Development and Application of Embedding Methods for the Simulation of Large Chemical Systems

Thesis by

Taylor A. Barnes

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended April 13, 2015)

To my parents

Acknowledgments

I would like to thank my advisor, Prof. Thomas F. Miller III, for his support, advice, and professional example. Throughout my graduate career, his guidance has equipped me to become a more rigorous investigator and a more cogent communicator, and his dedicated mentorship underpins all of the work presented within this dissertation.

I also wish to thank the faculty of the Division of Chemistry and Chemical Engineering. In particular, I thank the members of my thesis committee, Prof. William Goddard III, Prof. Mitchio Okumura, and Prof. Brent Fultz, for their supportive advice and insight. I would also like to express my memories of Prof. Aron Kuppermann, who enhanced my development as a scientist and as a person through his thoughtful regard for his students and his willingness to share his unique experiences as a pioneer within the field of quantum mechanics.

I thank all of the members of the Caltech staff, without whom my studies would not have been possible. I especially thank Priscilla Boon and Agnes Tong for their diligent administrative support, and Tom Dunn, Zailo Leite, and Naveed Near-Ansari for providing invaluable technical expertise.

My graduate experience would have been incomplete without the aid of the other members of the Miller group, both past and present: Jason, Bin, Nick, Nandini, Romelia, Artur, Josh, Connie, Mike, Fran, Kuba, Frank, Michiel, Joonho, Mark, Matt, Brett, Eric, Reid, Ralph, and Lila. I have been continually impressed by the breadth of their knowledge and the depth of their scientific insights, and offer my thanks for all of their discussions and encouragement throughout my studies.

Finally, I thank my family for their support, and especially wish to thank my parents for their selfless dedication to my development and for teaching me to value knowledge from a young age. They have laid the foundation for all that I am and ever will become.

Abstract

The high computational cost of correlated wavefunction theory (WFT) calculations has motivated the development of numerous methods to partition the description of large chemical systems into smaller subsystem calculations. For example, WFT-in-DFT embedding methods facilitate the partitioning of a system into two subsystems: a subsystem A that is treated using an accurate WFT method, and a subsystem B that is treated using a more efficient Kohn-Sham density functional theory (KS-DFT) method.[1–3] A primary challenge of WFT-in-DFT embedding is the accurate representation of the embedding potential, which describes the inter-subsystem interactions. In many WFT-in-DFT embedding methods, the embedding potential is treated using the approximations of orbital-free DFT, which can lead to large errors in the resulting energies. This dissertation describes the development and application of improved embedding methods that enable accurate and efficient calculation of the properties of large chemical systems.

This work builds upon several key studies. In Ref. 4, a method was described for calculating numerically exact KS-DFT embedding potentials through the use of an optimized effective potential (OEP) step. The first general implementation of this method for many-electron systems was introduced in Ref. 5, with additional challenges associated with the treatment of particularly complex systems being addressed in Ref. 6. Several variations of the OEP-based approach have also been proposed.[7–10] More recently, the projection-based WFT-in-DFT embedding method has been developed, enabling the calculation of numerically exact embedding potentials without the

need for a computationally challenging OEP step. The projection-based embedding method has been published in Ref. 11 and in Ref. 12.

Chapter 1 introduces a generalization of the projection-based WFT-in-DFT embedding method that is suitable for application to the study of large chemical systems. In the original implementation of the projection-based embedding method, all calculations on subsystem A were performed using the basis set of the full system. When the wavefunction of subsystem A is represented using the basis set of the full system, the computational cost of the WFT calculation on subsystem A increases with respect to the size of the full system; this characteristic imposes practical limitations on the applicability of the projection-based embedding method to large chemical systems. In principle, by truncating the basis set representation of the subsystem A wavefunction to include only a subset of the basis functions associated with the full system, it is possible to ensure that the cost of the WFT calculation on subsystem A is independent of the size of the full system. We show that naïve truncation of the basis set associated with subsystem A can lead to large numerical artifacts, and present a method for truncating the basis set associated with subsystem A that enables systematic control of these artifacts. The truncation method is applied to both covalently and non-covalently bound test cases, including water clusters and polypeptide chains, and it is demonstrated that errors associated with basis set truncation are controllable to well within chemical accuracy. This work has been published in Ref. 13.

Chapter 2 describes the application of the projection-based embedding method to investigate the oxidative stability of lithium-ion batteries. A central challenge in the refinement of lithium-ion batteries is the control of cathode-induced oxidative decomposition of electrolyte solvents, such as ethylene carbonate (EC) and dimethyl carbonate (DMC). We study the oxidation potentials of neat EC, neat DMC, and 1:1 mixtures of EC and DMC by using the projection-based embedding approach to

accurately calculate the vertical ionization energy (IE) of individual molecules at the CCSD(T) level of theory, while explicitly accounting for the solvent using a combination of DFT and molecular mechanics interactions. We find that the ensemble-averaged distributions of vertical IEs are consistent with a linear response interpretation of the statistics of the solvent configurations, enabling determination of both the intrinsic oxidation potential of the solvents and the corresponding solvent reorganization energies. Interestingly, we reveal that large contributions to the solvation properties of DMC originate from quadrupolar interactions, resulting in a much larger solvent reorganization energy than that predicted using simple dielectric continuum models. Demonstration that the solvation properties of EC and DMC are governed by fundamentally different intermolecular interactions provides insight into key aspects of lithium-ion batteries, with relevance to electrolyte decomposition processes, solid-electrolyte interphase formation, and the local solvation environment of lithium cations. This work has been published in Ref. 14.

Bibliography

[1] Cortona, P. Self-Consistently Determined Properties of Solids Without Band-Structure Calculations. *Phys. Rev. B* **1991**, *44*, 8454

[2] Wesolowski, T. A.; Warshel, A. Frozen Density-Funtional Approach for *Ab-Initio* Calculations of Solvated Molecules. *J. Phys. Chem.* **1993**, *97*, 8050

[3] Govind, N.; Wang, Y. A.; da Silva, A. J. R.; Carter, E. A. Accurate *Ab Initio* Energetics of Extended Systems Via Explicit Correlation Embedded in a Density Functional Environment. *Chem. Phys. Lett.* **1998**, *295*, 129

[4] Goodpaster, J. D.; Ananth, N.; Manby, F. R.; Miller, T. F., III Exact Non-additive Kinetic Potentials for Embedded Density Functional Theory. *J. Chem. Phys.* **2010**, *133*, 084103

[5] Goodpaster, J. D.; Barnes, T. A.; Miller, T. F., III Embedded Density Functional Theory for Covalently Bonded and Strongly Interacting Subsystems. *J. Chem. Phys.* **2011**, *134*, 164108

[6] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III Density Functional Theory Embedding for Correlated Wavefunctions: Improved Methods for Open-Shell Systems and Transition Metal Complexes. *J. Chem. Phys.* **2012**, *137*, 224113

[7] Roncero, O.; de Lara-Castells, M. P.; Villarreal, P.; Flores, F.; Ortega, J.; Pani-

agua, M.; Aguado, A. An Inversion Technique for the Calculation of Embedding Potentials. *J. Chem. Phys.* **2008**, *129*, 184104

[8] Fux, S.; Jacob, C. R.; Neugebauer; Visscher, L.; Reiher, M. Accurate Frozen-Density Embedding Potentials as a First Step Towards a Subsystem Description of Covalent Bonds. *J. Chem. Phys.* **2010**, *132*, 164101

[9] Nafziger, J.; Wasserman, A. Density-Based Partitioning Methods for Ground-State Molecular Calculations. *J. Phys. Chem. A* **2014**, *118*, 7623

[10] Libisch, F.; Huang, C.; Carter, E. A. Embedded Correlated Wavefunction Schemes: Theory and Applications. *Acc. Chem. Res.* **2014**, *47*, 2768

[11] Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F., III A Simple, Exact Density-Functional-Theory Embedding Scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564

[12] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III Accurate and Systematically Improvable Density Functional Theory Embedding for Correlated Wavefunctions. *J. Chem. Phys.* **2014**, *140*, 18A507

[13] Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller, T. F., III Accurate Basis Set Truncation for Wavefunction Embedding. *J. Chem. Phys.* **2013**, *139*, 024103

[14] Barnes, T. A.; Kaminski, J.; Borodin, O.; Miller, T. F., III *Ab Initio* Characterization of the Electrochemical Stability and Solvation Properties of Condensed-Phase Ethylene Carbonate and Dimethyl Carbonate Mixtures. *J. Phys. Chem.* **2015**, *119*, 3865

Contents

List of Figures

List of Tables

Chapter 1

Accurate basis set truncation for wavefunction embedding

1.1 Introduction

The computational cost of electronic structure calculations has motivated the development of methods to partition the description of large systems into smaller subsystem calculations. Among these are the QM/MM,[1–6] ONIOM,[7,8] fragment molecular orbital (FMO),[9–15] and WFT-in-DFT embedding[16–30] approaches, which allow for the treatment of systems that would not be practical using conventional wavefunction theory (WFT) approaches. In particular, WFT-in-DFT embedding utilizes the theoretical framework of density functional theory (DFT) embedding to enable the WFT description of a given subsystem in the effective potential that is created by the remaining electronic density of the system.[16–30] We recently introduced a simple, projection-based method for performing accurate WFT-in-DFT embedding calculations[30] that avoids the need for a numerically challenging optimized effective potential (OEP) calculation[24,25,31–34] via the introduction of a level-shift operator. It was shown that this method enables the accurate calculation of WFT-in-DFT subsystem correlation energies, as well as many-body expansions (MBEs) of the total WFT correlation energy.[30]

In our original implementation, projection-based embedding was performed in

the supermolecular basis, such that the embedded subsystem electronic structure calculation is performed in the atomic orbital (AO) basis set of the full system.[30] From a computational efficiency standpoint, this is not ideal. Although the embedded subsystem calculation has fewer occupied MOs than that performed over the full system, the number of virtual MOs is not reduced. The cost of traditional WFT methods typically depends more strongly on the number of virtual MOs than on the number of occupied MOs; for example, the CCSD(T) method scales as o^3v^4, where o and v indicate the number of occupied and virtual MOs, respectively.[35] Truncation of the AO basis set in which the embedded subsystem is represented would lead to a reduction in the number of virtual MOs, thus significantly reducing the computational cost of the embedded subsystem calculation.

In the current work, we present a method for accurately truncating the AO basis set for embedded subsystem calculations, and we demonstrate its accuracy for both covalently and non-covalently bound systems. It is shown that this approach provides a means of controlling truncation errors and of systematically switching between existing approximate embedding methods and rigorous projection-based embedding. Furthermore, we present both embedded WFT calculations and embedded MBE calculations for molecular clusters and polypeptides.

1.2 Projection-Based Embedding

We now review the projection-based embedding method,[30] which provides a rigorous framework for embedding either a WFT subsystem description in a self-consistent field (SCF) environment (WFT-in-SCF embedding) or an SCF subsystem description in an SCF environment (SCF-in-SCF embedding). The method builds upon earlier ideas to maintain orthogonality between subsystem orbitals, including frozen-core approximations,[36] the Philips-Kleinman pseudopotential approach,[37] the incre-

mental scheme of Stoll *et al.*,[38] the region method of Mata *et al.*,[39] and Henderson's embedding scheme.[40]

In projection-based embedding, an SCF calculation (either HF or Kohn-Sham (KS)-DFT) is first performed over the full system. The resulting set of occupied MOs, $\{\phi_i\}$, is then optionally rotated before it is partitioned into sets $\{\phi_i\}_{\mathrm{A}}$ and $\{\phi_i\}_{\mathrm{B}}$, which correspond to subsystems A and B, respectively. These two sets of orbitals are used to construct the respective subsystem density matrices in the AO basis set, γ^{A} and γ^{B}.

In the embedded subsystem calculation, orthogonality between the subsystem MOs is enforced via the addition of a projection operator, $\mathbf{P}^{\mathrm{B}}$, to the subsystem A embedded Fock matrix, such that

$$\mathbf{f}^{\mathrm{A}} = \mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] + \mathbf{g}[\gamma^{\mathrm{A}}_{\mathrm{emb}}], \tag{1.1}$$

where the embedded core Hamiltonian is

$$\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] = \mathbf{h} + \mathbf{g}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] - \mathbf{g}[\gamma^{\mathrm{A}}] + \mu\mathbf{P}^{\mathrm{B}}, \tag{1.2}$$

$\mathbf{h}$ is the standard one-electron core Hamiltonian, $\mathbf{g}$ includes all two-electron terms, and μ is a level-shift parameter; $\gamma^{\mathrm{A}}_{\mathrm{emb}}$ is the density matrix associated with the MO eigenfunctions of $\mathbf{f}^{\mathrm{A}}$. The projection operator is given by

$$P^{\mathrm{B}}_{\alpha\beta} \equiv \langle b_\alpha | \left\{ \sum_{i \in \mathrm{B}} |\phi_i\rangle\langle\phi_i| \right\} |b_\beta\rangle, \tag{1.3}$$

where the b_α are the AO basis functions and the summation spans the MOs in $\{\phi_i\}_{\mathrm{B}}$. In the limit of $\mu \to \infty$, the MOs of subsystem A are constrained to be mutually orthogonal with those of subsystem B.[36–43] Enforcement of this orthogonality condition eliminates the need for an OEP calculation, since non-additive contributions to the

kinetic energy vanish in this limit. The embedded SCF calculation using the Fock matrix in Eq. 2.4 is iterated to self-consistency with respect to $\gamma^{\mathrm{A}}_{\mathrm{emb}}$. The energy of the resulting SCF-in-SCF embedding calculation is then

$$\begin{aligned} E_{\mathrm{SCF}}[\gamma^{\mathrm{A}}_{\mathrm{emb}};\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] = E_{\mathrm{SCF}}[\gamma^{\mathrm{A}}_{\mathrm{emb}}] + E_{\mathrm{SCF}}[\gamma^{\mathrm{B}}] + E^{\mathrm{nad}}_{\mathrm{SCF}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] \\ + \mathrm{tr}\left[(\gamma^{\mathrm{A}}_{\mathrm{emb}} - \gamma^{\mathrm{A}})(\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] - \mathbf{h})\right], \end{aligned} \tag{1.4}$$

where E_{SCF} is the SCF energy and $E^{\mathrm{nad}}_{\mathrm{SCF}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]$ is the non-additive interaction energy between the densities γ^{A} and γ^{B}. The last term in Eq. 1.4 is a first-order correction to the difference between $E^{\mathrm{nad}}_{\mathrm{SCF}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]$ and $E^{\mathrm{nad}}_{\mathrm{SCF}}[\gamma^{\mathrm{A}}_{\mathrm{emb}},\gamma^{\mathrm{B}}]$.[25] For $\mu \to \infty$, the SCF-in-SCF embedding energy is identical to the energy of the corresponding SCF calculation performed over the full system; as a result, the projection-based approach is numerically exact for SCF-in-SCF embedding calculations. In our previous work,[30] we introduced an additional perturbative correction to the SCF-in-SCF energy to account for the finite value of μ in a given computation; this correction is typically far smaller than the energy differences discussed in the current paper and is thus neglected throughout.

