langle {}^{a}_{i}|A'|{}^{b}_{i}\rangle$, in analogy with the previously mentioned effective operators. The local perturbed amplitude solutions can be written as $X^{l}_{q\,2} = \sum_i X^{l\,i}_{q\,2}$ and used in Eqn. 4–7 to determine effective bond polarizabilities,

$$\alpha_{q\,q'\,i} \leftarrow \Lambda^i q^i T^i X^i_{q'}. \tag{4–21}$$

4.3 Dispersion Interactions

The PES of a system of molecules is dominated at small internuclear distances by quantum mechanical effects such as coulomb, exchange, and correlation which cause large distortions in the electron densities of the molecules. In the limit of large internuclear distances, that is the distance is large relative to the sizes of the constituents and therefore the overlap is negligible, the intermolecular potential is dominated by attractive quantum mechanical London dispersion or instantaneous dipole interactions [101]. The total long-range interaction energy between two chemical species, atoms and/or molecules A and B, can be represented as,

$$E^{AB} = E^{AB}_{es} + E^{AB}_{ind} + E^{AB}_{disp}, \tag{4–22}$$

with E^{AB}_{es} the usual electrostatic energy, E^{AB}_{ind} the induction energy, and E^{AB}_{disp} the dispersion energy. Consider the case for which the two chemical species are charge neutral and have no permanent dipole moment, requiring that E^{AB}_{es} and E^{AB}_{ind} vanish. The E^{AB}_{disp} can be represented in terms of dipole-dipole, dipole-quadrupole, etc. terms as in the following multipole expansion,

$$E^{AB}_{disp} = -\frac{C^{AB}_6}{R^6} - \frac{C^{AB}_7}{R^7} - \frac{C^{AB}_8}{R^8} + \cdots, \tag{4–23}$$

where the convention of assuming positive dispersion coefficients is used. The lowest-order term, the dipole-dipole term,

$$E^{AB}_{dip-dip} = \frac{\vec{\mu}_A \cdot \vec{\mu}_B - 3\left(\vec{\mu}_A \cdot \vec{R}\right)\left(\vec{\mu}_B \cdot \vec{R}\right)}{R^3}, \tag{4–24}$$

will be the focus herein [101]. Note that R is the AB distance. Using perturbation theory gives the leading order term in the energy, $E^{AB}_{dip-dip}$, as $-\frac{C_6^{AB}}{R^6}$ with,

$$C_6^{AB} = 6 \sum_k \sum_{k'} \frac{|\mu^A_{0k}|^2 |\mu^B_{0k'}|^2}{\omega^A_k + \omega^B_{k'}}, \tag{4–25}$$

where k, k' label the excited states, μ_{0k} is the transition dipole moment, and ω_k are the excitation energies.

4.3.1 Casimir-Polder Equation

Casimir and Polder [118] transformed this expression for the dispersion coefficient into an integral over imaginary frequencies. In this way Eqn. 4–25 can be rewritten in terms of fictitious polarizabilities evaluated at imaginary frequencies,

$$C_6^{AB} = \frac{3}{\pi} \int_0^\infty d\omega \left(\sum_k \frac{|\mu^A_{0k}|^2 \omega_{0k}}{\omega^2_{0k} + \omega^2} \right) \left(\sum_{k'} \frac{|\mu^B_{0k'}|^2 \omega_{0k'}}{\omega^2_{0k'} + \omega^2} \right), \tag{4–26}$$

or,

$$C_6^{AB} = \frac{3}{\pi} \int_0^\infty d\omega \alpha_A(i\omega) \alpha_B(i\omega). \tag{4–27}$$

Polarizabilities at imaginary frequencies can be understood to be the response of a molecule to an exponentially increasing electric field [101]. Due to the fact that these polarizabilities experience no poles and monotonically decrease in magnitude with increasing field strength, they can be found quite easily and suffer little from convergence difficulties. This is to be contrasted with determining the shape of the poles around an excitation in real frequency calculations. The effective Hamiltonian approach for polarizabilities at imaginary frequencies has been developed in detail elsewhere [109].

4.3.2 Corrective Dispersion Potential

One question to address is to what extent can the dispersion coefficient, C_6^{AB}, be decomposed into orbital dispersion coefficient contributions, ${C_6^{AB}}_{ij}$, corresponding to all possible intermolecular bond pairs,

$$C_6^{AB} = \sum_i^{n_{occ\,A}} \sum_j^{n_{occ\,B}} {C_6^{AB}}_{ij}. \tag{4–28}$$

As such correlated dispersion coefficients can be determined for long-ranged interactions in large molecules by taking advantage of the transferability of the QM regions in a NLSCC calculation. The polarizability at imaginary frequencies is well behaved as previously mentioned, thereby requiring only a small number of grid-points.

4.3.3 Anisotropic Contributions

The Casimir-Polder equation, as written above, uses isotropic polarizabilities and therefore only provides isotropic dispersion coefficients. To assess how the dispersion interaction between otherwise very distant functional groups changes as a function of orientation the anisotropic polarizabilities need to be incorporated into the Casimir-Polder equation [6, 12]. Furthermore, the same techniques used here can be applied to higher C_n coefficients in the event that such terms are necessary.

4.4 Implementation

All calculations are performed using a modified version of the ACES II [16], quantum chemistry software package. These modifications include an interface to Flocke's highly efficient and general NLMO package [15]. Results are obtained using a 375 MHz power3 processor using a maximum of 4 of the available 8 GB of shared memory over four processors, as well as an 8 dual-core ia64 processor SGI Altix machine with 256 GB of shared memory using multithreaded libraries. Such calculations make use of some recent advances in the ACES II [16] code to take advantage of large shared memory machines. These advances include updating some old memory limitations to allow for large effective Hamiltonians and in-core contractions in the context of many-body methods. Using in-core contractions are quite useful in the post-CC part of the calculation in evaluating the CC density matrix and subsequent use in calculating the polarizability and dispersion coefficients. All correlated polarizability and dispersion coefficient calculations are performed at the CCSD level using the effective Hamiltonian approach with a non-canonical RHF NLMO reference function. Dispersion coefficients are determined using an ω_0 value of 0.4 [109]. The atomic units of polarizability are volume in Bohr units

cubed, a_0^3. The results are expected to be accurate for molecules whose excited states are dominated by single excitations for which EOM-CCSD [75] is known to be quite good.

Geometries for poly-glycine are taken from reference [36]. Tryptophan and di-glycine molecules are created using Molden [30]. Geometries of test systems are available upon request.

Poly-glycine is meant to represent the simplest peptide that still contains a variety of functional groups, for example, it has a 3-center peptide-bond. Tryptophan has a 9-center indole-bond. A cc-pVDZ basis is used unless specified otherwise, however, for more accurate calculations more advanced basis sets with high angular momentum and diffuse functions should be used. Localized core occupied orbitals have been dropped unless specified otherwise along with localized core virtual orbitals. For example Rydberg functions of angular momentum s have been removed. Other Rydberg functions have also been dropped depending on the type of calculation.