For the special case of DFT-in-DFT embedding, the two-electron potential terms include contributions from the electron-electron electrostatic repulsion and exchange-correlation, such that

$$\mathbf{g}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] = \mathbf{J}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] + \mathbf{v}_{\mathrm{xc}}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}]. \tag{1.5}$$

The associated non-additive interaction energy is

$$E^{\mathrm{nad}}_{\mathrm{SCF}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] = J^{\mathrm{nad}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] + E^{\mathrm{nad}}_{\mathrm{xc}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}], \tag{1.6}$$

where

$$J^{\mathrm{nad}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] = \int \mathrm{d}r_1 \int \mathrm{d}r_2 \frac{\gamma^{\mathrm{A}}(1)\gamma^{\mathrm{B}}(2)}{r_{12}} \tag{1.7}$$

and

$$E_{\mathrm{xc}}^{\mathrm{nad}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] = E_{\mathrm{xc}}[\gamma^{\mathrm{A}}+\gamma^{\mathrm{B}}] - E_{\mathrm{xc}}[\gamma^{\mathrm{A}}] - E_{\mathrm{xc}}[\gamma^{\mathrm{B}}]. \tag{1.8}$$

Evaluation of $J^{\mathrm{nad}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]$ is straightforward, and although the exact form of $E_{\mathrm{xc}}^{\mathrm{nad}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]$ is not known, approximate exchange-correlation (XC) functionals are well-established. Eq. 1.6 does not include any contributions from the non-additive kinetic energy (NAKE), $T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]$, as this term vanishes due to the explicit mutual orthogonalization of the subsystem MOs. Similarly, the special case of HF-in-HF embedding is obtained by replacing the exchange-correlation potential and energy functionals, $\mathbf{v}_{\mathrm{xc}}[\gamma^{\mathrm{A}}+\gamma^{\mathrm{B}}]$ in Eq. 1.5 and $E_{\mathrm{xc}}^{\mathrm{nad}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]$ in Eq. 1.6, with the corresponding HF exchange terms.[30]

Projection-based embedding also allows for WFT-in-SCF embedding, in which subsystem A is treated at the WFT level and subsystem B is described at the SCF level.[30] This simply involves replacing the standard one-electron core Hamiltonian in a WFT calculation with the embedded core Hamiltonian of Eq. 2.5. The electronic energy from the WFT-in-SCF approach is

$$\begin{aligned} E_{\mathrm{WFT}}[\Psi^{\mathrm{A}};\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] = {} & \langle\Psi^{\mathrm{A}}|\hat{H}^{\mathrm{A\ in\ B}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]|\Psi^{\mathrm{A}}\rangle \\ & + E_{\mathrm{SCF}}[\gamma^{\mathrm{B}}] + E_{\mathrm{SCF}}^{\mathrm{nad}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] \\ & - \mathrm{tr}\left[\gamma^{\mathrm{A}}(\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}] - \mathbf{h})\right], \end{aligned} \tag{1.9}$$

where $|\Psi^{\mathrm{A}}\rangle$ is the embedded wavefunction from the WFT-in-SCF embedding calculation and $\hat{H}^{\mathrm{A\ in\ B}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]$ is the Hamiltonian resulting from replacing the standard core Hamiltonian with the embedded core Hamiltonian. Because the term

$\mathrm{tr}\left[\gamma^{\mathrm{A}}_{\mathrm{emb}}(\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^{\mathrm{A}},\gamma^{\mathrm{B}}]-\mathbf{h})\right]$ is included in the first term of Eq. 2.7, it does not appear in the last term, unlike Eq. 1.4.

1.3 AO Basis Set Truncation

1.3.1 The Challenges of AO Basis Set Truncation

Practical implementation of WFT-in-DFT embedding for large systems requires truncation of the AO basis set for the subsystem that is described at the WFT level of theory. We now illustrate the challenges of this task by analyzing the errors that arise from truncation of the AO basis set; in particular, we show that significant numerical errors can arise due to the difficulty of constructing MOs in the truncated AO basis set that are sufficiently orthogonal to the projected MOs in subsystem B.

Calculations utilizing the truncated AO basis set are referred to as truncated embedding calculations, as opposed to supermolecular embedding calculations for which the AO basis set is not truncated. Specifically, the truncated embedding calculation for subsystem A is performed within an AO basis set, $\{b_\alpha\}_{\mathrm{A}}$, that is a subset of the AO basis set for the full system, $\{b_\alpha\}$. All calculations are performed using the implementation of projection-based embedding in the Molpro software package.[44]

As a starting point, we present a set of HF-in-HF supermolecular embedding calculations against which truncated embedding calculations can be compared. A closed-shell HF calculation is performed on a water hexamer in the BK-1 geometry[45] using the cc-pVDZ basis set;[46,47] all geometries employed in this study are provided in the appendices (Section 1.6-1.8). We number the molecules of the water hexamer as shown in Fig. 2.1(a). Following Pipek-Mezey localization of the canonical HF MOs,[48] subsystem partitioning is performed by assigning the five MOs with the largest Mulliken population on water molecule 1 to $\{\phi_i\}_{\mathrm{A}}$; the remaining MOs are assigned to $\{\phi_i\}_{\mathrm{B}}$. A HF-in-HF embedding calculation is then performed over a range of values

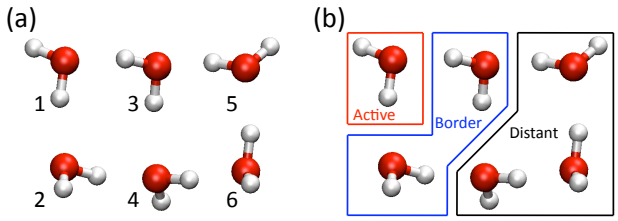

Figure 1.1: (a) The BK-1 water hexamer, with molecule numbering indicated. (b) Illustration of the atom sets defined in Section 1.3.2, with one possible choice of the active, border, and distant atoms indicated.

for the level-shift parameter μ.

The solid line in Fig. 1.2(a) presents the μ-dependence of the HF-in-HF embedding error,

$$E_{\mathrm{err}}^{\mathrm{HF}} \equiv E_{\mathrm{emb}}^{\mathrm{HF}} - E_{\mathrm{full}}^{\mathrm{HF}}, \tag{1.10}$$

where $E_{\mathrm{emb}}^{\mathrm{HF}}$ is the energy of the HF-in-HF embedding calculation, and $E_{\mathrm{full}}^{\mathrm{HF}}$ is the energy of the HF calculation performed over the full system. As previously observed,[30] the error in the SCF-in-SCF supermolecular embedding calculations is sub-microhartree and varies little with respect to μ over several orders of magnitude.

The dashed line in Fig. 1.2(a) shows the results of a naive HF-in-HF truncated embedding calculation, in which $\{b_\alpha\}_{\mathrm{A}}$ is defined to include only the AO basis functions centered on the atoms in water molecules 1, 2, and 3. Calculation of the HF MOs, $\{\phi_i\}$, and the subsystem density matrices, γ^{A} and γ^{B}, is performed in the supermolecular basis, $\{b_\alpha\}$. The embedded core Hamiltonian in Eq. 2.5 is initially constructed in the supermolecular basis, after which all matrix elements in $\mathbf{h}^{\mathrm{A\ in\ B}}$ that do not correspond to the truncated AO basis are discarded. The embedded calculation for subsystem A is then performed in the truncated AO basis. Unlike the supermolecular case, Fig. 1.2(a) illustrates that these naive truncated embedding calculations (solid) produce energies which strongly vary with respect to μ.

The dashed-dotted line and the crosses in Fig. 1.2(a) show the dependence of errors in the truncated embedding calculations with respect to the choice of which

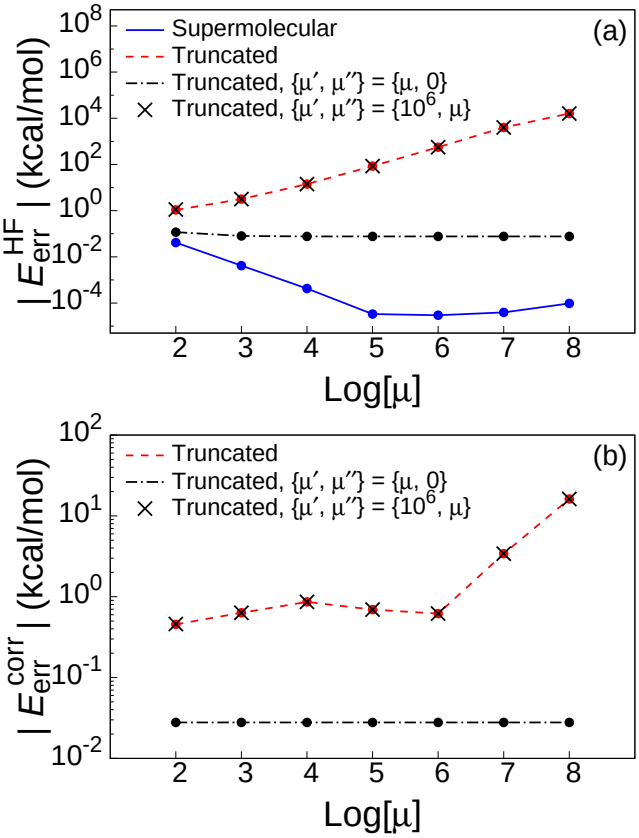

Figure 1.2: (a) HF-in-HF embedding error for molecule 1 of the BK-1 water hexamer. The solid curve provides the supermolecular embedding results, while the results of naive truncation of the AO basis set are shown in the dashed curve. Also shown is the effect of partitioning the projection operator for HF-in-HF embedding in the truncated basis set, with either $\{\mu', \mu''\} = \{\mu, 0\}$ (dashed-dotted) or $\{\mu', \mu''\} = \{10^6, \mu\}$ (crosses). (b) The corresponding truncation error for the CCSD(T)-in-HF truncated embedding calculations.

MOs in subsystem B are projected. In these results, the projection operator is partitioned into two parts, $P^{\mathrm{B}'}_{\alpha\beta}$ and $P^{\mathrm{B}''}_{\alpha\beta}$, each with a different level-shift parameter. The partitioned projection operators are defined as

$$P^{\mathrm{B}'}_{\alpha\beta} \equiv \langle b_\alpha | \left\{ \sum_{i \in \mathrm{B}'} |\phi_i\rangle\langle\phi_i| \right\} |b_\beta\rangle \tag{1.11}$$

and

$$P^{\mathrm{B}''}_{\alpha\beta} \equiv P^{\mathrm{B}}_{\alpha\beta} - P^{\mathrm{B}'}_{\alpha\beta}. \tag{1.12}$$

The summation in Eq. 1.11 is over the set of MOs $\{\phi_i\}_{\mathrm{B}'}$, which is a subset of $\{\phi_i\}_{\mathrm{B}}$. Eq. 1.12 corresponds to the projection of the set of MOs, $\{\phi_i\}_{\mathrm{B}''}$, that consists of all subsystem B MOs that are not included in $\{\phi_i\}_{\mathrm{B}'}$. The resulting embedded core Hamiltonian (from Eq. 2.5) is

$$\begin{aligned} \mathbf{h}^{\mathrm{A\ in\ B}} = \mathbf{h} + \mathbf{g}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] - \mathbf{g}[\gamma^{\mathrm{A}}] + \mu'\mathbf{P}^{\mathrm{B}'} \\ + \mu''\mathbf{P}^{\mathrm{B}''}. \end{aligned} \tag{1.13}$$

In these calculations, a particular MO in $\{\phi_i\}_{\mathrm{B}}$ is assigned to $\{\phi_i\}_{\mathrm{B}'}$ only if its combined Mulliken population on the basis functions centered on water molecules 2 and 3 is greater than 0.5, such that only the 10 (doubly-occupied) MOs in subsystem B that are localized on water molecules 2 and 3 are included. Setting μ' to a positive value while $\mu'' = 0$ corresponds to projecting only the MOs that are localized within the truncated AO basis set, $\{b_\alpha\}_{\mathrm{A}}$.

As illustrated by the dashed-dotted curve in Fig. 1.2(a), the error in the truncated embedding calculation exhibits very little dependence on μ', which suggests that the μ-dependence observed in the dashed curve is caused primarily by projection of the subsystem B MOs that are not localized within the AO basis set accessible to subsystem A. This conclusion is also supported by the set of crosses, which shows the

effect of changing μ'' while leaving μ' fixed at 10^6.

The results from Fig. 1.2(a) may seem counterintuitive, since the overlap between $\{\phi_i\}_{\mathrm{B}''}$ and the truncated AO basis set is much smaller than the overlap between $\{\phi_i\}_{\mathrm{B}'}$ and the truncated AO basis set; it might be expected that projection of the MOs in $\{\phi_i\}_{\mathrm{B}''}$ would have little impact on the truncated embedding calculation. However, the observed behavior can be understood in terms of the difficulty of constructing MOs that are orthogonal to $\{\phi_i\}_{\mathrm{B}''}$ within the truncated Hilbert space of subsystem A. Because the orbitals that are projected by μ'' do not strongly overlap with the basis functions accessible to subsystem A, achieving orthogonality between the subsystem A MOs and $\{\phi_i\}_{\mathrm{B}''}$ places severe demands on the diffuse functions of the truncated AO basis set; in the supermolecular basis set, this difficulty is eliminated. For cases in which the truncated basis set is insufficiently flexible to construct MOs that are effectively orthogonal to $\{\phi_i\}_{\mathrm{B}''}$, the error in the truncated embedding calculation increases linearly with the level-shift parameter μ''.

Fig. 1.2(b) shows that the same trends hold for WFT-in-HF embedding. The figure plots the truncation error in the correlation energy of the WFT-in-HF embedding calculations,

$$E_{\mathrm{err}}^{\mathrm{corr}} \equiv E_{\mathrm{trunc}}^{\mathrm{corr}} - E_{\mathrm{super}}^{\mathrm{corr}}, \tag{1.14}$$

where $E_{\mathrm{trunc}}^{\mathrm{corr}}$ is the correlation energy of a WFT-in-HF truncated embedding calculation (i.e., the difference between the WFT-in-HF and HF-in-HF embedding energies) and $E_{\mathrm{super}}^{\mathrm{corr}}$ is the correlation energy of a WFT-in-HF supermolecular embedding calculation obtained with the same choices of $\{\phi_i\}_{\mathrm{B}}$ and μ'. In the supermolecular embedding calculation, all members of $\{\phi_i\}_{\mathrm{B}}$ are assigned to $\{\phi_i\}_{\mathrm{B}'}$. The correlation energy is defined in the standard way:

$$E^{\mathrm{corr}} \equiv E^{\mathrm{WFT}} - E^{\mathrm{HF}}. \tag{1.15}$$

The WFT calculations in Fig. 1.2(b) are performed at the CCSD(T) level of theory,[49] and the subsystems are partitioned as in the corresponding HF-in-HF embedding calculations. As observed for the HF-in-HF truncated embedding calculations, the errors of the CCSD(T)-in-HF truncated embedding calculations exhibit very little dependence on μ' and strong dependence on μ''.

Taken together, the results in Fig. 1.2 illustrate that significant numerical artifacts arise from the enforcement of orthogonality between the MOs of subsystem A in the truncated basis set and the MOs of subsystem B that are localized outside of the truncated AO basis set. Projection of $\{\phi_i\}_{\mathrm{B}''}$ leads to significant errors, as well as dependence upon the level-shift parameter (Figs. 1.2(a) and 1.2(b), crosses). This problem is avoided by setting $\mu'' = 0$ in Eq. 1.13, resulting in truncated embedding calculations that exhibit both good accuracy and very little dependence on the remaining level-shift parameter, μ' (Figs. 1.2(a) and 1.2(b), dashed-dotted curve).

1.3.2 An Improved AO Basis Set Truncation Algorithm

Incorporating the observations from Section 1.3.1, we now present an algorithm for AO basis set truncation in projection-based embedding that avoids dependence on the level-shift parameters and that yields controllable error with respect to the size of the truncated basis set. Truncated embedding calculations require specification of (*i*) the subsystem B MOs, $\{\phi_i\}_{\mathrm{B}}$, (*ii*) the set of AO basis functions in which subsystem A is solved, $\{b_\alpha\}_{\mathrm{A}}$, and (*iii*) the set of subsystem B MOs that are to be projected, $\{\phi_i\}_{\mathrm{B}'}$. In the new algorithm, these specifications are made via the respective selection of (*i*) a set of "active atoms" that are associated with subsystem A, (*ii*) a set of "border atoms" that lie at the interface of subsystems A and B, and (*iii*) an MO overlap threshold parameter, τ.

The set of active atoms is used to determine $\{\phi_i\}_{\mathrm{B}}$. An SCF calculation is performed over the full system using either HF theory or KS-DFT, followed by localiza-

tion of the MOs; we employ the Pipek-Mezey localization method throughout this paper. An MO is assigned to $\{\phi_i\}_{\mathrm{B}}$ if and only if the atom on which the MO has the largest Mulliken population is not an active atom. For the BK-1 water hexamer, one example of a choice of active atoms is provided in Fig. 2.1(b).

The set of border atoms is used to determine $\{b_\alpha\}_{\mathrm{A}}$. Only AO basis functions centered on either an active atom or a border atom are included in $\{b_\alpha\}_{\mathrm{A}}$. Any atom that is not assigned to either the set of active atoms or the set of border atoms is assigned to the set of "distant atoms." The special case in which no atoms are included in the set of border atoms is equivalent to using the monomolecular basis, while the special case in which no atoms are included in the set of distant atoms corresponds to using the supermolecular basis. An example of one possible choice of border atoms is given in Fig. 2.1(b).

The overlap threshold parameter τ is used to determine $\{\phi_i\}_{\mathrm{B}'}$. A given MO in $\{\phi_i\}_{\mathrm{B}}$ is assigned to $\{\phi_i\}_{\mathrm{B}'}$ if it exhibits a combined electronic population on the border atoms, N_i, such that $|N_i| > \tau$; for the purpose of determining the electronic population on individual atoms, we employ Mulliken population analysis throughout this paper. For the special case of $\tau = 0$, all MOs in $\{\phi_i\}_{\mathrm{B}}$ are assigned to $\{\phi_i\}_{\mathrm{B}'}$, whereas sufficiently large values of τ correspond to assigning no MOs to $\{\phi_i\}_{\mathrm{B}'}$.

Figs. 1.3(a) and 1.3(b) illustrate the effect of τ on the number of projected MOs and on the accuracy of HF-in-HF truncated embedding calculations, respectively. The calculations are performed using the BK-1 water hexamer geometry, and the sets of active and border atoms correspond to the case shown in Fig. 2.1(b). The level-shift parameters are set to $\{\mu', \mu''\} = \{10^6, 0\}$, and HF-in-HF truncated embedding calculations using the cc-pVDZ basis set (i.e., HF-in-HF/cc-pVDZ truncated embedding calculations) are performed over a range of τ. These calculations correspond to changing the number of projected MOs, while leaving the size of the truncated AO basis set unchanged. As τ approaches zero, the number of MOs in $\{\phi_i\}_{\mathrm{B}'}$ approaches the total

number of MOs in $\{\phi_i\}_{\mathrm{B}}$ (Fig. 1.3(a)). As more MOs are added to $\{\phi_i\}_{\mathrm{B}'}$, the error increases substantially (Fig. 1.3(b)); this is consistent with the previous observation that projection of the subsystem B MOs not localized within $\{b_\alpha\}_{\mathrm{A}}$ results in large errors (Fig. 1.2, crosses). For very large values of τ, the error in Fig. 1.3(b) increases substantially due to "charge leakage," which is discussed later in this section and in Section 1.3.3.

Fig. 1.3(c) illustrates the sensitivity of HF-in-HF truncated embedding calculations to the size of the truncated AO basis set. The calculations use the set of active atoms indicated in Fig. 2.1(b) and $\{\mu', \mu'', \tau\} = \{10^6, 0, 0.5\}$. The set of border atoms for each calculation is determined through the use of a cutoff parameter, $R_{\text{O-O}}$. If the oxygen atom of a particular water molecule is within a distance $R_{\text{O-O}}$ of an active oxygen atom, all atoms in that water molecule are included in the set of border atoms; the set of border atoms for each value of $R_{\text{O-O}}$ in Fig. 1.3(c) is indicated in Table 1.1. Fig. 1.3(c) illustrates that the truncated embedding calculation converges rapidly with respect to the number of border atoms.

$R_{\text{O-O}}$	Molecules
2.0 Å	
3.0 Å	2, 3
4.2 Å	2, 3, 4
5.0 Å	2, 3, 4, 5
6.0 Å	2, 3, 4, 5, 6

Table 1.1: List of water molecules, the atoms of which comprise the set of border atoms for each value of $R_{\text{O-O}}$ in Fig. 1.3(c). At $R_{\text{O-O}} = 3.0$ Å, the set of border atoms is the same as that shown in Fig. 2.1(b).

Although the algorithm described in this section works well for relatively compact AO basis sets, such as the cc-pVDZ basis set used in all calculations up this point, it exhibits convergence failure for calculations that employ more diffuse basis sets, such as the aug-cc-pVDZ basis set. This is due to the well-known problem of charge leakage, in which the neglect of repulsive interactions in an embedding calculation

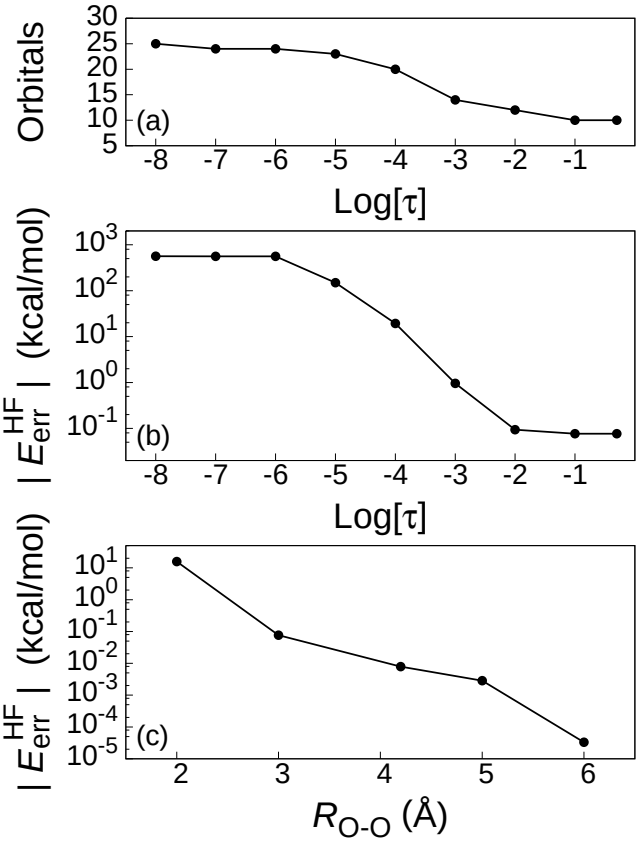

Figure 1.3: (a) The number of MOs assigned to $\{\phi_i\}_{\mathrm{B}'}$ as a function of τ for the BK-1 conformation of the water hexamer. The sets of active and border atoms correspond to the case shown in Fig. 2.1(b). (b) The absolute error in the HF-in-HF embedding calculation as a function of τ. The data point on the far right is equivalent to the dashed-dotted curve in Fig. 1.2(a) at $\mu' = 10^6$, while the data point on the far left is equivalent to the cross at $\mu'' = 10^6$. Thus changing τ corresponds to switching between the dashed-dotted curve and the set of crosses in Fig. 1.2(a). (c) The absolute error in the HF-in-HF embedding calculation as a function of the border atom cutoff, $R_{\text{O-O}}$.

allows for the improper transfer of electron density from the embedded subsystem to the surrounding environment.[50–52] As we show in the Section 1.3.3, this problem can be remedied in the context of truncated projection-based embedding.

1.3.3 Switching Between Orbital Projection and Approximation of the NAKP

To address the problem of charge leakage in truncated embedding calculations employing diffuse basis sets, we include a simple modification to the truncated embedding algorithm from Section 1.3.2. Because that algorithm does not fully enforce mutual orthogonality between the subsystem A MOs and the MOs in $\{\phi_i\}_{\mathrm{B}''}$, the NAKE between the corresponding electronic densities is non-zero. Accounting for this NAKE contribution requires modification of the embedded core Hamiltonian in Eq. 1.13, such that

$$\begin{aligned}\mathbf{h}^{\mathrm{A\ in\ B}} \approx \mathbf{h} + \mathbf{g}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] - \mathbf{g}[\gamma^{\mathrm{A}}] + \mu' \mathbf{P}^{\mathrm{B}'} \\ + \mathbf{v}^{\mathrm{A}}_{\mathrm{NAKP}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}],\end{aligned} \tag{1.16}$$

where $\gamma^{\mathrm{B}''}$ is the density matrix corresponding to the subsystem B MOs in $\{\phi_i\}_{\mathrm{B}''}$, and the non-additive kinetic potential (NAKP) is

$$\mathbf{v}^{\mathrm{A}}_{\mathrm{NAKP}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}] = \mathbf{v}_{\mathrm{T_s}}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}''}] - \mathbf{v}_{\mathrm{T_s}}[\gamma^{\mathrm{A}}]. \tag{1.17}$$

The corresponding SCF-in-SCF energy from Eq. 1.4 is then

$$\begin{aligned}E_{\mathrm{SCF}}[\gamma^{\mathrm{A}}_{\mathrm{emb}}; \gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] \approx E_{\mathrm{SCF}}[\gamma^{\mathrm{A}}_{\mathrm{emb}}] + E_{\mathrm{SCF}}[\gamma^{\mathrm{B}}] \\ + E^{\mathrm{nad}}_{\mathrm{SCF}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] + T^{\mathrm{nad}}_{\mathrm{s}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}] \\ + \mathrm{tr}\left[(\gamma^{\mathrm{A}}_{\mathrm{emb}} - \gamma^{\mathrm{A}})(\mathbf{h}^{\mathrm{A\ in\ B}} - \mathbf{h})\right],\end{aligned} \tag{1.18}$$

where

$$T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}] = T_{\mathrm{s}}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}''}] - T_{\mathrm{s}}[\gamma^{\mathrm{A}}] - T_{\mathrm{s}}[\gamma^{\mathrm{B}''}], \tag{1.19}$$

and the corresponding WFT-in-SCF energy from Eq. 2.7 is

$$\begin{aligned} E_{\mathrm{WFT}}[\Psi^{\mathrm{A}}; \gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] \approx & \langle \Psi^{\mathrm{A}} | \hat{H}^{\mathrm{A\ in\ B}} | \Psi^{\mathrm{A}} \rangle \\ & + E_{\mathrm{SCF}}[\gamma^{\mathrm{B}}] + E_{\mathrm{SCF}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] + T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}] \\ & - \mathrm{tr}\left[\gamma^{\mathrm{A}}(\mathbf{h}^{\mathrm{A\ in\ B}} - \mathbf{h})\right]. \end{aligned} \tag{1.20}$$

By construction, the overlap between the MOs in subsystem A and $\{\phi_i\}_{\mathrm{B}''}$ is small; it can thus be expected that currently available approximations to the kinetic energy functional will provide an adequate description of the NAKE.

If all atoms are included in either the set of active or border atoms and if τ is sufficiently small, this approach corresponds to supermolecular projection-based embedding and involves no approximate KE functionals. In the other extreme, if no atoms are included in the set of border atoms, then no MOs are projected and the approach corresponds to the familiar case of monomolecular DFT embedding with the use of an approximate KE functional. The protocol in Eqs. 1.16-1.19 thus allows for the systematic switching between monomolecular DFT embedding and projection-based supermolecular embedding via modulation of τ and the set of border atoms.

To demonstrate this switching, Fig. 1.4 presents a series of truncated embedding calculations on the BK-1 water hexamer using the cc-pVDZ basis set. In each calculation, the active atoms correspond to one of the water molecules, $\{\mu', \tau\} = \{10^6, 0.5\}$, and $\mathbf{v}_{\mathrm{NAKP}}^{\mathrm{A}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}]$ is obtained using the Thomas-Fermi (TF) functional;[53,54] the border atoms are determined using a range of $R_{\mathrm{O\text{-}O}}$. Figs. 1.4(a) and 1.4(b) show the effect of truncation in the HF-in-HF embedding calculations and in the CCSD(T)-in-HF embedding calculations, respectively. In both cases, the results are seen to quickly converge with respect to the number of border atoms.

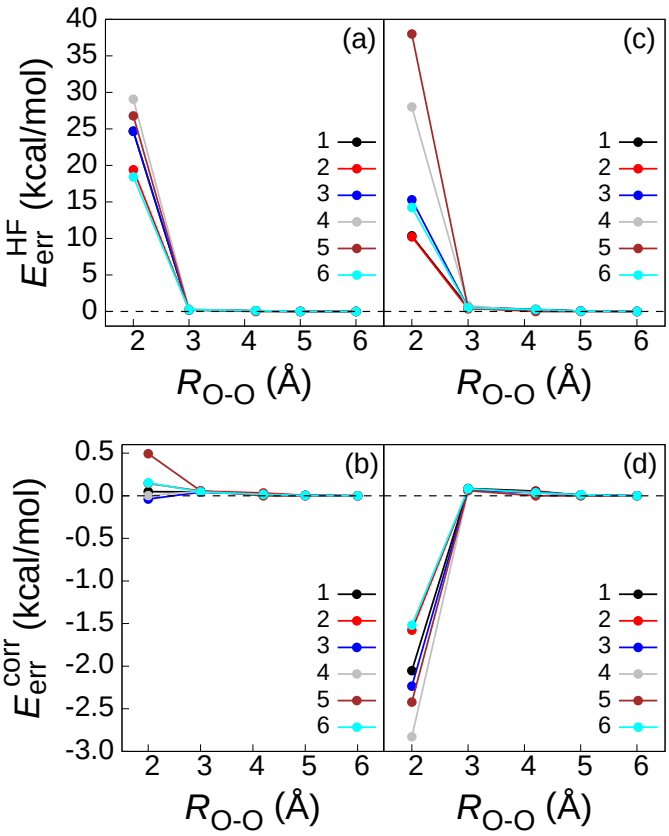

Figure 1.4: (a) The effect of varying $R_{\text{O-O}}$ on the HF-in-HF/cc-pVDZ embedding energy of the BK-1 conformation of the water hexamer. Each curve corresponds to assigning the constituent atoms of the indicated molecule as the set of active atoms. For a cutoff of 2.0 Å, the calculation is equivalent to a monomolecular calculation using TF embedding and no projection operator. At 6.0 Å, all of the calculations are performed in the supermolecular basis, and the projection operator is used exclusively with no approximate functionals. (b) The corresponding CCSD(T)-in-HF/cc-pVDZ results. (c) The corresponding HF-in-HF/aug-cc-pVDZ results. (d) The corresponding CCSD(T)-in-HF/aug-cc-pVDZ results.

In Figs. 1.4(c) and 1.4(d), these calculations are repeated using the larger aug-cc-pVDZ basis set. Again, the results converge rapidly with respect to the number of border atoms. However, the results in Figs. 1.4(c) and 1.4(d) contrast with those discussed in Section 1.3.2, for which truncated embedding with the larger basis set failed due to charge leakage. We thus find that inclusion of the NAKP between the subsystem A MOs and the MOs in $\{\phi_i\}_{\mathrm{B}''}$ helps to mitigate the issue of charge leakage when basis set truncation is employed. This finding is consistent with earlier observations that monomolecular DFT embedding is a useful strategy for mitigating charge leakage.[55–57]

Finally, we note that Eqs. 1.16 and 1.18 can be regarded as a pairwise approximation,[31,32] such that

$$\begin{aligned} T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}'} + \gamma^{\mathrm{B}''}] \approx T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}'}] \\ + T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}]. \end{aligned} \tag{1.21}$$

In the limit of $\mu' \to \infty$, the embedded subsystem MOs and the MOs in $\{\phi_i\}_{\mathrm{B}'}$ are constrained to be mutually orthogonal for all γ^{A}; subject to this constraint, $T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}'}] = 0$ for all γ^{A}, and

$$\mathbf{v}_{\mathrm{NAKP}}^{\mathrm{A}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}'}] = \frac{\delta T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}'}]}{\delta \gamma^{\mathrm{A}}} = 0. \tag{1.22}$$

Therefore, the only nonzero contribution to the NAKP is $\mathbf{v}_{\mathrm{NAKP}}^{\mathrm{A}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}]$ (Eq. 1.16), and the only contribution to the NAKE is $T_{\mathrm{s}}^{\mathrm{nad}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}]$ (Eq. 1.18).

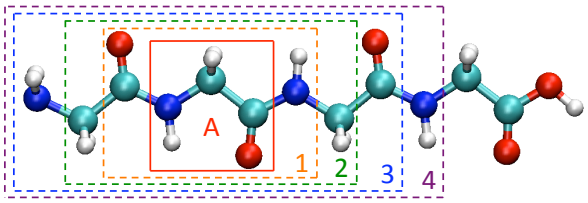

Figure 1.5: The Gly-Gly-Gly-Gly tetrapeptide, with the set of active atoms composing the Gly2 residue (solid red box). Each of the dashed boxes indicates the union of the sets of active and border atoms for the corresponding value of n_t; any atoms outside of the boxes are included in the set of distant atoms.

1.4 Applications

1.4.1 WFT-in-HF Truncated Embedding for Polypeptides

For a more demanding illustration of the truncated embedding approach presented in Section 1.3.3, we consider the Gly-Gly-Gly-Gly tetrapeptide. The optimized geometry of the tetrapeptide is determined at the HF/cc-pVDZ level, with all backbone dihedral angles constrained to 180°. For each of the truncated embedding calculations in this section, the set of active atoms consists of the atoms of one of the four glycine residues. The set of border atoms for each truncated embedding calculation is specified by a cutoff, n_t. If a backbone atom is within n_t bonds of an active atom, then it is included in the set of border atoms; if a non-backbone moiety (i.e., H, O, or OH) is bonded to a border atom, then its associated atoms are likewise included in the set of border atoms. Several sets of border atoms, each corresponding to a different value of n_t, are illustrated in Fig. 1.5 for the case in which the atoms of the Gly2 residue comprise the set of active atoms.