4.5 Applications

4.5.1 Motivation

As mentioned above, a decomposition of the CCSD wavefunction into functional group contributions is possible by virtue of its size-extensivity, unlike CISD. Diagonal T_2 from the ground state CCSD wavefunction, Λ_2 from the ground state CCSD density matrix, and C_2 from the ground state CISD wavefunction amplitudes for a σ_{CH}-bond belonging to a methyl group as a function of alkane size were previously shown in Fig. 2-3. The figure shows that the CISD amplitudes, unlike the CCSD amplitudes, are not transferable. Additionally, it is seen that the left-hand eigenstate of $\bar{H}$, Λ_2, is transferable despite the fact that it is parameterized by a linear, size-inextensive, as in the context of EOM-CCSD [75], CI-like operator. This operator provides important relaxation terms to the CC density matrix and thus correlated properties. The CISD amplitudes decrease in magnitude with increasing system size which is consistent with the fact that in the infinite limit the correlation energy per particle becomes zero.

Fig. 4-1 shows the transferability of the response matrix from Eqn. 4–6 for a σ_{CH}-bond/anti-bond pair among a series of substituted alkanes. The response matrix plotted excludes the constant $E + (-1)^l\omega$ scalar. The plot demonstrates that the localized methyl group is screened from the perturbation. The left plot contains the single excitation component while the double excitation component is on the right. When there is no screening present the response matrix components are quite different, however, these differences become negligible as the number of screening regions increases. The rate at which this happens obviously depends on the functional group, for example, the methyl substituent is found to change only slightly with screening while the fluorine, not surprisingly, perturbs the density significantly. Note that not only does each functional group independently approach a constant but that the different functional groups approach the same constant. The general conclusions are the same for both the singles and doubles components. Given that the effective Hamiltonian from CC theory contains information about the excited states and can be represented for a large target system in terms of many smaller Hamiltonians, it is possible to obtain the excitation spectrum of a large molecule via locally transferable QM regions.

A plot showing the transferability of the methyl group contribution to the average static polarizability among a series of substituted alkanes is shown in Fig. 4-2. These methyl group contributions are highly transferable in contrast to the effective Hamiltonian elements which are found to require more regions of screening to reach the same level of transferability. This is because the polarizability takes advantage of some internal cancellations among the excitation, de-excitation, etc. operators. When there are no screening regions the methyl group polarizability is quite dependent on the substituted group, however, with as little as one screening region it becomes approximately independent of the substituent.

The methyl group contribution to the average polarizability of ethanol and ethylamine versus frequency is shown in Fig. 4-3. For smaller frequencies this contribution is

independent of substituent, while as the frequency increases toward an excitation energy the transferability of the methyl group polarizability breaks down. The example here is for the first excitation energy of ethylamine which is given by excitation energy (EE-) EOM-CCSD at the dotted vertical line. The breakdown in transferability is a manifestation of the delocalized nature of the excited state wavefunctions which occur at different frequencies. Each group polarizability comprising the molecular polarizability will experience a pole in the response at the same excitation energy because each contribution contains a piece of the corresponding denominator.

Despite these technical complications among the polarizability contributions in that they all experience singularities at the excitation energy of the system, Figs. 4-4 and 4-5 show that the rate at which this infinity is approached is quite different among the functional groups thereby providing useful information about the excited states. Figs. 4-4 and 4-5 show a complete decomposition of the correlated polarizability in terms of functional group contributions as a function of frequency for three excited states of ethanol and ethylamine, respectively. These excited states are verified with EE-EOM-CCSD and are given by the corresponding dotted vertical lines. For ethanol the first two excited states are localized on the hydroxyl group, which has a strong response, with residual tail contributions from the methylene group. One of the σ_{CC}-bonds, the one neighboring the hydroxyl group, which are included in the methylene groups for convenience, not surprisingly dominates this tail methylene contribution. These polarizability contributions are opposite in sign, reflecting their different hyperpolarizabilities, and their sum gives the total polarizability given that the methyl group contributions are negligible. If the σ_{CC}-bond are moved from the methylene group to the hydroxyl group it would be seen that the new hydroxyl group dominates the polarizability while the contributions from the methylene and methyl groups are both negligible. This means that the excitation is dominant on that group and, to such an extent, is localized. The third excited state is more delocalized than the first two because other than the slight dominance of the methylene

group all the functional groups make comparable contributions. The third state is most likely a delocalized Rydberg state.

From Fig. 4-5 in analogy with ethanol the first and second excited states of ethylamine are strongly dominated by the unique functional group, in this case the amino group, with residual contributions from the methyl and methylene groups. The dominance of the methylene group for the third excited state of ethylamine is greater than it is for ethanol.

4.5.2 Tryptophan Dynamic Polarizabilities

Table 4-1 shows the decomposition of the average static polarizability into functional group and NLMO contributions for tryptophan. NLMOs are ordered per column as lp, σ, and π orbitals as well as a 9-center indole-bond given by $1\pi_9 - 5\pi_9$. For each type of NLMO within each column the polarizability contributions are ordered by increasing magnitude. The molecule tryptophan is shown on the left in Fig. 4-6 along with the atomic indices. The molecule is decomposed into functional groups corresponding to carboxyl, amino, indole, and the remaining aliphatic part. Comparing the group polarizabilities shows that the indole group has the largest polarizability followed in order by the aliphatic, carboxyl, and amino groups. It is well known that the polarizability of a molecule is proportional to its volume, therefore the following averages are considered, an orbital average in which this quantity is divided by the total number of orbitals, and an atom average in which it is divided by the total number of atoms. For the averages it is found that the indole group is largest followed by the aliphatic region. The ordering of the carboxyl versus amino groups depends on the averaging method used. The largest NLMO lp polarizability contribution comes from the carbonyl oxygen in the carboxyl group, while the largest σ-bond contributions come from the central σ_{CC}-bond of the aliphatic region. Somewhat surprisingly the σ and π contributions to the carbonyl group of the carboxyl region are the smallest NLMO contributions among the molecule and are even comparable to one another.

Fig. 4-7 shows the decomposition of the polarizability into functional groups used in Table 4-1 as a function of frequency. Excitation energies for the first two excited states as given by EE-EOM-CCSD are shown by the dotted vertical lines. It is easy to see that the first excited state is localized to the indole group due to its very large response in comparison to the other functional groups. The second excited state is localized to the carboxyl and aliphatic groups with constant contributions from the amino and indole groups. In this case because the carboxyl polarizability contribution is negative it indicates that the local dipole is decreasing in magnitude with increasing frequency. This is in contrast with the dipole of the aliphatic group which is increasing in magnitude, suggesting how the density rearranges in response to the electric field.

4.5.3 Poly-Glycine Dynamic Polarizabilities

Fig. 2-7 shows the QM regions defining the NLSCCSD effective Hamiltonian calculations on poly-glycine. The functional groups of interest are represented by the QM1 regions, the QM label is to specify that this region is treated quantum mechanically as opposed to by more classical interactions, like an ESP, etc. Such a potential could be incorporated into regions outside of QM2, but will not in the current formulation. There is one region for the methylene group, one for the peptide group, one for the N-terminus, and one for the C-terminus. This region is embedded in the QM2 region, which in turn could be embedded in an even larger QM region [29].