Fig. 1.6 illustrates that WFT-in-HF truncated embedding calculations on this system exhibit significant τ-dependence, since localization of the HF MOs yields orbitals with significant population on two or more backbone atoms. These calculations are performed using MP2-in-HF/aug-cc-pVDZ embedding, with the set of active atoms comprising those in the Gly2 residue, with the set of border atoms associated with

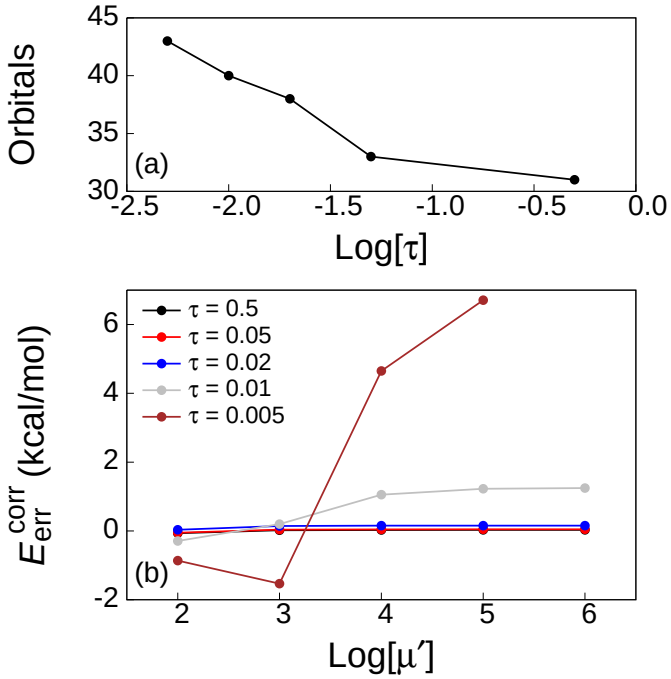

Figure 1.6: (a) τ-dependence of the number of projected orbitals within MP2-in-HF/aug-cc-pVDZ truncated embedding calculations on the Gly-Gly-Gly-Gly tetrapeptide with $n_t = 3$. The choice of active and border atoms is indicated in Fig. 1.5. (b) μ'-dependence of the truncation error of this calculation for several values of τ.

$n_t = 3$, and with $\mathbf{v}^{A}_{NAKP}[\gamma^{A}, \gamma^{B''}]$ obtained using the TF functional. Fig. 1.6(a) shows the number of projected MOs for several values of τ, and the μ'-dependence for each value of τ is shown in part Fig. 1.6(b). For $\tau \geq 0.02$, there is very little μ'-dependence, whereas smaller values of τ lead to greater dependence on the level-shift parameter.

In general, it is preferable to set τ as small as possible without introducing significant μ'-dependence, since this results in fewer orbitals being treated at the level of the approximate KE functional. For all systems considered in this paper, we find that $\tau = 0.05$ results in small μ'-dependencies; all remaining calculations reported in this paper thus employ $\{\mu', \tau\} = \{10^6, 0.05\}$ and utilize the TF functional to approximate $\mathbf{v}^{A}_{NAKP}[\gamma^{A}, \gamma^{B''}]$.

Fig. 1.7 presents additional MP2-in-HF embedding calculations using different sets of active atoms and using a range of values for the border atom cutoff, n_t. The

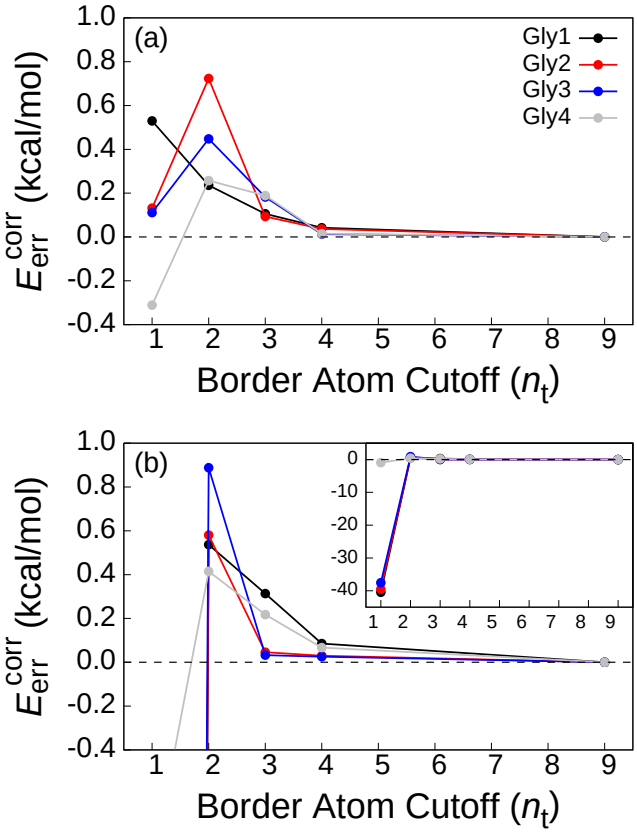

Figure 1.7: (a) Convergence of the truncation error of embedding calculations on the Gly-Gly-Gly-Gly tetrapeptide using the cc-pVDZ basis set and several values of n_t. In each curve, the set of active atoms corresponds to the indicated residue. For $n_t = 9$, there are no distant atoms in any of the calculations. (b) The corresponding calculation using the aug-cc-pVDZ basis set. The inset shows the same results on a larger scale.

set of active atoms associated with each curve corresponds to a different residue in the tetrapeptide. The results in Fig. 1.7(a) and Fig. 1.7(b) are obtained using the cc-pVDZ and aug-cc-pVDZ basis sets, respectively. Both sets of results converge rapidly with respect to the number of border atoms, although it is clear that a minimum of $n_t = 2$ is needed for the calculations with the aug-cc-pVDZ basis set; more diffuse basis functions in the augmented basis set lead to greater overlap between the subsystem A MOs and the MOs in $\{\phi_i\}_{B''}$, thus increasing the contribution from the approximate NAKP functional and yielding a stronger dependence on the border atom cutoff.

1.4.2 Embedded MBE

A promising application domain for projection-based WFT-in-HF embedding is the accurate MBE calculation of WFT energies.[30] This approach has the advantage of avoiding many of the challenges of more traditional MBE methods,[9,58–72] including sensitivity to the parameterization of point charges[73] or the need for "cap-atom" approximations.[74–79] As described previously, we perform the embedded MBE (EMBE) expansion in the correlation energy;[30] inclusion of the 1-body and 2-body terms yields the EMBE2 expression

$$E^{\mathrm{EMBE2}} = \sum_i E_i^{\mathrm{corr}} - \sum_{i>j} (E_{ij}^{\mathrm{corr}} - E_i^{\mathrm{corr}} - E_j^{\mathrm{corr}}), \tag{1.23}$$

where E_i^{corr} is the WFT-in-HF correlation energy of monomer i and E_{ij}^{corr} is the WFT-in-HF correlation energy of the dimer ij.

1.4.2.1 Water Hexamers

EMBE2 calculations at the CCSD(T)-in-HF level are performed on a test set of 11 conformations of the water hexamer, using the 6-31G,[80,81] cc-pVDZ, and aug-cc-pVDZ basis sets. The calculations are performed with $\{\mu', \tau\} = \{10^6, 0.05\}$, and $\mathbf{v}_{\mathrm{NAKP}}^{\mathrm{A}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}''}]$ is obtained using the TF functional. Three of the hexamer geometries are taken from Ref. 82 and correspond to the (1) book, (2) cage, and (3) prism conformations; the other eight are taken from Ref. 45 and correspond to the (4) cyclic boat-1, (5) cyclic boat-2, (6) cyclic chair, (7) cage, (8) book-1, (9) book-2, (10) bag, and (11) an additional prism conformation. This test set includes a mixture of planar (Conf. 4, 5, and 6), quasi-planar (Conf. 1, 8, and 9), and three-dimensional (Conf. 2, 3, 7, and 11) conformations. Each monomer in the EMBE2 calculations corresponds to a set of active atoms comprising the atoms associated with one of the water molecules.

The relative energy of the water hexamer conformations are provided in Fig. 1.8(a), obtained using supermolecular EMBE2 calculations; full CCSD(T) calculations are also reported for comparison. Energies are reported with respect to that of Conf. (11), obtained using the corresponding level of theory and basis set. Fig. 1.8(b) presents the MBE error for each calculation, obtained using

$$E_{\mathrm{err}}^{\mathrm{MBE}} \equiv E^{\mathrm{EMBE2}} - E^{\mathrm{corr}}. \tag{1.24}$$

The mean unsigned MBE error, $\left\langle |E_{\mathrm{err}}^{\mathrm{MBE}}| \right\rangle$, of the EMBE2 calculations performed on this test set is 0.10 kcal/mol for the 6-31G basis set, 0.12 kcal/mol for the cc-pVDZ basis set, and 0.06 kcal/mol for aug-cc-pVDZ. The EMBE2 calculations are thus seen to produce smaller values of $\left\langle |E_{\mathrm{err}}^{\mathrm{MBE}}| \right\rangle$ than similar calculations using point-charge embedding;[67] equally important, however, is the fact that the embedding approach provided here rigorously avoids the problem of charge leakage and the use of arbitrary parameters, and allows for full basis set convergence.

Fig. 1.9 presents the relative energies for the corresponding truncated embedding calculations in the aug-cc-pVDZ basis set. The border atoms are determined in the manner described in Section 1.3.2, using both $R_{\text{O-O}} = 0$ Å (i.e., monomolecular DFT embedding using the TF KE functional) and $R_{\text{O-O}} = 3$ Å. The truncated embedding calculations with $R_{\text{O-O}} = 3$ Å are in far better agreement with the reference supermolecular calculations, thus illustrating the potential of using truncated projection-based embedding to significantly improve upon the accuracy of DFT embedding with approximate KE functionals.

Table 1.2 presents a summary of the EMBE2 results for all three basis sets (6-31G, cc-pVDZ, and aug-cc-pVDZ). The truncated embedding results using a non-empty set of border atoms (i.e., $R_{\text{O-O}} > 0$ Å) consistently produce smaller mean unsigned MBE errors than those obtained in the monomolecular basis (i.e., $R_{\text{O-O}} = 0$ Å).

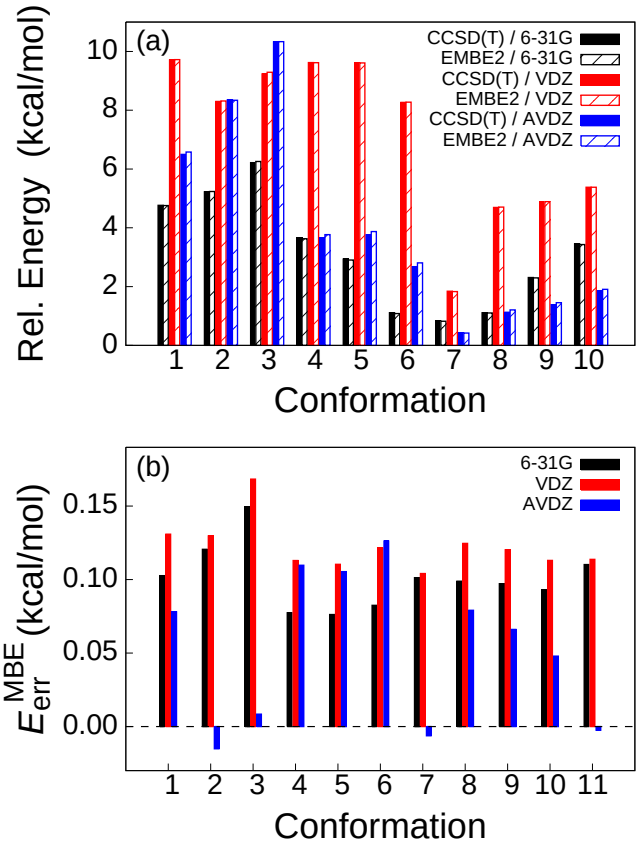

Figure 1.8: (a) Energies of water hexamer conformations obtained using both CCSD(T) over the full system and CCSD(T)-in-HF supermolecular EMBE2 calculations. Three different basis sets are employed, with the cc-pVDZ and aug-cc-pVDZ basis sets abbreviated as VDZ and AVDZ, respectively. Conformation energies are reported with respect to the corresponding calculation for Conf. 11. (b) Error in the energy of the EMBE2 calculations.

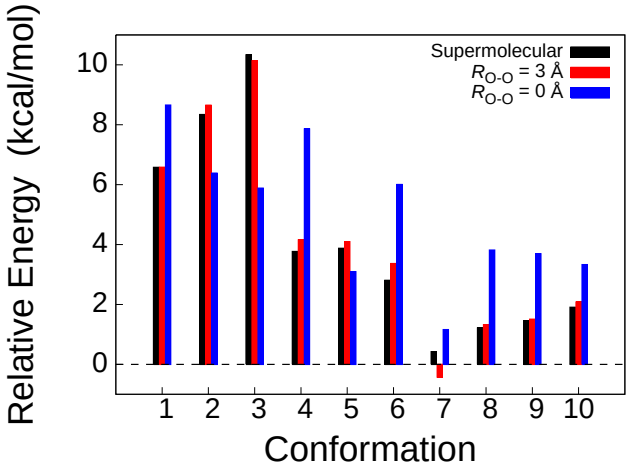

Figure 1.9: Energies of water hexamer conformations obtained using both CCSD(T)-in-HF supermolecular EMBE2 calculations and CCSD(T)-in-HF truncated EMBE2 calculations. The embedding calculations employ truncated embedding with a border atom cutoff of either $R_{\text{O-O}} = 0$ Å or $R_{\text{O-O}} = 3$ Å. Conformation energies are reported with respect to the corresponding calculation for Conf. 11.

The standard deviation of the errors for each set of calculations, $\sigma[E_{\text{err}}^{\text{MBE}}]$, is also provided; this quantity reports on errors in the relative conformation energies, which may be of greater practical relevance than $\left\langle |E_{\text{err}}^{\text{MBE}}| \right\rangle$. For the embedding calculations employing a cutoff of $R_{\text{O-O}} = 0$ Å, $\sigma[E_{\text{err}}^{\text{MBE}}]$ exceeds 1 kcal/mol when the correlation-consistent basis sets are employed. For the embedding calculations employing a cutoff of $R_{\text{O-O}} = 3$ Å, $\sigma[E_{\text{err}}^{\text{MBE}}]$ is approximately 0.4 kcal/mol for each basis set, which is significantly smaller than the errors associated with the finite size of the basis sets (Fig. 1.8). Furthermore, the greater consistency of $\sigma[E_{\text{err}}^{\text{MBE}}]$ across the three basis sets for calculations that employ $R_{\text{O-O}} = 3$ Å rather than $R_{\text{O-O}} = 0$ Å indicates that truncated projection-based embedding provides more consistent errors in the relative energies than DFT embedding with approximate KE functionals. Finally, we note that in the limit of large $R_{\text{O-O}}$ (i.e., supermolecular projection-based embedding) the precision of the results is further improved, in agreement with the expectation of controllable accuracy with respect to the choice of embedding parameters.

	$\langle \lvert E_{\text{err}}^{\text{MBE}} \rvert \rangle$		
Truncation level	6-31G	cc-pVDZ	aug-cc-pVDZ
$R_{\text{O-O}} = 0$ Å	0.54	3.40	6.25
$R_{\text{O-O}} = 3$ Å	0.39	1.03	0.95
Supermolecular	0.10	0.12	0.06
	$\sigma[E_{\text{err}}^{\text{MBE}}]$		
Truncation level	6-31G	cc-pVDZ	aug-cc-pVDZ
$R_{\text{O-O}} = 0$ Å	0.35	6.38	2.53
$R_{\text{O-O}} = 3$ Å	0.42	0.40	0.40
Supermolecular	0.02	0.02	0.05

Table 1.2: Summary of the EMBE2 results on the water hexamer test set. Results are listed using truncated embedding with a cutoff of $R_{\text{O-O}} = 0$ Å, truncated embedding with a cutoff of $R_{\text{O-O}} = 3$ Å, and supermolecular embedding. All values are in kcal/mol.

1.4.2.2 Polypeptides

EMBE2 calculations at the MP2-in-HF level are performed on several conformations of the Gly-Gly-Gly tripeptide using the cc-pVDZ and aug-cc-pVDZ basis sets. The calculations are performed with $\{\mu', \tau\} = \{10^6, 0.05\}$, and $\mathbf{v}_{\text{NAKP}}^{\text{A}}[\gamma^{\text{A}}, \gamma^{\text{B}''}]$ is obtained using the TF functional. The geometries are obtained via optimization at the HF/cc-pVDZ level, with the Gly1-Gly2 bond dihedral (Ω) constrained to several values, and with all other backbone dihedral angles constrained to 180°. Several of these geometries are shown at left in Fig. 1.10. Each monomer in the EMBE2 calculations corresponds to a set of active atoms comprised of one of the tripeptide residues. The sets of border atoms employed in the EMBE2 calculations are defined in terms of the n_{t} cutoff, as described in Section 1.4.1.