Table 4-2 has total and unit cell NLSCCSD dynamic polarizabilities for translationally periodic poly-glycine where the unit cell is chosen as one glycine residue. The four frequencies are chosen as zero, one, two, and three times that of the sodium D-line in analogy with Stanton's work [109]. Approximate dynamic polarizabilities calculated using NLSCC are compared to the conventional effective Hamiltonian approach for tetra-glycine where it is seen that for zero frequency over 99% of the correlated polarizability is recovered. As the frequency approaches an excitation energy of tetra-glycine the QM regions become less transferable giving slightly less agreement with conventional results. In the

infinite limit, the polarizability of a unit cell of glycine is $\approx 17.1\ a_0^3$ at zero frequency and $\approx 21.6\ a_0^3$ at three times the sodium D-line frequency ($3 \times 0.0773\ h$).

4.5.4 Alkane Dispersion Coefficients

Eqn. 4–28 shows that the total dispersion coefficient can be represented in terms of inter-bond contributions. Fig. 4-8 shows the transferability of three pair contributions to the dispersion coefficients for the interaction of two alkanes. Eight grid-points are used for the numerical integration. Dimers of propane through heptane are used and values for the methyl-methyl, methyl-methylene, and methylene-methylene dispersion coefficients are determined. It is seen that the methyl-methyl contribution to dispersion is approximately constant in this isotropic approximation from one alkane dimer to another while the methyl-methylene and methylene-methylene contributions require a few units of screening before becoming transferable. The definition of the methylene group used herein includes the adjacent σ_{CC}-bonds and so it is not surprising that those contributions are larger than the methyl group contributions which only contain σ_{CH}-bonds.

In Table 4-3 dispersion coefficients per methylene unit cell are shown for the interaction of two alkanes given on the right side of the table. The column on the left gives the QM2 region used in constructing the target results shown on the right. Note that the diagonal elements are the exact result with the basis set used. The value per unit cell shown are determined by taking the total dispersion coefficient from Eqn. 4–28 and dividing by the total number of carbons. The rate of convergence of the methylene unit cell dispersion coefficient is clearest for heptane which when represented in terms of a propane QM2 region is in error by $7.3\ a_0^6 h$. This error is reduced to $0.4\ a_0^6 h$ when QM2 is hexane. Also shown is the methylene unit cell dispersion coefficient determined in the infinite limit. The converged values for the methylene unit cell are comparable to those from Fig. 4-8.

4.5.5 Di-Glycine Dispersion Coefficients

The convergence with respect to the number of grid-points for the total and diagonal contributions to the dispersion coefficient for the interaction of two di-glycines is shown

in Fig. 4-9. The di-glycine molecule is shown on the right in Fig. 4-6. The coefficients are normalized to the values corresponding to the largest number of grid-points which are shown in the legend. Di-glycine is decomposed into carboxyl, peptide, amino, and two types of methylene groups. It can be seen that all contributions rapidly converge with respect to the number of grid-points. It is seen that the peptide-peptide contribution is largest, while the amino group has the smallest contribution, and that the two methylene groups, not surprisingly, are approximately equivalent.

A decomposition of the dispersion coefficient into diagonal NLMO contributions per functional group for the two interacting di-glycines is shown in Table 4-4. NLMOs are ordered per column as lp and σ orbitals as well as two 3-center carboxyl- and peptide-bonds given by $1\pi_3 - 2\pi_3$. For each type of NLMO within each column the orbital dispersion coefficient contributions are ordered by increasing magnitude. A sixteen point grid is used. The largest lp contribution is from the nitrogen of the amino group while the largest σ-bond contributions are from the σ_{CC}- and σ_{CH}-bonds. The 3-center peptide-bond has by far the largest contribution. The more appropriate way to describe long-range dispersion effects between structural regions is to evaluate anisotropic coefficients which will be the subject of future work. The current isotropic coefficients illustrate the computational approach to be used.

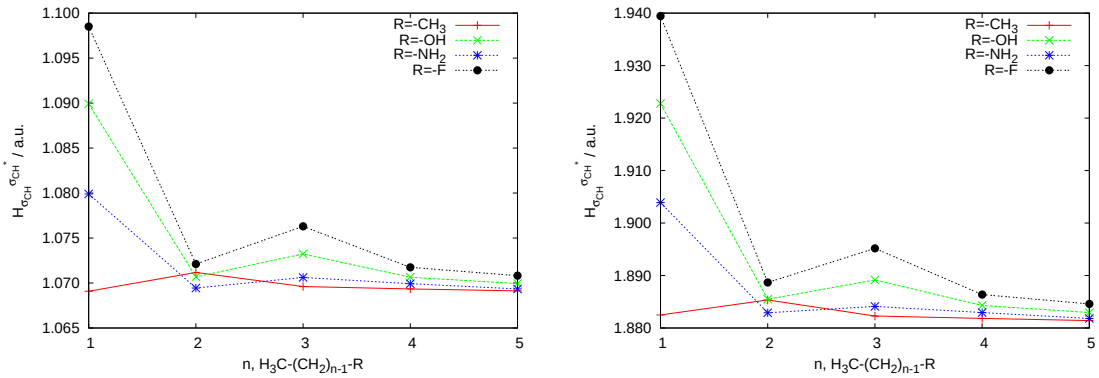

Figure 4-1: Transferability of the effective Hamiltonian. Figure showing the transferability of the response matrix from Eqn. 4–6 for a σ_{CH}-bond among a series of substituted alkanes. The σ_{CH}-bond belongs to a methyl group which is on the opposite side of the substituent and the x-axis represents increasing distance from the methyl group in terms of the number of methylene groups. The left plot contains the single excitation component, $\langle {}^{\sigma^*_{CH}}_{\sigma_{CH}} | \bar{H}_N | {}^{\sigma^*_{CH}}_{\sigma_{CH}} \rangle$, while the double excitation component is on the right. H in the figure means the CC effective Hamiltonian, $\bar{H}_N$, which is normal ordered with respect to the reference function $|0\rangle$.

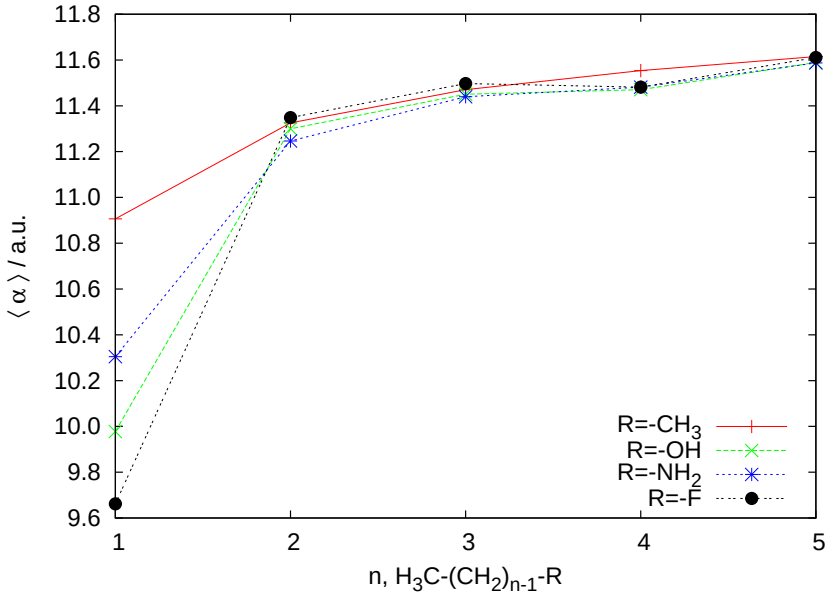

Figure 4-2: Transferability of CCSD bond polarizabilities. The figure is for the methyl group contribution to the average static polarizability among a series of substituted alkanes. The methyl group is on the opposite side of the substituent and the x-axis represents increasing distance from the methyl group in terms of the number of methylene groups.

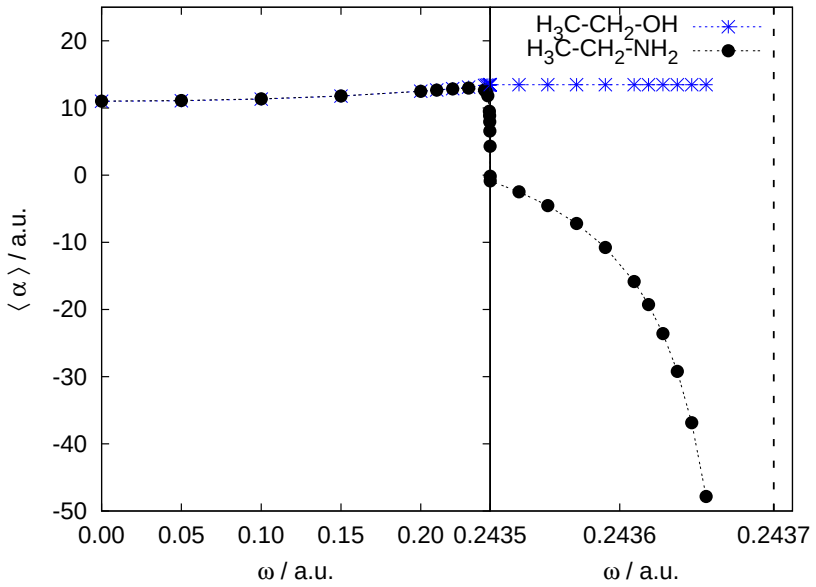

Figure 4-3: Frequency dependent bond polarizabilities and associated poles. The methyl group contribution to the average polarizability of ethanol and ethylamine is shown. The plot shows the gradual breakdown of the transferability of the methyl group polarizability as the frequency approaches the excitation energy. The example here is for the first excitation energy of ethylamine which is given by EE-EOM-CCSD at the dotted vertical line. Note the change in the domain on the left- and right-hand sides of the plot.

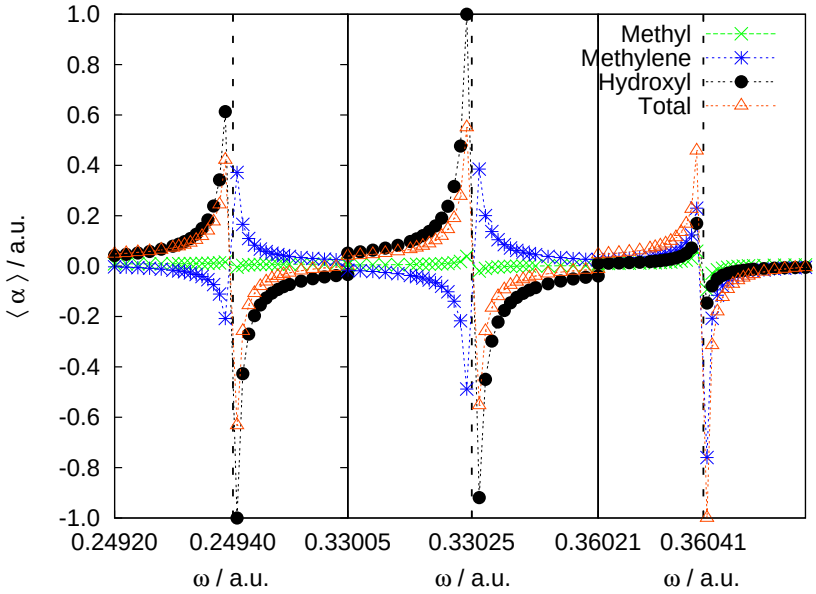

Figure 4-4: Frequency dependent bond polarizabilities for ethanol. The average polarizability is used. The frequencies are chosen symmetrically about three excited states determined using EE-EOM-CCSD and are shown as the dotted vertical line in each case.

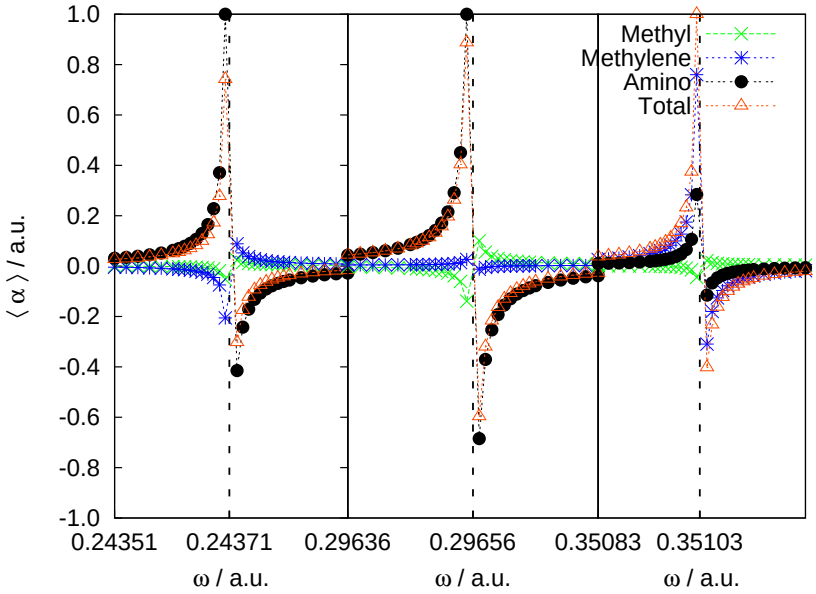

Figure 4-5: Frequency dependent bond polarizabilities for ethylamine. The average polarizability is used. The frequencies are chosen symmetrically about three excited states determined using EE-EOM-CCSD and are shown as the dotted vertical line in each case.