Figs. 1.11(a) and 1.11(b) present the correlation energies from EMBE2 calculations on the Gly-Gly-Gly tripeptide conformations using the cc-pVDZ and aug-cc-pVDZ basis sets, respectively; each correlation energy is reported relative to that of the corresponding calculation on the conformation with Ω =180°. It is seen that the truncated embedding calculations reproduce the trends in the relative energies of the reference MP2 calculations, and that the accuracy improves with the number of

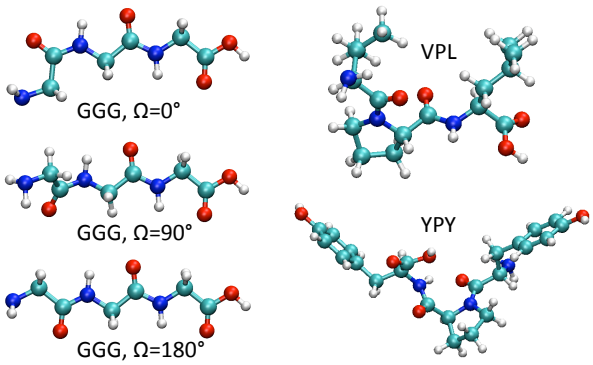

Figure 1.10: Three of the Gly-Gly-Gly (GGG) tripeptide conformations are presented on the left for several different dihedral angles. The geometries of the Val-Pro-Leu (YPL) and Tyr-Pro-Tyr (YPY) tripeptides are presented on the right.

border atoms. Table 1.3 lists the corresponding values of $\left\langle |E_{\mathrm{err}}^{\mathrm{MBE}}| \right\rangle$ and $\sigma[E_{\mathrm{err}}^{\mathrm{MBE}}]$. In agreement with the results in Fig. 1.7, sets of border atoms associated with $\mathrm{n_t} \geq 2$ are needed to achieve suitable accuracy with the aug-cc-pVDZ basis set. Both $\left\langle |E_{\mathrm{err}}^{\mathrm{MBE}}| \right\rangle$ and $\sigma[E_{\mathrm{err}}^{\mathrm{MBE}}]$ are generally found to improve with increasing numbers of border atoms. These results demonstrate that truncated EMBE2 calculations yield accurate results for systems in which embedding is performed across covalently bound monomers. Furthermore, since Ω is associated with rotation of a bond that connects different monomers, these results indicate that the EMBE2 calculations are relatively robust with respect to changes in the electronic environment in the inter-subsystem covalent bonds.

To illustrate the corresponding calculations for tripeptides with different side-chains, additional EMBE2 calculations are performed on the Val-Pro-Leu and Tyr-Pro-Tyr tripeptides. These tripeptides include both hydrophobic and hydrophilic side-chains, including residues with aromatic rings; in particular, we note that the pro-line side-chains present an interesting challenge to the accuracy of the truncated em-bedding calculations, since they exhibit covalent bonds to multiple backbone atoms. Geometries for these tripeptides are optimized at the HF/cc-pVDZ level of theory, with the initial position of the heavy atoms obtained from reported crystal structures

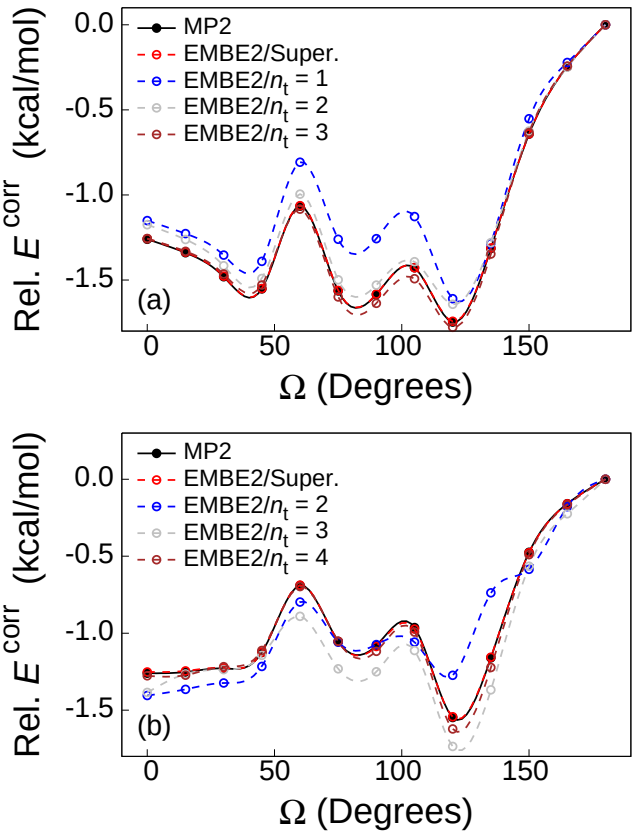

Figure 1.11: (a) Gly-Gly-Gly tripeptide conformation energies obtained using MP2-in-HF EMBE2 calculations and employing the cc-pVDZ basis. Conformation energies are reported with respect to the corresponding calculation for the Ω =180° conformation. The results using $n_t = 4$ are not shown for this basis set, as they are nearly indistinguishable from the supermolecular results. (b) The corresponding results employing the aug-cc-pVDZ basis. The results using $n_t = 1$ are not shown for this basis set, as they are highly inaccurate.

	$\langle \lvert E_{\text{err}}^{\text{MBE}} \rvert \rangle$				
Basis	$n_{\text{t}} = 1$	$n_{\text{t}} = 2$	$n_{\text{t}} = 3$	$n_{\text{t}} = 4$	Super.
VDZ	0.106	0.345	0.103	0.007	0.050
AVDZ	18.549	0.678	0.216	0.054	0.056
	$\sigma[E_{\text{err}}^{\text{MBE}}]$				
Basis	$n_{\text{t}} = 1$	$n_{\text{t}} = 2$	$n_{\text{t}} = 3$	$n_{\text{t}} = 4$	Super.
VDZ	0.112	0.033	0.028	0.008	0.002
AVDZ	19.148	0.167	0.080	0.026	0.005

Table 1.3: Summary of the EMBE2 results for the Gly-Gly-Gly tripeptide. All calculations use either the cc-pVDZ (VDZ) basis set or the aug-cc-pVDZ (AVDZ) basis set. Results are provided for several values of n_{t}, as well as for the supermolecular basis set (Super.). Both the mean unsigned MBE error over all values of Ω and the standard deviation of the MBE error are provided. All values are reported in kcal/mol.

(Fig. 1.10, right).[83,84] For the truncated embedding calculations, the atoms of side-chain moieties are only included in the set of border atoms if all backbone atoms to which the side-chain moieties are bonded are border atoms.

Table 1.4 presents the results of EMBE2 calculations for the Val-Pro-Leu and Tyr-Pro-Tyr tripeptides, as well as for the Gly-Gly-Gly-Gly tetrapeptide from Section 1.4.1. Due to the computational cost of the reference calculations, results employing the aug-cc-pVDZ basis set are not included for these more complex tripeptides. As with the Gly-Gly-Gly tripeptide calculations (Fig. 1.11 and Table 1.3), the results yield small (sub kcal/mol) errors that systematically decrease with the number of border atoms.

Peptide/Basis	$n_{\text{t}} = 1$	$n_{\text{t}} = 2$	$n_{\text{t}} = 3$	$n_{\text{t}} = 4$	Super.
VPL/VDZ	-0.029	-1.041	-0.413	-0.205	0.037
YPY/VDZ	0.604	-0.767	-0.092	-0.277	0.076
GGGG/VDZ	-0.175	-0.821	-0.751	-0.066	0.095
GGGG/AVDZ	26.439	-1.513	-1.118	-0.234	0.108

Table 1.4: Summary of the EMBE2 results for the Val-Pro-Lue tripeptide, the Tyr-Pro-Tyr tripeptide, and the Gly-Gly-Gly-Gly tetrapeptide. All calculations use either the cc-pVDZ (VDZ) basis set or the aug-cc-pVDZ (AVDZ) basis set. Results are provided for several values of n_{t}, as well as for the supermolecular basis set (Super.). All values are reported in kcal/mol.

1.5 Conclusions

In this paper, we have presented an extension of our projection-based embedding method to allow for truncation of the AO basis set for subsystem calculations. The truncation approach involves combining highly accurate projection-based embedding for nearby interactions with an approximate treatment of the NAKP between distant MOs. Application of this approach to both molecular clusters and polypeptides illustrates that the errors introduced by truncation of the AO basis set are both small and systematically controllable with respect to the extent of truncation. EMBE calculations on these systems yield accurate total and relative conformational energies, even when the monomers in the expansion are connected by covalent bonds. Furthermore, we have demonstrated that this approach offers a means of switching between accurate projection-based embedding and DFT embedding using approximate KE functionals, such that it both benefits from previous research on the development of approximate KE functionals and allows for systematic improvement upon those functionals in practical applications. These results establish that our method exhibits the essential elements necessary for efficient WFT-in-SCF embedding calculations on large molecular systems.

1.6 Appendix: Water Hexamer Conformations

Each conformation is numbered in the manner described in the main text. All geometries are reported in units of Ångströms.

Conf. 1

O	-0.01444	-1.40688	0.86698
H	-0.82663	-1.49669	0.34972
H	-0.12523	-1.87516	1.68772
O	0.00588	1.34004	0.84909
H	0.08181	0.39642	1.02379
H	0.79287	1.61382	0.36582
O	2.39229	-1.33294	-0.43486
H	1.59349	-1.47091	0.08523
H	2.24183	-1.68508	-1.30567
O	2.39226	1.39375	-0.42232
H	2.55075	0.44513	-0.49263
H	3.12911	1.78535	0.03323
O	-2.40639	-1.30367	-0.41807
H	-2.58996	-0.35940	-0.32660
H	-2.41112	-1.51135	-1.34633
O	-2.36824	1.39716	-0.44640
H	-1.47660	1.53891	-0.10056
H	-2.97126	1.91930	0.07093

Conf. 2

O	0.67982	1.66847	0.22469
H	1.54472	1.23557	0.26190
H	0.80199	2.60883	0.28673
O	-0.92115	-0.00176	1.65020
H	-0.44806	0.77331	1.33150
H	-1.76238	-0.05171	1.18135
O	0.63241	-1.64984	0.20904
H	0.08923	-1.22268	0.89198
H	0.53282	-2.59214	0.27953
O	-0.42478	0.02956	-1.62376
H	-0.09616	0.81284	-1.17004
H	0.00023	-0.73782	-1.22438
O	2.74478	-0.02765	-0.08537
H	2.20281	-0.78880	0.15326
H	2.94467	-0.09143	-1.01363
O	-2.76330	-0.07803	-0.30968
H	-2.08455	-0.14812	-0.99436
H	-3.30763	0.67610	-0.50478

Conf. 3

O	1.38240	-0.79588	1.21410
H	0.45622	-0.77606	1.50601
H	1.91210	-1.20090	1.89061
O	1.48104	1.37700	-0.30523
H	1.66217	0.81313	0.45827
H	2.14568	2.05442	-0.35742
O	1.06670	-1.01572	-1.39720
H	1.28587	-1.35829	-0.52354
H	1.27925	-0.07519	-1.39115
O	-1.27339	-0.50500	1.53311
H	-1.33645	0.44562	1.36259
H	-1.58576	-0.94802	0.73654
O	-1.20461	1.67808	0.04663
H	-0.27064	1.81552	-0.13712
H	-1.54482	1.05338	-0.60204
O	-1.58571	-0.71089	-1.07832
H	-0.70068	-0.95230	-1.38904
H	-2.23434	-1.09209	-1.65835

Conf. 4

O	18.0712	-51.4753	1.0577
H	18.3158	-52.4080	1.0702
H	18.9200	-50.9885	0.9316
O	20.4013	-50.1151	0.6787
H	20.3203	-49.3282	0.0893
H	20.8628	-49.7855	1.4593
O	20.1820	-47.9418	-0.9528
H	19.3727	-47.3927	-0.8299
H	20.2353	-48.1019	-1.9024
O	15.8359	-50.5064	-0.1649
H	15.7879	-50.8915	-1.0479
H	16.6469	-50.8938	0.2396
O	17.9552	-46.4504	-0.4572
H	17.7010	-45.6711	-0.9652
H	17.1102	-46.9274	-0.2792
O	15.6330	-47.7931	0.0179
H	15.7046	-48.7745	-0.0499
H	15.1884	-47.6406	0.8605

Conf. 5

O	26.0243	-46.8132	0.6983
H	25.9101	-46.6716	1.6460
H	26.0029	-47.7919	0.5870
O	26.0716	-49.5194	0.3763
H	26.7867	-49.9034	-0.1830
H	25.3029	-50.0816	0.2267
O	28.1035	-50.5775	-1.1078
H	29.0003	-50.2744	-0.8287
H	28.1161	-50.5173	-2.0704
O	28.2510	-45.6677	-0.3766
H	27.9994	-45.4294	-1.2768
H	27.4402	-46.0768	0.0100
O	30.5786	-49.7113	-0.3571
H	30.9359	-50.1242	0.4385
H	30.6024	-48.7444	-0.1688
O	30.5444	-47.0387	0.1802
H	29.7277	-46.5346	-0.0448
H	31.2721	-46.4282	0.0153

Conf. 6

O	-0.1141	-45.4756	0.4357
H	-0.1212	-45.0081	1.2795
H	-0.9900	-45.9275	0.3939
O	-2.6533	-49.4119	0.4391
H	-1.8471	-49.9163	0.1772
H	-2.9576	-49.8399	1.2483
O	2.1002	-46.9006	-0.2471
H	2.4044	-46.4744	-1.0573
H	1.2956	-46.3941	0.0156
O	-0.4370	-50.8340	-0.2436
H	0.4396	-50.3836	-0.2009
H	-0.4308	-51.2999	-1.0883
O	1.9963	-49.6175	-0.1859
H	2.6080	-49.8611	0.5191
H	2.0244	-48.6317	-0.2130
O	-2.5466	-46.6945	0.3830
H	-3.1568	-46.4524	-0.3238
H	-2.5749	-47.6802	0.4115

Conf. 7

O	10.1313	-39.6546	-1.2837
H	10.9834	-39.2940	-0.9427
H	10.3710	-40.2693	-1.9876
O	8.5206	-37.2415	-1.1214
H	8.9694	-38.0370	-1.4609
H	7.6674	-37.5876	-0.7864
O	10.0161	-37.0305	1.0891
H	9.4323	-36.9616	0.2835
H	9.9503	-36.1767	1.5337
O	8.7588	-39.6895	1.2947
H	9.2025	-39.9141	0.4565
H	9.1929	-38.8551	1.5467
O	12.1674	-38.3372	-0.0356
H	11.5483	-37.7652	0.4699
H	12.7194	-38.7610	0.6334
O	6.4748	-38.4556	0.3410
H	7.1523	-39.0355	0.7568
H	5.7121	-39.0216	0.1750

Conf. 8

O	15.9468	-39.5823	0.5548
H	15.3045	-40.0871	0.0416
H	16.8221	-39.7683	0.1280
O	18.3471	-39.7460	-0.6662
H	19.1305	-40.0034	-0.1359
H	18.5029	-38.8042	-0.8557
O	20.6467	-40.0714	0.9070
H	21.4549	-40.4532	0.5435
H	20.8337	-39.1110	0.9827
O	20.8368	-37.2970	0.9070
H	20.0333	-37.0136	0.4221
H	20.8440	-36.7677	1.7135
O	16.1147	-36.8858	0.5596
H	15.9307	-37.8581	0.5885
H	15.9523	-36.5744	1.4583
O	18.5450	-36.8241	-0.6030
H	18.5146	-36.2064	-1.3439
H	17.6621	-36.7483	-0.1590

Conf. 9

O	29.7954	-37.1884	-0.3910
H	30.6865	-36.9230	-0.1328
H	29.8886	-38.1244	-0.7014
O	29.8265	-39.8168	-1.0122
H	29.6466	-40.1441	-1.9019
H	29.0506	-40.1097	-0.4687
O	27.7083	-40.3069	0.6051
H	27.6403	-39.4409	1.0439
H	26.8649	-40.3743	0.1095
O	25.3507	-40.1183	-0.9006
H	24.5083	-40.4902	-0.6120
H	25.2501	-39.1476	-0.7875
O	25.3917	-37.3962	-0.3791
H	25.5108	-36.7584	-1.0937
H	26.1882	-37.2944	0.1837
O	27.6626	-37.4465	1.2384
H	27.7398	-36.9721	2.0750
H	28.4945	-37.2512	0.7373

Conf. 10

O	7.5344	-47.1952	0.4765
H	8.4160	-47.2312	0.0120
H	7.2660	-46.2693	0.4279
O	8.1630	-50.9274	0.5473
H	8.2441	-50.4060	1.3688
H	7.3554	-50.5591	0.1389
O	9.8927	-47.4450	-0.7542
H	10.6734	-47.2098	-0.2369
H	10.0133	-48.4102	-0.9487
O	10.1564	-50.1025	-1.0813
H	10.0613	-50.5472	-1.9320
H	9.4474	-50.4883	-0.5041
O	8.2958	-48.9198	2.5834
H	7.7974	-48.8686	3.4078
H	7.9420	-48.2014	2.0231
O	6.1151	-49.3715	-0.6444
H	6.4601	-48.5223	-0.3019
H	5.1679	-49.3538	-0.4625

Conf. 11

O	-1.9241	-39.9369	0.2244
H	-1.7339	-40.0619	-0.7214
H	-2.0308	-38.9602	0.2590
O	-1.6252	-37.1872	-0.1835
H	-1.3632	-37.4801	-1.0740
H	-0.7751	-37.0474	0.2718
O	-0.7346	-39.0353	-2.2858
H	-0.6724	-39.1225	-3.2446
H	0.1927	-39.1034	-1.9556
O	1.6853	-39.1868	-1.0050
H	1.7953	-38.3281	-0.5647
H	1.4640	-39.7654	-0.2520
O	0.4498	-40.0965	1.4209
H	-0.4725	-40.1381	1.0403
H	0.4880	-40.7832	2.0974
O	0.9394	-37.3782	1.1917
H	1.2480	-36.8472	1.9361
H	0.8015	-38.2813	1.5485

1.7 Appendix: Gly-Gly-Gly Tripeptide Conformations

The value of Ω is provided for each conformation. All geometries are reported in units of Ångströms.