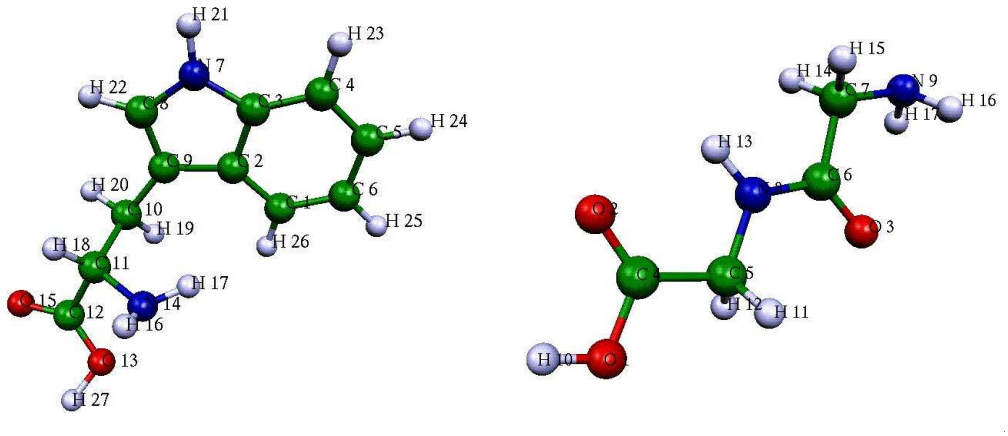

Figure 4-6: Ball and stick models of tryptophan and di-glycine. The molecules are shown on the left and right, respectfully along with atomic indices.

Table 4-1: Static bond polarizabilities for tryptophan. The average static polarizability is used. NLMOs are ordered per column as lp, σ, and π orbitals as well as a 9-center indole-bond given by $1\pi_9 - 5\pi_9$. For each type of NLMO within each column the polarizability contributions are ordered by increasing magnitude. Also shown are two types of polarizability densities.

carboxyl		*amino*		*aliphatic*		*indole*	
$O13$	0.6425	$N14$	0.7464	$C11N14$	1.4988	$C8C9$	0.9998
$O15$	0.6762	$N14H17$	1.2272	$C10H19$	1.7382	$C1C6$	1.1163
$O13$	1.2822	$N14H16$	1.2913	$C11H18$	1.7850	$C4C5$	1.1168
$O15$	2.5920			$C10H20$	1.8908	$C3N7$	1.1142
$C12O15$	0.4309			$C11C12$	2.0785	$C5C6$	1.1446
$C12O13$	0.7029			$C9C10$	2.2084	$C1C2$	1.1787
$O13H27$	1.3575			$C10C11$	2.7265	$C3C4$	1.1847
$C12O15$	0.4887					$N7C8$	1.1696
						$C2C3$	1.2690
						$N7H21$	1.4704
						$C2C9$	1.6608
						$C1H26$	1.6810
						$C4H23$	1.7444
						$C8H22$	1.7679
						$C6H25$	1.7876
						$C5H24$	1.8183
						$1\pi_9$	0.7087
						$2\pi_9$	1.2645
						$3\pi_9$	6.8956
						$5\pi_9$	10.9716
						$4\pi_9$	13.7593
Total	8.1732		3.2650		13.9266		55.8248
Orb Avg	1.0217		1.0883		1.9895		2.6583
Atm Avg	2.0433		1.0883		2.7853		3.7217

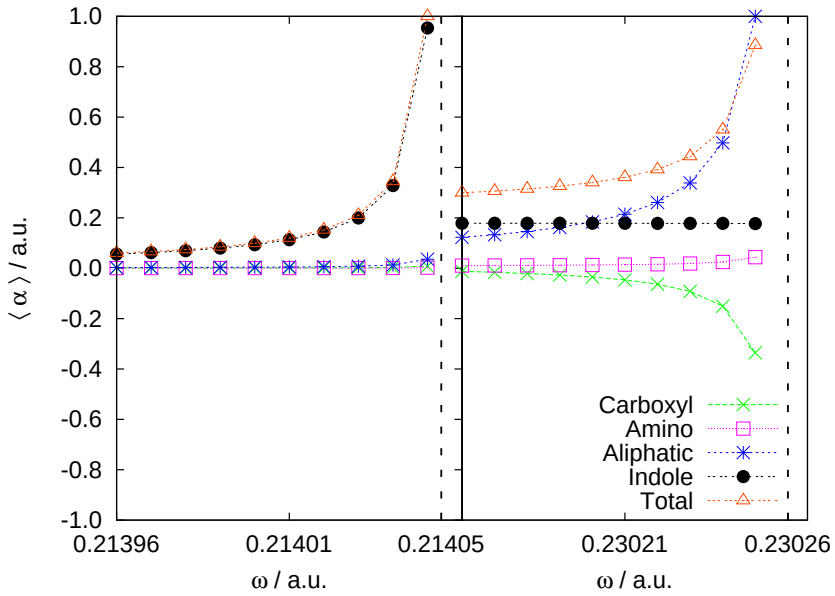

Figure 4-7: Frequency dependent bond polarizabilities for tryptophan. The average polarizability is used. Excitation energies for the first two excited states as given by EE-EOM-CCSD are shown by the dotted vertical lines.

Table 4-2: Frequency dependent NLSCCSD polarizabilities for poly-glycine. The table shows the total and unit cell values for translationally periodic poly-glycine. The unit cell is a glycine residue. Frequencies are chosen as multiples of the sodium D-line frequency.

$n_{glycine}$	0.0000		0.0773		0.1547		0.2321	
n_{cell}	$\langle\alpha\rangle$	$\langle\alpha\rangle/n_{cell}$	$\langle\alpha\rangle$	$\langle\alpha\rangle/n_{cell}$	$\langle\alpha\rangle$	$\langle\alpha\rangle/n_{cell}$	$\langle\alpha\rangle$	$\langle\alpha\rangle/n_{cell}$
3	52.3967	17.4656	53.2957	17.7652	56.3886	18.7962	64.0370	21.3457
4	69.4862(99.4%)	17.3715	70.7215(99.4%)	17.6804	74.9951(99.2%)	18.7488	85.6499(98.7%)	21.4125
5	86.5756	17.3151	88.1474	17.6295	93.6015	18.7203	107.2627	21.4525
8	137.8439	17.2305	140.4250	17.5531	149.4209	18.6776	172.1013	21.5127
10	172.0228	17.2023	175.2767	17.5277	186.6338	18.6634	215.3271	21.5327
15	257.4701	17.1647	262.4060	17.4937	279.6661	18.6444	323.3914	21.5594
20	342.9173	17.1459	349.5353	17.4768	372.6984	18.6349	431.4557	21.5728
25	428.3645	17.1346	436.6647	17.4666	465.7307	18.6292	539.5201	21.5808
50	855.6007	17.1120	872.3112	17.4462	930.8922	18.6178	1079.8417	21.5968
75	1282.8368	17.1045	1307.9578	17.4394	1396.0538	18.6141	1620.1634	21.6022
100	1710.0730	17.1007	1743.6044	17.4360	1861.2153	18.6122	2160.4851	21.6049
250	4273.4899	17.0940	4357.4839	17.4299	4652.1844	18.6087	5402.4151	21.6097
500	8545.8514	17.0917	8713.9497	17.4279	9303.7996	18.6076	10805.6318	21.6113
∞		17.0894		17.4259		18.6065		21.6129