$\Omega = 0^\circ$

N	-1.5267487178	0.0003318359	-4.9581655443
H	-1.1002439103	0.8044479587	-5.3878527168
H	-1.0766836580	-0.7843778331	-5.3995494210
C	-1.2013879770	-0.0055991494	-3.5553735666
H	-1.6567427374	0.8638552633	-3.0737322771
H	-1.6531761866	-0.8816382678	-3.0818132984
C	0.2933923286	-0.0070391498	-3.2527855893
O	1.1066714526	-0.0211875883	-4.1333563149
N	0.6588362176	0.0064590075	-1.9467914031
H	1.6425886714	0.0044649127	-1.7577423341
C	-0.2040106866	0.0221195162	-0.8023482048
H	-0.8392739455	0.9118360564	-0.7687288001
H	-0.8665049732	-0.8470868370	-0.7667798820
C	0.6393348427	0.0115623440	0.4636227850
O	1.8372146090	0.0017144527	0.4408726419
N	-0.0699934110	0.0151224710	1.6090423115
H	-1.0689271497	0.0226279731	1.6044777432
C	0.5787916540	0.0089080194	2.8912626172
H	1.2221420998	-0.8626612585	3.0175453187
H	1.2113573763	0.8862824928	3.0333087599
C	-0.4639103385	-0.0058759144	3.9753557149
O	-1.6357218524	-0.0058512157	3.7876339327
O	0.0837393243	-0.0177313263	5.1770809492
H	-0.6116830908	-0.0257222035	5.8265257531

$\Omega = 15^\circ$

N	0.0014367247	-1.5291038900	-4.9561492742
H	-0.7533584891	-1.0569089114	-5.4258220198
H	0.8326013201	-1.1256513950	-5.3557003956
C	-0.0472579142	-1.2027336778	-3.5541251055
H	-0.9570297247	-1.6197333500	-3.1138602241
H	0.7856605325	-1.6906300720	-3.0405338755
C	-0.0085968400	0.2909200200	-3.2504825524
O	0.0586204096	1.1034647395	-4.1283720161
N	-0.0961906635	0.6591750238	-1.9447505748
H	-0.0066444171	1.6412577362	-1.7656879432
C	0.0777444654	-0.1934941034	-0.8040184799
H	-0.6927915675	-0.9650291024	-0.7497530468
H	1.0469673607	-0.7041129690	-0.7970047451
C	0.0102021774	0.6439178266	0.4636889191
O	0.0136027100	1.8419098222	0.4443933648
N	-0.0332701547	-0.0684933585	1.6065737553
H	-0.0144163050	-1.0672269214	1.5998719971
C	-0.0330717843	0.5774179416	2.8902887157
H	0.8282985389	1.2348499392	3.0150479918
H	-0.9200922121	1.1952321972	3.0363954344
C	0.0022766767	-0.4668868235	3.9723033811
O	0.0298300068	-1.6380367807	3.7824222222
O	-0.0016684237	0.0784746066	5.1751851608
H	0.0209782076	-0.6179295411	5.8232200367

$\Omega = 30°$

N	0.0209323510	-1.5175297593	-4.9684376108
H	-0.6730326311	-0.9996915760	-5.4817089147
H	0.8990259183	-1.1615651720	-5.3082019512
C	-0.1028719023	-1.2008935668	-3.5687374666
H	-1.0614368730	-1.5675583610	-3.1919929155
H	0.6659941405	-1.7400760951	-3.0087784643
C	-0.0122405920	0.2852767036	-3.2430544339
O	0.1449741564	1.1037132045	-4.1017060483
N	-0.2100742540	0.6450870818	-1.9393995416
H	-0.0822626892	1.6256311631	-1.7685896767
C	0.1086137454	-0.1834269615	-0.8067754232
H	-0.5686554882	-1.0350175744	-0.7279321503
H	1.1294471523	-0.5814770995	-0.8405065756
C	0.0102297530	0.6462839965	0.4635132096
O	0.0046923503	1.8445301984	0.4466445009
N	-0.0357378622	-0.0684637904	1.6045235850
H	-0.0128110024	-1.0669836515	1.5959727437
C	-0.0378578377	0.5754296185	2.8892913665
H	0.8175988265	1.2408500651	3.0120784163
H	-0.9301788213	1.1847733575	3.0387675405
C	0.0099685222	-0.4697741378	3.9698810588
O	0.0455299286	-1.6405005381	3.7786866260
O	0.0059775411	0.0741579724	5.1734862164
H	0.0358939489	-0.6228957252	5.8205179729

$\Omega = 45°$

N	0.0359023616	-1.4897395191	-4.9990271227
H	-0.5146027370	-0.8736625433	-5.5736688105
H	0.9895003359	-1.2593024075	-5.2236450333
C	-0.2127726050	-1.1838727051	-3.6140516838
H	-1.2465610349	-1.4313490788	-3.3590782254
H	0.4120033028	-1.8248687228	-2.9866161067
C	0.0006243574	0.2757268100	-3.2323533320
O	0.2891833194	1.0993501613	-4.0484404194
N	-0.2993064597	0.6214487484	-1.9345557102
H	-0.1410230528	1.5999905634	-1.7691467863
C	0.1134572257	-0.1792856157	-0.8068327951
H	-0.4978900565	-1.0772788816	-0.7098314962
H	1.1600704394	-0.5008714239	-0.8687418281
C	-0.0078652321	0.6457010224	0.4643971705
O	-0.0504062669	1.8431740719	0.4475949825
N	-0.0231368621	-0.0700312882	1.6057079808
H	0.0256182578	-1.0675216496	1.5963217785
C	-0.0400061620	0.5726170423	2.8907681082
H	0.8020979252	1.2545121252	3.0165785097
H	-0.9442611969	1.1643976253	3.0386708264
C	0.0266022121	-0.4724825374	3.9705212469
O	0.0907176750	-1.6418615151	3.7784093626
O	0.0054734543	0.0697537930	5.1747662379
H	0.0493444510	-0.6273150346	5.8209735100

$\Omega = 60°$

N	-1.1248465263	-0.5058783252	-5.1033718830
H	-0.5868096283	0.1158314347	-5.6836814763
H	-0.7063613835	-1.4107165856	-5.2409789224
C	-0.9771890399	-0.1181746331	-3.7238086951
H	-1.4313383231	0.8644124677	-3.5630177852
H	-1.5465869446	-0.8066628902	-3.0967282580
C	0.4571610861	-0.0523349885	-3.2107878778
O	1.3706469549	-0.3402646528	-3.9065027770
N	0.6735420709	0.4026725389	-1.8743750217
H	0.3989472523	1.3682956856	-1.7857518351
C	0.0617474803	-0.3458276048	-0.7974220256
H	-0.9713320314	-0.6634674962	-0.9862764235
H	0.6433323499	-1.2450055522	-0.5898454617
C	-0.0039066614	0.5372061901	0.4368739478
O	-0.1213536829	1.7286581334	0.3506844363
N	0.0214456556	-0.1094347475	1.6167159408
H	0.1639287634	-1.0964951775	1.6653804345
C	-0.0981629559	0.5997709445	2.8604331763
H	0.7103391009	1.3189774217	3.0002878742
H	-1.0295776054	1.1640835019	2.9174680685
C	-0.0620435016	-0.3811388743	4.0000492062
O	0.0644583207	-1.5549808920	3.8808867256
O	-0.1861887407	0.2248506271	5.1676213998
H	-0.1526039377	-0.4331423724	5.8541055123

$\Omega = 75°$

N	-1.2270726420	0.1104897221	-5.0604360714
H	-0.4747107352	0.2824329904	-5.7063102378
H	-1.3743672498	-0.8843372393	-5.0967070805
C	-0.8112456152	0.4812243800	-3.7327203797
H	-0.6146836249	1.5568818380	-3.6940978909
H	-1.6338960526	0.3146410078	-3.0336483638
C	0.4115258608	-0.2529717177	-3.1961297607
O	0.9319481397	-1.1216010635	-3.8109072548
N	0.8554173758	0.0924748058	-1.8872760462
H	0.8499194741	1.0958404609	-1.7853764497
C	0.0418849878	-0.4266560996	-0.8013526686
H	-1.0331260934	-0.4490589797	-1.0203818609
H	0.3415550332	-1.4470905095	-0.5615734114
C	0.2000074751	0.4711939450	0.4134706443
O	0.4408761233	1.6421478144	0.3040069580
N	0.0008481888	-0.1242791767	1.6033149262
H	-0.1508158460	-1.1088616283	1.6725018846
C	0.0638826038	0.6193359106	2.8311761199
H	1.0399233401	1.0844861973	2.9766810278
H	-0.6713809482	1.4244182284	2.8566500784
C	-0.1988794959	-0.3047777982	3.9884084497
O	-0.4118252517	-1.4681160202	3.8924111064
O	-0.1614179081	0.3380071248	5.1421959617
H	-0.3281852826	-0.2860615046	5.8408368111

$\Omega = 90°$

N	-1.0624592857	0.2428260996	-5.1556855485
H	-0.2759022467	0.2707543501	-5.7829047428
H	-1.3501576726	-0.7215345659	-5.1455158446
C	-0.6349102635	0.6208467638	-3.8344065291
H	-0.2342584845	1.6382766751	-3.8509394304
H	-1.4988035510	0.6614276532	-3.1638788262
C	0.3726951998	-0.3095856773	-3.1742533964
O	0.6549550332	-1.3606603840	-3.6486725318
N	0.8320457985	0.0679998258	-1.8942558638
H	0.8226704401	1.0707929514	-1.8012942382
C	0.0380363064	-0.4490198083	-0.7931138495
H	-1.0382324460	-0.4744288242	-1.0097155572
H	0.3407902112	-1.4670823845	-0.5507056349
C	0.2038589167	0.4599889257	0.4121592053
O	0.4416084861	1.6306164140	0.2906114290
N	0.0122401825	-0.1250944268	1.6080637307
H	-0.1419059974	-1.1087002723	1.6859745794
C	0.0759377773	0.6295148751	2.8291047228
H	1.0553986659	1.0873165269	2.9757855353
H	-0.6520410094	1.4413449220	2.8428032413
C	-0.2005903691	-0.2814154569	3.9937023058
O	-0.4188999171	-1.4444874286	3.9077505632
O	-0.1688929257	0.3729441352	5.1411170960
H	-0.3440982288	-0.2431768661	5.8447418506

$\Omega = 105^{\circ}$

N	-0.2713337949	-0.8315994304	-5.2995694200
H	-0.1733398408	0.0063978360	-5.8482849268
H	0.6517210225	-1.2316601752	-5.2707486478
C	-0.6730502348	-0.4848621785	-3.9621793456
H	-1.6344009885	0.0353667984	-3.9846313161
H	-0.8462590270	-1.3956416444	-3.3795756171
C	0.3356357853	0.3226298054	-3.1582838219
O	1.4642175420	0.4454588177	-3.5159323199
N	-0.0851458706	0.7507196383	-1.8962868980
H	-1.0862434527	0.7120830374	-1.8085037387
C	0.4744501365	0.0237141488	-0.7716794993
H	0.5737399153	-1.0509698206	-0.9706009958
H	1.4675313533	0.3975093336	-0.5274824798
C	-0.4551812645	0.1569817190	0.4212018128
O	-1.6339409944	0.3433973296	0.2867081791
N	0.1252551995	0.0023082380	1.6247758842
H	1.1138724943	-0.1074973525	1.7142878961
C	-0.6447367950	0.0446208139	2.8371533726
H	-1.1542111350	1.0010364071	2.9636587875
H	-1.4178995612	-0.7243142263	2.8543896113
C	0.2665474634	-0.1665776262	4.0150304912
O	1.4400882786	-0.3262878624	3.9442823868
O	-0.4009304791	-0.1497731425	5.1552064905
H	0.2156132628	-0.2822968765	5.8677349896

$\Omega = 120^{\circ}$

N	-1.4139178753	-0.7894612062	-4.4093892816
H	-0.8085152203	-0.8070610243	-5.2131998774
H	-1.3367345287	-1.7088059709	-4.0073389300
C	-0.9278800462	0.1770755727	-3.4590479302
H	-0.9477852582	1.1764846557	-3.9027249976
H	-1.5892280245	0.2227353238	-2.5919873844
C	0.4537474935	-0.1090252200	-2.8915896658
O	0.9720430137	-1.1780158131	-3.0020652662
N	1.0188571576	0.8954427359	-2.1087311861
H	0.4418193472	1.7168396846	-2.1158212720
C	1.3403847403	0.5524775922	-0.7423964882
H	1.7895912921	-0.4386420775	-0.7238185760
H	2.0861760376	1.2487053560	-0.3525704837
C	0.1320689319	0.5941789934	0.1828501412
O	-0.9369737227	1.0107185748	-0.1745736895
N	0.3582169528	0.1555680165	1.4356955874
H	1.2541094418	-0.1905058888	1.7110198104
C	-0.6733718621	0.1760877206	2.4356647980
H	-1.0764737424	1.1789029703	2.5818222844
H	-1.5165949032	-0.4620270754	2.1670939463
C	-0.1115439057	-0.3078104406	3.7443731220
O	1.0100926577	-0.6556316302	3.9138572390
O	-1.0263283999	-0.3021646073	4.6979121802
H	-0.6263420175	-0.6174491342	5.5016339739

$\Omega = 135^\circ$

N	-1.2593196802	-0.6859529280	-4.6159301349
H	-0.5353936139	-0.8970303782	-5.2825267553
H	-1.4259327601	-1.5523909177	-4.1317923643
C	-0.7759558123	0.2882726910	-3.6722995717
H	-0.5374913041	1.2212347911	-4.1901661357
H	-1.5587296288	0.5299798830	-2.9494374947
C	0.4150857549	-0.1527082959	-2.8381386118
O	0.7555269563	-1.2957093132	-2.7663683863
N	1.0210068599	0.8414811651	-2.0911623981
H	0.5249648634	1.7103605993	-2.1520939679
C	1.3536379781	0.5286436784	-0.7239237795
H	1.8104147260	-0.4577606243	-0.6903534638
H	2.0888452012	1.2427287808	-0.3481107453
C	0.1360673179	0.5614504522	0.1903184867
O	-0.9452626086	0.9134212712	-0.1963963155
N	0.3679773324	0.1910659170	1.4646222740
H	1.2690936767	-0.1259238494	1.7570666787
C	-0.6827788913	0.1762646362	2.4440786593
H	-1.1290297737	1.1627077016	2.5758855327
H	-1.4941534992	-0.4970574748	2.1634245492
C	-0.1282305921	-0.2757761237	3.7669746547
O	1.0022327831	-0.5795023204	3.9612001612
O	-1.0605737103	-0.2970157975	4.7034972182
H	-0.6639765025	-0.5913568555	5.5167425123