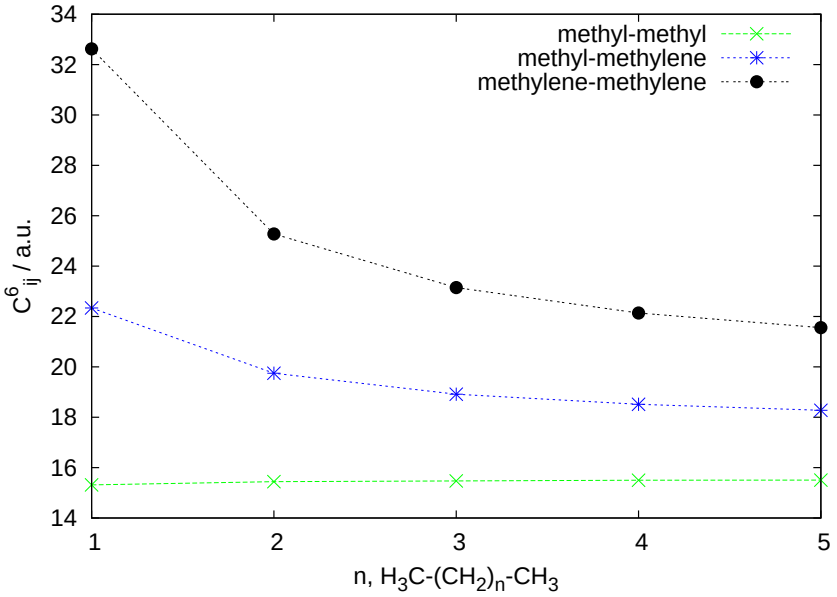

Figure 4-8: Transferability of dispersion coefficients for alkanes. The figure shows the three pair contributions for the interaction of two alkanes. Eight grid-points are used for the numerical integration.

Table 4-3: Dispersion coefficients per methylene unit cell. Values shown are for the interaction of two alkanes shown on the right side of the table. The column on the left gives the QM2 region used in constructing the target results shown on the right. Note that the diagonal elements are the exact result within the basis set used. Eight grid-points are used for the numerical integration.

$QM2$	C_3H_8	C_4H_{10}	C_5H_{12}	C_6H_{14}	C_7H_{16}	$\cdots$	$C_\infty H_\infty$
C_3H_8	20.3592	23.1539	24.9177	26.1297	27.0132	$\cdots$	32.6242
C_4H_{10}		20.0543	21.0499	21.7272	22.2176	$\cdots$	25.2763
C_5H_{12}			19.8873	20.4129	20.7926	$\cdots$	23.1456
C_6H_{14}				19.7885	20.1155	$\cdots$	22.1356
C_7H_{16}					19.7237	$\cdots$	21.5582

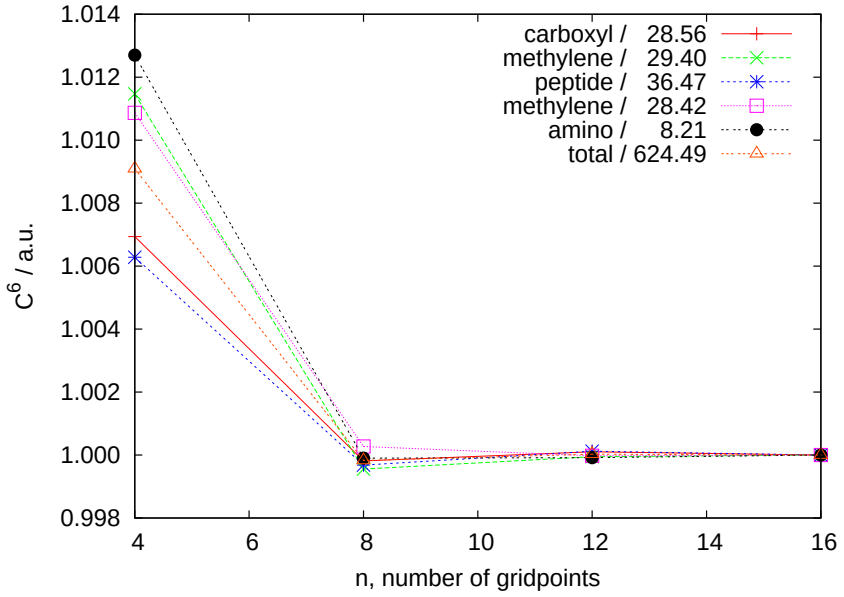

Figure 4-9: Convergence of di-glycine dispersion coefficients with grid-points. The total and diagonal contributions for the interaction of two di-glycines are shown. The coefficients are normalized to the values corresponding to the largest number of grid-points which are shown in the legend.

Table 4-4: Bond dispersion coefficients for di-glycine. Diagonal contributions are shown for two interacting di-glycines. NLMOs are ordered per column as lp and σ orbitals as well as two 3-center carboxyl- and peptide-bonds given by $1\pi_3 - 2\pi_3$. For each type of NLMO within each column the orbital dispersion coefficient contributions are ordered by increasing magnitude. A 16 point grid is used.

carboxyl		*methylene*		*peptide*		*methylene*		*amino*	
$O1$	0.2399	$C5N8$	1.6104	$O3$	0.2743	$C7N9$	1.3419	$N9$	0.6779
$O2$	0.2570	$C4C5$	1.8782	$O3$	0.3130	$C7H14$	1.9014	$N9H16$	1.0548
$O2$	0.3490	$C5H11$	1.9359	$O3C6$	0.1149	$C7H15$	1.9014	$N9H17$	1.0555
$O2C4$	0.0932	$C5H12$	1.9359	$C6N8$	0.7885	$C6C7$	2.0032		
$O1C4$	0.4252			$N8H13$	1.0697				
$O1H10$	0.7193			$1\pi_3$	0.7048				
$1\pi_3$	0.3183			$2\pi_3$	3.8378				
$2\pi_3$	2.0970								

REFERENCES

[1] R. J. Bartlett, Annu. Rev. Phys. Chem. **32**, 359 (1981).

[2] R. J. Bartlett and M. Musiał, Rev. Mod. Phys. **79**, 291 (2007).

[3] J. J. McClellan, T. F. Hughes, and R. J. Bartlett, Int. J. Quantum Chem. **105**, 914 (2005).

[4] M. A. Spackman, J. Chem. Phys. **94**, 1295 (1991).

[5] Y. K. Kang and M. S. Jhon, Theoret. Chim. Acta **61**, 41 (1982).

[6] T. P. Haley, E. R. Graybill, and S. M. Cybulski, J. Chem. Phys. **124**, 204301 (2006).

[7] A. D. Becke and E. R. Johnson, J. Chem. Phys. **122**, 154104 (2005).

[8] P. Jin, J. S. Murray, and P. Politzer, Int. J. Quantum Chem. **106**, 2347 (2006).

[9] K. J. Miller, J. Am. Chem. Soc. **112**, 8533 (1990).

[10] Q. Wu and W. Yang, J. Chem. Phys. **116**, 515 (2002).

[11] A. T. Amos and R. J. Crispin, Mol. Phys. **31**, 159 (1976).