$\Omega = 150^{\circ}$

N	-0.1916019688	-0.4805097381	-5.4894607427
H	0.2099977048	0.3712620256	-5.8445237378
H	0.5810558299	-1.1221459621	-5.4230991638
C	-0.7157323668	-0.2486075360	-4.1685415698
H	-1.4896121844	0.5215836678	-4.2044994523
H	-1.2082697723	-1.1562929707	-3.8066886878
C	0.3229985482	0.1153057927	-3.1169114401
O	1.4946436830	-0.0702880290	-3.2889091243
N	-0.1771436916	0.5812711122	-1.9387440575
H	-1.1708547542	0.5390510572	-1.8169501869
C	0.5470786082	0.3522553639	-0.7145765669
H	1.1629515528	-0.5461437872	-0.7882768749
H	1.2240072907	1.1776457628	-0.4852880689
C	-0.4307941269	0.1884450462	0.4343285277
O	-1.6204603785	0.1600996766	0.2692541503
N	0.1365620333	0.0692994696	1.6495159145
H	1.1278057279	0.1149052482	1.7660469567
C	-0.6584867410	-0.0958277062	2.8345791953
H	-1.3712343337	0.7197556932	2.9620487083
H	-1.2415721933	-1.0177484825	2.8116707001
C	0.2417141862	-0.1347683046	4.0389141424
O	1.4242874592	-0.0456862615	4.0050691364
O	-0.4479230918	-0.2816903341	5.1565838650
H	0.1618772178	-0.3009925293	5.8867509095

$\Omega = 165^{\circ}$

N	-0.0735534164	-0.1278046356	-5.5988222439
H	0.4236706332	0.7320188232	-5.7613739336
H	0.6389988673	-0.8386777196	-5.6012713788
C	-0.6833868256	-0.0852883855	-4.2948227894
H	-1.4023279706	0.7358128407	-4.2478887592
H	-1.2582507900	-1.0016531889	-4.1313844217
C	0.2942375283	0.0413808248	-3.1336408796
O	1.4786856915	-0.0926285533	-3.2750989445
N	-0.2695086927	0.2851224788	-1.9265476468
H	-1.2626778835	0.2283125801	-1.8200325640
C	0.4835415243	0.1113025760	-0.7151518171
H	1.0737036943	-0.8078018519	-0.7421966014
H	1.1876170253	0.9311507256	-0.5629123065
C	-0.4697237292	0.0443927335	0.4628902696
O	-1.6626878640	0.0090060821	0.3241503609
N	0.1213719472	0.0128306370	1.6719868169
H	1.1143929461	0.0698815684	1.7659740575
C	-0.6513324174	-0.0582912847	2.8809596955
H	-1.3468684809	0.7773886376	2.9691393603
H	-1.2509523869	-0.9682603254	2.9295930758
C	0.2730385253	-0.0343261148	4.0672398850
O	1.4554735606	0.0397959491	4.0045065145
O	-0.3948837087	-0.1051966506	5.2052977533
H	0.2294741971	-0.0855590712	5.9230416874

$\Omega = 180^\circ$

N	-0.0269163945	-0.0003669973	-5.6185831999
H	0.5714960797	0.8057293587	-5.6878423688
H	0.5988200817	-0.7865317202	-5.6747167940
C	-0.6691792140	-0.0004135318	-4.3299529134
H	-1.3205009335	0.8733349665	-4.2447530573
H	-1.3222321350	-0.8729010352	-4.2442458752
C	0.2829384544	-0.0001902358	-3.1399093201
O	1.4770616019	-0.0004021832	-3.2712393308
N	-0.3103442048	-0.0002321155	-1.9273323967
H	-1.3045236274	-0.0010712146	-1.8278776451
C	0.4550670986	-0.0000398520	-0.7143865871
H	1.1048562652	-0.8756293613	-0.6530794975
H	1.1038902078	0.8763750316	-0.6530191145
C	-0.4887630540	-0.0001434093	0.4726209523
O	-1.6834573950	-0.0002002073	0.3460622534
N	0.1133293086	0.0000202885	1.6770662164
H	1.1089865063	-0.0012817929	1.7600347228
C	-0.6489538670	-0.0000504657	2.8946046756
H	-1.2969697284	0.8741706656	2.9700780066
H	-1.2967377301	-0.8745234412	2.9702960791
C	0.2880282786	0.0001695503	4.0711491323
O	1.4720359809	0.0003698137	3.9956032092
O	-0.3705491948	0.0001130930	5.2168946799
H	0.2616940103	0.0002281862	5.9279667429

1.8 Appendix: Additional Polypeptides

All geometries are reported in units of Ångströms.

Gly-Gly-Gly-Gly Tetrapeptide

N	-0.1381852021	-0.0000206193	-7.4229345830
H	0.4563390224	0.8074964311	-7.5082101125
H	0.4893348729	-0.7838678993	-7.4916000070
C	-0.7473407180	-0.0003399168	-6.1180661359
H	-1.3981910256	0.8719997096	-6.0163972754
H	-1.3957295477	-0.8744949861	-6.0154608640
C	0.2353553970	0.0001663079	-4.9529754492
O	1.4256732424	-0.0021192756	-5.1165043438
N	-0.3264031636	-0.0001282097	-3.7261064747
H	-1.3177157713	0.0095065890	-3.6013922283
C	0.4676682616	0.0002801600	-2.5313017822
H	1.1182183902	-0.8755449804	-2.4857744828
H	1.1180917124	0.8763360238	-2.4848843545
C	-0.4488387558	-0.0001981941	-1.3224527580
O	-1.6471011600	-0.0005189907	-1.4246354130
N	0.1796692937	0.0001243735	-0.1344154200
H	1.1770761927	0.0033883588	-0.0658734854
C	-0.5481271058	-0.0002562393	1.1030768941
H	-1.1944982795	0.8758481133	1.1839170214
H	-1.1930052150	-0.8774824986	1.1837400785
C	0.4379490050	0.0002517706	2.2559248645
O	1.6264745862	0.0008683186	2.0824090304
N	-0.1197425862	-0.0000406320	3.4803033203
H	-1.1118074578	-0.0005551970	3.5997031660
C	0.6856388935	0.0003736132	4.6703533310
H	1.3350461651	0.8751932693	4.7226604783
H	1.3358064345	-0.8738923161	4.7226979669
C	-0.2107477804	-0.0000943094	5.8784540974
O	-1.3965963579	-0.0007089000	5.8418951048
O	0.4864131509	0.0002640236	7.0005323306
H	-0.1202150859	-0.0000524132	7.7336468398

Val-Pro-Leu Tripeptide

N	64.1711561477	51.2834939821	52.9481710861
H	63.7320072415	51.7175191771	52.1536519743
H	64.6877418786	50.5054884195	52.5735108191
C	63.1371800321	50.7994103346	53.8475283388
H	63.6158557317	50.1259911586	54.5568018214
C	62.5612365567	51.9763329622	54.6770381388
H	63.4468229045	52.4200469289	55.1398983837
C	61.8924285105	53.0639412441	53.8316696175
H	61.6276532193	53.9089203820	54.4719012868
H	62.5596220168	53.4409519362	53.0556359900
H	60.9812377548	52.7021930133	53.3549444293
C	61.6242926071	51.5075863394	55.7937706555
H	61.3179591731	52.3594287214	56.4048052189
H	60.7184698778	51.0424457974	55.3992262079
H	62.1200513360	50.7945050123	56.4580606602
C	62.0448267733	50.0531770783	53.0740160374
O	61.6530667914	50.4896632688	52.0169982844
N	61.5133560270	48.9101815530	53.5680038559
C	61.9356046371	48.1210148413	54.7232906025
H	63.0187691336	48.0982665004	54.8182796381
H	61.5173459003	48.5217767634	55.6477756705
C	60.4131289140	48.2919642517	52.8415990225
H	60.7523289672	47.9884763541	51.8496929701
C	60.0493727469	47.0713461744	53.7019342583
H	59.6592713753	46.2488076485	53.1036945599
H	59.2884463599	47.3585041235	54.4283653634
C	61.3571693072	46.7418263389	54.4190800325
H	61.2148336888	46.1422709708	55.3180189221
H	62.0293339658	46.1963052237	53.7519574905
C	59.2281932979	49.2474488847	52.7137298384
O	58.8358647348	49.8806994140	53.6530479693
C	56.4560216764	49.4515905527	50.4308778631
O	56.9303292667	48.6863933513	49.4482089284
H	56.1905593925	48.2827429041	49.0071074120
O	55.2960331109	49.5623572144	50.6450356294
N	58.6184573687	49.2368966792	51.5046834214
H	59.1372227131	48.8684751344	50.7356233343
C	57.5550448339	50.1708368989	51.1859753505
H	57.1284669577	50.4960667765	52.1260915394
C	58.0890061038	51.3721653285	50.3867310396
H	58.8834726998	51.8192130446	50.9888293947
H	58.5665740104	50.9913813207	49.4789068866
C	57.0609748446	52.4462174019	50.0000267789
H	56.2906611992	51.9815351080	49.3745239394
C	56.3624231649	53.0604607023	51.2164764095

H	55.6955286422	53.8681813347	50.9061012010
H	55.7570886647	52.3272769717	51.7504782041
H	57.0908036607	53.4790228005	51.9176207082
C	57.7432959664	53.5274637522	49.1575921312
H	57.0213842532	54.2769699303	48.8255881411
H	58.5185582885	54.0421388330	49.7326730213
H	58.2154625735	53.1016381617	48.2687405215

Tyr-Pro-Tyr Tripeptide

N	17.0223234808	32.3803608336	25.2474684599
H	16.2193162877	31.8016856835	25.0612213487
H	17.7672183478	31.7461875945	25.4807228741
C	16.7195342805	33.2347739709	26.3760646713
H	17.6550365211	33.6474852860	26.7485537407
C	15.8095718160	34.4028620265	25.9342845485
H	14.9879174160	33.9771192618	25.3533554394
H	15.3510969559	34.8577013550	26.8140547639
C	16.5222300547	35.4763789131	25.1392408405
C	16.9400231986	35.2648668776	23.8212724590
H	16.7592552550	34.3070901440	23.3570900261
C	17.5921842183	36.2491145917	23.1045393250
H	17.9107777105	36.0787795436	22.0857762410
C	17.8488761296	37.4855268150	23.6902675640
O	18.4902294782	38.4120168435	22.9447263763
H	18.6003552461	39.2107024873	23.4379052389
C	16.7832348254	36.7178584732	25.7006976345
H	16.4648017549	36.9275069240	26.7147019307
C	17.4414380063	37.7188052870	24.9926045923
H	17.6282894338	38.6785220023	25.4600551123
C	15.9984966250	32.4694287857	27.4900848634
O	15.2660154939	31.5477339834	27.2039393016
N	16.1586322973	32.8758420333	28.7676020320
C	17.1366883399	33.8626698317	29.2423214893
H	18.1524297547	33.5420934743	28.9979028605
H	16.9723173354	34.8383133522	28.7883452197
C	15.5483498675	32.1275875535	29.8771298414
H	15.5437903527	31.0674880646	29.6323378565
C	16.4535861851	32.4596155789	31.0619457925
H	17.3014090502	31.7703330521	31.0649257570
H	15.9319937731	32.3783417120	32.0109268925
C	16.9235965895	33.8810598573	30.7549811970
H	17.8313266560	34.1623152113	31.2884081218
H	16.1370416081	34.5874400720	31.0179387129
C	14.0993697162	32.5763396484	30.1107350553
O	13.7790687370	33.2466109112	31.0559736139
C	11.6908807953	34.0844536882	29.3184980311
O	12.5337196843	34.7319167785	28.5180301340
H	12.3729773198	35.6619085694	28.6350025966
O	10.8939468265	34.6493046433	29.9866490822
N	13.2169522231	32.1582505891	29.1829202146
H	13.5841215949	31.7204615065	28.3611810682
C	11.8275848557	32.5713872093	29.1998468315
H	11.4371984781	32.3324594622	28.2092780753
C	10.9888466504	31.8261461116	30.2554247547

H	11.4031772009	30.8196055456	30.3209553002
H	11.1386527217	32.2976212904	31.2239619088
C	9.5108755476	31.7238668329	29.9310382948
C	8.5669274204	32.5897488606	30.4845992273
H	8.8964514683	33.3767221162	31.1467487520
C	7.2195062764	32.4726982844	30.1937088507
H	6.4948117309	33.1463724175	30.6287609406
C	6.7780330938	31.4790453743	29.3280941183
O	5.4521769958	31.4063452716	29.0747515961
H	5.2719788831	30.6949465406	28.4793267217
C	9.0461971552	30.7344950679	29.0729165051
H	9.7451139773	30.0324323393	28.6330622876
C	7.6975168734	30.6072737796	28.7649993153
H	7.3665294280	29.8229796852	28.0941435991

Bibliography

[1] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).

[2] P. Sherwood, A. H. de Vries, S. J. Collins, S. P. Greatbanks, N. A. Burton, M. A. Vincent, and I. H. Hillier, Faraday Discuss. **106**, 79 (1997).

[3] J. L. Gao, P. Amara, C. Alhambra, and M. J. Field, J. Phys. Chem. A **102**, 4714 (1998).

[4] H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2007).

[5] H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).

[6] L. Hu, P. Söderhjelm, and U. Ryde, J. Chem. Theory Comput. **7**, 761 (2011).

[7] S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, THEOCHEM **461-462**, 1 (1999).

[8] F. Maseras and K. Morokuma, J. Comp. Chem. **16**, 1170 (1995).

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

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

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

[12] P. Arora, W Li, P. Piecuch, J. W. Evans, M. Albao, and M. S. Gordon, J. Phys. Chem. C **114**, 12649 (2010).

[13] S. R. Pruitt, M. A. Addicoat, M. A. Collins, and M. S. Gordon, Phys. Chem. Chem. Phys. **14**, 7752 (2012).

[14] K. R. Brorsen, N. Minezawa, F. Xu, T. L. Windus, and M. S. Gordon, J. Chem. Theory Comput. **8**, 5008 (2012).

[15] A. Gaenko, T. L. Windus, M. Sosonkina, and M. S. Gordon, J. Chem. Theory Comput. **9**, 222 (2013).

[16] N. Govind, Y. A. Yang, A. J. R. da Silva, and E. A. Carter, Chem. Phys. Lett. **295**, 129 (1998).

[17] N. Govind, Y. A. Wang, and E. A. Carter, Phys. Rev. B **110**, 7677 (1999).

[18] T. Klüner, N. Govind, Y. A. Yang, and E. A. Carter, J. Chem. Phys. **116**, 42 (2002).

[19] P. Huang and E. A. Carter, J. Chem. Phys. **125**, 084102 (2006).

[20] S. Sharifzadeh, P. Huang, and E. Carter, J. Phys. Chem. C **112**, 4649 (2008).

[21] A. S. P. Gomes, C. R. Jacob, and L. Visscher, Phys. Chem. Chem. Phys. **10**, 5353 (2008).

[22] T. A. Wesolowski, Phys. Rev. A **77**, 012504 (2008).

[23] Y. G. Khait and M. R. Hoffmann, J. Chem. Phys. **133**, 044107 (2010).

[24] C. Huang, M. Pavone, and E. A. Carter, J. Chem. Phys. **134**, 154110 (2011).

[25] C. Huang and E. A. Carter, J. Chem. Phys. **135**, 194104 (2011).

[26] S. Hofener, A. S. P. Gomes, and L. Visscher, J. Chem. Phys. **136**, 044104 (2012).

[27] O. Roncero, A. Zanchet, P. Villarreal, and A. Aguado, J. Chem. Phys. **131**, 234110 (2009).

[28] A. Severo Pereira Gomes and C. R. Jacob, Annu. Rep. Prog. Chem., Sect. C: Phys. Chem. **108**, 222 (2012).

[29] J. D. Goodpaster, T. A. Barnes, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **137**, 224113 (2012).

[30] F. R. Manby, M. Stella, J. D. Goodpaster, and T. F. Miller III, J. Chem. Theory Comput. **8**, 2564 (2012).

[31] J. D. Goodpaster, N. Ananth, F. R. Manby, and T. F. Miller III, J. Chem. Phys. **133**, 084103 (2010).

[32] J. D. Goodpaster, T. A. Barnes, and T. F. Miller III, J. Chem. Phys. **134**, 164108 (2011).

[33] S. Fux, C. R. Jacob, J. Neugebauer, L. Visscher, and M. Reiher, J. Chem. Phys. **132**, 164101 (2010).

[34] J. Nafziger, Q. Wu, and A. Wasserman, J. Chem. Phys. **135**, 234101 (2011).

[35] P. D. Dedíková, P. Neogrády, and M. Urban, J. Phys. Chem. A **115**, 2350 (2011).

[36] P. G. Lykos and R. G. Parr, J. Chem. Phys. **24**, 1166 (1956).

[37] J. C. Phillips and L. Kleinman, Phys. Rev. **116**, 287 (1959).