[12] F. Pirani, D. Cappelletti, and G. Liuti, Chem. Phys. Lett. **350**, 286 (2001).

[13] A. Ye and J. Autschbach, J. Chem. Phys. **125**, 234101 (2006).

[14] A. E. Reed and F. Weinhold, J. Chem. Phys. **83**, 1736 (1985).

[15] N. Flocke and R. J. Bartlett, Chem. Phys. Lett. **367**, 80 (2003).

[16] ACES II is a program product of the Quantum Theory Project, University of Florida. Authors: J. F. Stanton, J. Gauss, J. D. Watts, M. Nooijen, N. Oliphant, S. A. Perera, P. G. Szalay, W. J. Lauderdale, S. A. Kucharski, S. R. Gwaltney, S. Beck, A. Balková, D. E. Bernholdt, K. K. Baeck, P. Rozyczko, H. Sekino, C. Hober, and R. J. Bartlett. Integral packages included are VMOL (J. Almlöf and P. R. Taylor); VPROPS (P. Taylor); ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, J. Olsen, and P. R. Taylor).

[17] K. Eskandari, M. Mandado, and R. A. Mosquera, Chem. Phys. Lett. **437**, 1 (2007).

[18] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi, Chem. Phys. Lett. **313**, 701 (1999).

[19] T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayasi, and K. Kitaura, Chem. Phys. Lett. **318**, 614 (2000).

[20] D. G. Fedorov and K. Kitaura, J. Phys. Chem. A **111**, 6904 (2007).

[21] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **120**, 6832 (2004).

[22] D. G. Fedorov, R. M. Olson, K. Kitaura, M. S. Gordon, and S. Koseki, J. Comp. Chem. **25**, 872 (2004).

[23] K. Fukuzawa, K. Kitaura, M. Uebayasi, K. Nakata, T. Kaminuma, and T. Nakano, J. Comp. Chem. **26**, 1 (2005).

[24] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **121**, 2483 (2004).

[25] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **123**, 134103 (2005).

[26] V. Lotrich, N. Flocke, M. Ponton, A. D. Yau, S. A. Perera, E. Deumens, and R. J. Bartlett, J. Chem. Phys. **128**, 194104 (2008).

[27] N. Flocke and V. Lotrich, J. Comp. Chem. **29**, 2722 (2008).

[28] N. Flocke and R. J. Bartlett, J. Chem. Phys. **121**, 10935 (2004).

[29] T. F. Hughes, N. Flocke, and R. J. Bartlett, J. Phys. Chem. A **112**, 5994 (2008).

[30] G. Schaftenaar and J. H. Noordik, J. Comput.-Aided Mol. Design **14**, 123 (2000).

[31] T. Janowski, A. R. Ford, and P. Pulay, J. Chem. Theory Comput. **3**, 1368 (2007).

[32] T. Kinoshita, O. Hino, and R. J. Bartlett, J. Chem. Phys. **119**, 7756 (2003).

[33] T. B. Pedersen, A. M. J. S. de Meras, and H. Koch, J. Chem. Phys. **120**, 8887 (2004).

[34] A. G. Taube and R. J. Bartlett, J. Chem. Phys. **128**, 164101 (2008).

[35] F. R. Manby, H.-J. Werner, T. B. Adler, and A. J. May, J. Chem. Phys. **124**, 094103 (2006).

[36] C. Hampel and H.-J. Werner, J. Chem. Phys. **104**, 6286 (1996).

[37] W. Kohn, Phys. Rev. A **133**, A171 (1964).

[38] M. Schütz, J. Chem. Phys. **113**, 9986 (2000).

[39] P. E. Maslen, A. D. Dutoi, M. S. Lee, Y. Shao, and M. Head-Gordon, Mol. Phys. **103**, 425 (2005).

[40] G. Stollhoff and P. Fulde, J. Chem. Phys. **73**, 4548 (1980).

[41] S. Hirata, I. Grabowski, M. Tobita, and R. J. Bartlett, Chem. Phys. Lett. **345**, 475 (2001).

[42] S. Li, J. Shen, W. Li, and Y. Jiang, J. Chem. Phys. **125**, 074109 (2006).

[43] M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).

[44] G. E. Scuseria and P. Y. Ayala, J. Chem. Phys. **111**, 8330 (1999).

[45] W. D. Laidig, G. D. Purvis, and R. J. Bartlett, J. Phys. Chem. **89**, 2161 (1985).

[46] B. Paulus, K. Rościszewski, H. Stoll, and U. Birkenheuer, Phys. Chem. Chem. Phys. **5**, 5523 (2003).

[47] N. Flocke and R. J. Bartlett, J. Chem. Phys. **118**, 5326 (2003).

[48] S. Li, J. Ma, and Y. Jiang, J. Comp. Chem. **23**, 237 (2002).

[49] W. Li and S. Li, J. Chem. Phys. **121**, 6649 (2004).

[50] A. A. Auer and M. Nooijen, J. Chem. Phys. **125**, 024104 (2006).

[51] S. Hirata, R. Podeszwa, M. Tobita, and R. J. Bartlett, J. Chem. Phys. **120**, 2581 (2004).

[52] H. Stoll, Phys. Rev. B **46**, 6700 (1992).

[53] H. Stoll, B. Paulus, and P. Fulde, J. Chem. Phys. **123**, 144108 (2005).

[54] P. E. Maslen, M. S. Lee, and M. Head-Gordon, Chem. Phys. Lett. **319**, 205 (2000).

[55] M. Schütz, J. Chem. Phys. **116**, 8772 (2002).

[56] M. Schütz and H.-J. Werner, Chem. Phys. Lett. **318**, 370 (2000).

[57] K. Raghavachari, G. Trucks, J. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[58] J. Watts, J. Gauss, and R. J. Bartlett, J. Chem. Phys. **98**, 8718 (1993).

[59] I. Marcotte, F. Separovic, M. Auger, and S. M. Gagné, Biophys. J. **86**, 1587 (2004).

[60] S. Abdali, T. A. Niehaus, K. J. Jalkanen, X. Cao, L. A. Nafie, T. Frauenheim, S. Suhai, and H. Bohr, Phys. Chem. Chem. Phys. **5**, 1295 (2003).

[61] D. G. Fedorov, T. Ishida, M. Uebayasi, and K. Kitaura, J. Phys. Chem. A **111**, 2722 (2007).

[62] T. M. Watson and J. D. Hirst, Phys. Chem. Chem. Phys. **6**, 2580 (2004).

[63] G. Stollhoff and P. Fulde, Z. Phys. B Con. Mat. **26**, 257 (1977).

[64] P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).

[65] H. Stoll, Chem. Phys. Lett. **191**, 548 (1992).

[66] P. E. Maslen and M. Head-Gordon, J. Chem. Phys. **109**, 7093 (1998).

[67] B. Paulus, Chem. Phys. Lett. **371**, 7 (2003).

[68] D. G. Fedorov and K. Kitaura, J. Chem. Phys. **124**, 079904 (2006).

[69] O. Christiansen, P. Manninen, P. Jörgensen, and J. Olsen, J. Chem. Phys. **124**, 084103 (2006).

[70] J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, J. Phys. Chem. **96**, 135 (1992).

[71] M. Jaszuński, A. Rizzo, and P. Jørgensen, Theor. Chem. Acc. **106**, 251 (2001).

[72] D. Maurice and M. Head-Gordon, J. Phys. Chem. **100**, 6131 (1996).

[73] S. R. Gwaltney, M. Nooijen, and R. J. Bartlett, Chem. Phys. Lett. **248**, 189 (1996).

[74] H. Nakatsuji, J.-Y. Hasegawa, and M. Hada, J. Chem. Phys. **104**, 2321 (1996).

[75] J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **98**, 7029 (1993).

[76] C. F. Bunge and R. Carbó-Dorca, J. Chem. Phys. **125**, 014108 (2006).

[77] C. F. Bunge, J. Chem. Phys. **125**, 014107 (2006).

[78] T. D. Crawford and R. A. King, Chem. Phys. Lett. **366**, 611 (2002).

[79] T. Korona and H.-J. Werner, J. Chem. Phys. **118**, 3006 (2003).

[80] D. Kats, T. Korona, and M. Schütz, J. Chem. Phys. **125**, 104106 (2006).

[81] K. Sadeghian and M. Schütz, J. Am. Chem. Soc. **129**, 4068 (2007).

[82] D. Kats, T. Korona, and M. Schütz, J. Chem. Phys. **127**, 064107 (2007).

[83] Y. Mochizuki, S. Koikegami, S. Amari, K. Segawa, K. Kitaura, and T. Nakano, Chem. Phys. Lett. **406**, 283 (2005).

[84] Y. Mochizuki, K. Tanaka, K. Yamashita, T. Ishikawa, T. Nakano, S. Amari, K. Segawa, T. Murase, H. Tokiwa, and M. Sakurai, Theor. Chem. Acc. **117**, 541 (2007).

[85] S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino, Mol. Phys. **103**, 2255 (2005).

[86] G. P. Das, D. S. Dudis, A. T. Yeates, and J.-P. Blaudeau, Chem. Phys. Lett. **438**, 89 (2007).

[87] I. Mayer, Chem. Phys. Lett. **437**, 284 (2007).

[88] P. Surján, Chem. Phys. Lett. **439**, 393 (2007).

[89] T. F. Hughes and R. J. Bartlett, in preparation (2008).

[90] A. Dreuw and M. Head-Gordon, Chem. Rev. **105**, 4009 (2005).

[91] M. Nooijen and R. J. Bartlett, J. Chem. Phys. **106**, 6441 (1997).

[92] J. F. Stanton, J. Chem. Phys. **101**, 8928 (1994).

[93] M. Nooijen and R. J. Bartlett, J. Chem. Phys. **102**, 3629 (1995).

[94] S. Hirata, M. Nooijen, and R. J. Bartlett, Chem. Phys. Lett. **326**, 255 (2000).

[95] M. Musiał, S. A. Kucharski, and R. J. Bartlett, J. Chem. Phys. **118**, 1128 (2003).

[96] T. D. Bouman, B. Voigt, and A. E. Hansen, J. Am. Chem. Soc. **101**, 550 (1979).

[97] G. Doggett, Mol. Phys. **29**, 313 (1975).

[98] J. Stanton and J. Gauss, J. Chem. Phys. **101**, 8938 (1994).

[99] G. S. Tschumper, M. L. Leininger, B. C. Hoffman, E. F. Valeev, H. F. Schaefer III, and M. Quack, J. Chem. Phys. **116**, 690 (2002).

[100] T. F. Hughes and R. J. Bartlett, J. Chem. Phys. **129**, 054105 (2008).

[101] A. J. Stone, *The Theory of Intermolecular Forces*, (Oxford University Press, New York, 1996), Vol. 32.

[102] A. D. Buckingham, Annu. Rev. Phys. Chem. **49**, xiii (1998).

[103] R. J. Bartlett and H. Sekino, *Nonlinear Optical Materials: Theory and Modeling*, (ACS, Washington, 1996), Vol. 628, Chap. 2.

[104] H. Sekino and R. J. Bartlett, J. Chem. Phys. **85**, 976 (1986).

[105] J. Linderberg and Y. Öhrn, *Propagators in Quantum Chemistry*, (Wiley Interscience, Hoboken, 2004).

[106] G. J. B. Hurst, M. Dupuis, and E. Clementi, J. Chem. Phys. **89**, 385 (1988).

[107] J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **99**, 5178 (1993).

[108] P. B. Rozyczko, S. A. Perera, M. Nooijen, and R. J. Bartlett, J. Chem. Phys. **107**, 6736 (1997).

[109] J. F. Stanton, Phys. Rev. A **49**, 1698 (1994).

[110] S. A. Perera, M. Nooijen, and R. J. Bartlett, J. Chem. Phys. **104**, 3290 (1996).

[111] H. Sekino and R. J. Bartlett, Adv. Quantum Chem. **35**, 149 (1999).

[112] C. Hättig, O. Christiansen, and P. Jørgensen, J. Chem. Phys. **107**, 10592 (1997).

[113] O. Christiansen, A. Halkier, H. Koch, P. Jørgensen, and T. Helgaker, J. Chem. Phys. **108**, 2801 (1998).

[114] T. Korona, K. Pflüger, and H.-J. Werner, Phys. Chem. Chem. Phys. **6**, 2059 (2004).

[115] N. J. Russ and T. D. Crawford, Chem. Phys. Lett. **400**, 104 (2004).

[116] T. Korona, M. Przybytek, and B. Jeziorski, Mol. Phys. **104**, 2303 (2006).

[117] Y. Mochizuki, T. Ishikawa, K. Tanaka, H. Tokiwa, T. Nakano, and S. Tanaka, Chem. Phys. Lett. **418**, 418 (2006).

[118] H. B. G. Casimir and D. Polder, Phys. Rev. **73**, 360 (1948).

BIOGRAPHICAL SKETCH

Thomas Frank Hughes was born in 1980 in Long Island, New York to Thomas James and Rose Ann Hughes. He has two older sisters, Erica Christine and Kelly Marie Hughes. He grew up in Spring Hill, Florida and graduated from Central High School in 1998. He received a Bachelors of Science degree in chemistry, along with minors in math and physics, from the University of North Florida in Jacksonville, Florida in 2002. His undergraduate thesis was on kinetic isotope effects. During that time he also did research at Mayo Clinic of Jacksonville, Florida on docking, ligand design, and molecular dynamics studies of cyclic peptides and acetylcholinesterase.

As a result of his high school and undergraduate experiences, he decided to go to graduate school for his doctorate in physical chemistry at the Quantum Theory Project of the University of Florida under the direction of Graduate Research Professor Rodney J. Bartlett. Professor Bartlett's great understanding of theoretical chemistry has lead him to be a very influential figure in developing and applying quantum chemical methods, for example as in his ACES suite of high-level electronic structure programs. Part of the focus of the author's graduate research has been to develop and implement quantum chemical methods for larger molecules.