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

[39] R. A. Mata, H.-J. Werner, and M. Schütz, J. Chem. Phys. **128**, 144106 (2008).

[40] T. M. Henderson, J. Chem. Phys. **125**, 014105 (2006).

[41] A. A. Cantu and S. Huzinaga, J. Chem. Phys. **55**, 5543 (1971).

[42] B. Swerts, L. F. Chibotaru, R. Lindh, L. Seijo, Z. Barandiaran, S. Clima, K. Pierloot, and M. F. A. Hendrickx, J. Chem. Theory Comput. **4**, 586 (2008).

[43] J. L. Pascual, N. Barros, Z. Barandiaran, and L. Seijo, J. Phys. Chem. A **113**, 12454 (2009).

[44] H.-J Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Shütz *et al.*, MOLPRO, version 2008.3, a package of *ab initio* programs, 2008, see www.molpro.net.

[45] B. Temelso, K. A. Archer, and G. C. Shields, J. Phys. Chem. A **115**, 12034 (2011).

[46] T. H. Dunning, Jr., J. Chem. Phys. **90**, 1007 (1989).

[47] R. A. Kendall, T. H. Dunning, and R. J. Harrison, J. Chem. Phys. **96**, 6796 (1992).

[48] J. Pipek and P. Mezey, J. Chem. Phys. **90**, 4916 (1989).

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

[50] E. V. Stefanovich and T. N. Truong, J. Chem. Phys. **104**, 2946 (1996).

[51] A. Laio, J. VandeVondele, and U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).

[52] K. Senthilkumar, J. I. Mujika, K. E. Ranaghan, F. R. Manby, A. J. Mulholland, and J. N. Harvey, J. R. Soc. Interface **5**, S207 (2008).

[53] L. H. Thomas, Proc. Cambridge Philos. Soc. **23**, 542 (1927).

[54] E. Fermi, Z. Phys. **48**, 73 (1928).

[55] C. R. Jacob, T. A. Wesolowski, and L. Visscher, J. Chem. Phys. **123**, 174104 (2005).

[56] M. Dulak and T. A. Wesolowski, J. Chem. Phys. **124**, 164101 (2006).

[57] C. R. Jacob, S. M. Beyhan, and L. Visscher, J. Chem. Phys. **126**, 234116 (2007).

[58] E. Dahlke and D. Truhlar, J. Phys. Chem. B **110**, 10595 (2006).

[59] S. Hirata, O. Sode, M. Keçeli, and T. Shimazaki, *Accurate Condensed Phase Quantum Chemistry* (Taylor and Francis, 2011).

[60] P. J. Bygrave, N. L. Allan, and F. R. Manby, J. Chem. Phys. **137**, 164102 (2012).

[61] C. R. Taylor, P. J. Bygrave, J. N. Hart, N. L. Allan, and F. R. Manby, Phys. Chem. Chem. Phys. **14**, 7739 (2012).

[62] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 46 (2007).

[63] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **3**, 1342 (2007).

[64] H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2007).

[65] E. E. Dahlke and D. G. Truhlar, J. Chem. Theory Comput. **4**, 1 (2008).

[66] A. Sorkin, E. E. Dahlke, and D. G. Truhlar, J. Chem. Theory Comput. **4**, 683 (2008).

[67] E. E. Dahlke, H. R. Leverentz, and D. G. Truhlar, J. Chem. Theory Comput. **4**, 33 (2008).

[68] L. D. Jacobson and J. M. Herbert, J. Chem. Phys. **134**, 094118 (2011).

[69] S. Hirata, Chem. Phys. Phys. Chem. **11**, 8397 (2009).

[70] R. M. Richard and J. M. Herbert, J. Chem. Phys. **137**, 064113 (2012).

[71] G. J. O. Beran, J. Chem. Phys. **130**, 164115 (2009).

[72] S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran, Phys. Chem. Chem. Phys. **14**, 7578 (2012).

[73] E. B. Kadossov, K. J. Gaskell, and M. A. Langell, J. Comput. Chem. **28**, 1240 (2007).

[74] U. C. Deev and P. A. Kollman, J. Comput. Chem. **7**, 718 (1986).

[75] M. J. Field, P. A. Bash, and M. Karplus, J. Comput. Chem. **11**, 700 (1990).

[76] V. Deev and M. A. Collins, J. Chem. Phys. **122**, 154102 (2005).

[77] M. A. Collins and V. A. Deev, J. Chem. Phys. **125**, 104104 (2006).

[78] D. W. Zhang, X. H. Chen, and J. Z. H. Zhang, J. Comput. Chem. **24**, 1846 (2003).

[79] D. W. Zhang and J. Z. H. Zhang, J. Chem. Phys. **119**, 3599 (2003).

[80] R. Ditchfield, W. J. Hehre, and J. A. Pople, J. Chem. Phys. **54**, 724 (1971).

[81] W. J. Hehre, R. Ditchfield, and J. A. Pople, J. Chem. Phys. **56**, 2257 (1972).

[82] M. D. Tissandier, S. J. Singer, and J. V. Coe, J. Phys. Chem. A **104**, 752 (2000).

[83] S. C. Graham and J. M. Guss, Arch. Biochem. Biophys. **469**, 200 (2008).

[84] Z. Zhang, M. Qian, Q. Huang, Y. Jia, Y. Tang, K. Wang, D. Cui, and M. Li, J. Protein Chem. **20**, 59 (2001).

Chapter 2

Ab initio characterization of the electrochemical stability and solvation properties of condensed-phase ethylene carbonate and dimethyl carbonate mixtures

2.1 Introduction

The commercial market for Li-ion batteries is currently dominated by cells that utilize carbonate-based electrolytes and have cathodes that operate at potentials in the range of 3.5-4.2 V vs. Li.[1–6] Higher operational voltages would enable Li-ion batteries with greater energy density. However, typical carbonate solvents in these batteries undergo oxidation by the charged cathode surface at voltages exceeding 4.5 V vs. Li.[5] Following oxidation, the solvent molecules further decompose via pathways that have been implicated in capacity fade upon cycling.[2,5,7–12]

These issues have motivated both the investigation of additives that expand the window of electrochemical stability of carbonate-based solvents,[5,13–15] as well as the exploration of classes of solvents with improved stability against oxidation. Among the latter are sulfone-based electrolytes, which have been shown to exhibit improved

stability on high voltage cathodes.[10,16,17] Similarly, ionic liquids have attracted considerable interest due to their high electrochemical stability.[18–20] Unfortunately, these alternatives to traditional carbonate-based solvents typically suffer from other problems, such as high viscosity or the inability to form a robust solid electolyte interphase (SEI).[17]

A central obstacle in the pursuit of expanded windows of electrochemical stability is elucidation of the mechanisms by which each electrolyte component influences the oxidation potential of the other components. For example, the oxidation potentials of carbonate solvents are known to be influenced by the presence of co-solvents[2] and solvated ions,[21–23] as well as by the electrode composition.[24,25] This makes experimental investigation of oxidation potentials challenging, since all of these factors must be considered.

To complement these experimental efforts, theoretical studies have attempted to identify the intrinsic oxidation potential of the individual components of battery electrolytes. The intrinsic oxidation potential corresponds to the oxidation potential of each electrolyte component in the absence of an electrode and any oxidation-inducing decomposition reactions. The majority of these theoretical studies have been conducted using DFT methods[10,11,19–21,23,26–40] or WFT methods[10,27,41–43] applied in either the gas phase[26,27,35] or using implicit solvation models.[10,11,27,31–39,41,42] Only a few studies on electrolytes have utilized an explicit representation of the environment; these include QM/MM calculations on large clusters,[20] and condensed-phase calculations using DFT with periodic boundary conditions.[19,23,30,40]

In this paper, we perform CCSD(T)-level[44] calculations to determine the vertical IE of solvent molecules in condensed-phase mixtures of EC and DMC. This level of accuracy is achieved using the projection-based WFT-in-DFT embedding method,[45,46] which facilitates explicit treatment of the solvent environment at the DFT level. Using over 3,000 WFT-in-DFT embedding calculations on condensed-

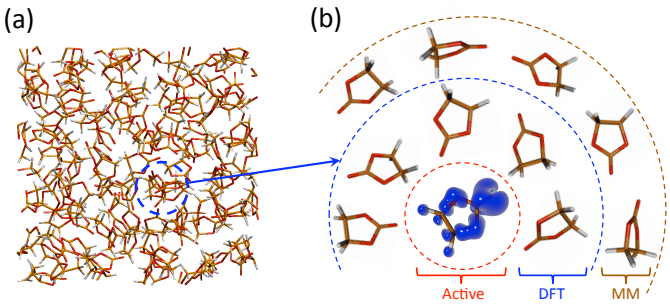

Figure 2.1: Summary of the embedding protocol. (a) MD simulations are performed to generate the equilibrium ensemble of solvent configurations. (b) An embedded CCSD(T) calculation is performed on a single molecule from the MD simulation (the "active region"), indicated by the red circle. The electron hole created upon oxidation of the active region is illustrated by the blue electron cloud. Nearby molecules are treated at the B3LYP level,[47,48] indicated by the blue circle. More distant molecules are treated using a point-charge molecular-mechanics (MM) model, indicated by the brown circle.

phase systems, we examine the effect of intermolecular interactions on the vertical IE of individual solvent molecules, and find that conventional implicit solvent models neglect interactions that are necessary to understand the electrochemistry and solvation properties of DMC. These observations enable a simple and intuitive explanation for experimentally-observed anomalies in the solvation structure of ions in carbonate-based electrolytes.

2.2 WFT-in-DFT Embedding for Condensed-Phase Systems

2.2.1 Overview of Embedding Strategy

Fig. 2.1 illustrates the overall embedding strategy employed for calculation of the oxidation potentials of EC and DMC mixtures, which consists of three main steps. The following is a brief overview of each step:

In step **(1)**, classical MD simulations are run to generate the ensemble of liquid-

phase configurations of the EC and DMC systems. Configurations are taken from these equilibrium simulations, with Fig. 2.1(a) showing a typical configuration of an MD simulation of 128 EC molecules. A more detailed description of these simulations is provided in Section 2.2.2.

In step **(2)**, WFT-in-DFT-in-MM embedding calculations are performed to calculate the vertical IE of a single molecule in each MD configuration. The oxidized molecule is treated at the WFT level, with the neighboring molecules being treated at the DFT level and more distant molecules being treated using an MM point-charge model. Fig. 2.1(b) illustrates the regions associated with each of these levels of theory, which we refer to as the active region, the DFT region, and the MM region, respectively. This step is described in more detail in Section 2.2.3.

In step **(3)**, we perform thermodynamic averages associated with the solvent reorganization energies and free energies of oxidation of the EC and DMC mixtures, using the ensemble of vertical IEs obtained in the preceding steps. From this ensemble, it is possible to confirm that the distribution of vertical IEs exhibit Gaussian fluctuations consistent with the predictions of linear response theory. Then, using this property,[49–53] the solvent reorganization energies can be calculated according to

$$\lambda = \frac{1}{2}\left(\langle \Delta E \rangle_0 - \langle \Delta E \rangle_{+1}\right), \tag{2.1}$$

where $\langle \Delta E \rangle_0$ is the vertical IE of a given solvent species, averaged over an ensemble of MD configurations of a given solvent mixture, and $\langle \Delta E \rangle_{+1}$ is the vertical IE of the same species, averaged over an ensemble of MD configurations in which the force field representation of the molecule in the active region has a net +1 charge. Similarly, the Helmholtz free energy of oxidation, which accounts for the effects of solvent

configurational relaxation, is

$$\Delta A = \frac{1}{2}\left(\langle \Delta E \rangle_0 + \langle \Delta E \rangle_{+1}\right), \tag{2.2}$$

from which the oxidation potential vs. Li^+/Li can be obtained as

$$\mathscr{E} = \frac{\Delta A}{F} - \mathscr{E}_{\text{ref}}, \tag{2.3}$$

where F is Faraday's constant and $\mathscr{E}_{\text{ref}}$=1.4 V is the reference oxidation potential of the Li^+/Li electrode.

2.2.2 MD Configurational Sampling

As described in step (1) of Section 2.2.1, the molecular configurations used in the vertical IE calculations are sampled using classical MD trajectories. MD simulations are performed for seven condensed-phase molecular systems, each having different combinations of EC and DMC molecules and their cations. Table 2.1 shows the molecular composition of each system.

Systems	EC	DMC	EC^+	DMC^+	Cell Volume / $Å^3$
EC	128	0	0	0	14384
EC^+	127	0	1	0	14384
DMC	0	128	0	0	19141
DMC^+	0	127	0	1	19141
EC:DMC	64	64	0	0	16503
EC^+:DMC	63	64	1	0	16503
EC:DMC^+	64	63	0	1	16503

Table 2.1: Number of Molecules in Each Simulated System, and the Volume of the Simulation Cell.

The MD trajectories are performed using the Lucretius^{TM} Molecular Dynamics simulation package and using previously published polarizable force-fields for neutral EC and DMC.[54,55] The bond lengths are constrained using the SHAKE algorithm,[56]

with a tolerance of 10^{-14} The oxidized solvent molecules, EC^+ and DMC^+, are treated using the same force field as the neutral molecules, except using the atomic point charges in Table 2.2; these charges are obtained by fitting to results generated at the MP2 level of theory[57] and using the aug-cc-pVTZ basis set,[58,59] in a manner similar to that by which the neutral charges were obtained.[54,55]

Atom	EC^+	DMC^+
C′	0.4943	0.6915
C	0.3150	0.0840
O′	-0.3277	-0.5323
O	0.3409	0.3113
H	0.0576	0.1169
Lp	-0.1772	-0.1628

Table 2.2: Point Charges for Each Atom in EC^+ and DMC^+ in the MD Simulations.

Values of the point charges of each atom in EC^+ and DMC^+ in MD simulations, with C′ and O′ representing the carbonyl carbon and carbonyl oxygen atom types, respectively. Two dummy atoms are attached to each of the sp^3 oxygen atoms, representing the lone-pair orbitals (Lp).

To generate liquid-phase configurations, a total of seven MD simulations are run in cubic simulation cells with periodic boundary conditions in the NVT ensemble. The Nosé-Hoover thermostat[60] is used to maintain a temperature of 313.0 K, using an associated damping frequency of 10^{13} s^{-1}. To determine the simulation volumes for the NVT trajectories, a total of three NPT simulations are performed on the charge-neutral EC, DMC, and EC:DMC systems. The NPT simulations are run for 3.0 ns using the Anderson-Hoover barostat[61] to maintain a pressure of 1.0 atm, using an associated damping frequency of 5×10^{11} s^{-1}. The NVT simulations on systems that include an oxidized molecule use the cell volumes of the corresponding neutral systems. The volume of each system is given in Table 2.1.

The MD numerical integration is performed using a multiple time-step integration scheme.[61] Following Ref. 55, the short time-step for bonded interactions is set to 0.5 fs, the central time-step for nonbonded interactions within 7.0 Å is set to 1.5 fs, and the

outer time-step for nonbonded interactions beyond 7.0 Å is set to 3.0 fs. Interactions between induced dipoles are self-consistently tapered to zero[62] beginning at 10.5 Å and ending at 11.0 Å, with a convergence threshold of 10^{-7} $Å^2e^2$. As in previously published MD simulations,[55] we employ Thole screening to smear induced dipoles and prevent over-polarization,[63] and the interaction between an induced dipole and a partial charge separated by three bonds is scaled by 0.8. The charge-charge and charge-induced dipole interactions are calculated using the Ewald summation method, with a short-range cutoff of 11.0 Å, $\alpha = 0.2$, and 5^3 k-vectors.

It has previously been noted that because of the approximately 10 kcal/mol barrier to cis-trans rotation in DMC, it is difficult to equilibrate DMC configurations, particularly when an ion is present.[55] To address this issue, we equilibrate systems that contain both DMC and an oxidized molecule in four phases, with the first phase corresponding to a 9.1 ns trajectory in which the MD force field is modified to reduce the barrier to conformational rotation of DMC to 3.0 kcal/mol, the second phase corresponding to an additional 9.1 ns trajectory in which the barrier is set to 5.0 kcal/mol, the third phase corresponding to an additional 9.1 ns trajectory in which the barrier is set to 7.0 kcal/mol, and the fourth phase corresponding to an additional 9.1 ns trajectory using the unmodified force field. All other systems are equilibrated for 1.0 ns, without any modifications to the MD force fields. Following equilibration, a total of 151 configurations are obtained from each simulation for use in subsequent embedding calculations, with 50 ps intervals between each configuration.

2.2.3 Projection-Based Embedding

As described in Section 2.2.1, the vertical IEs of EC and DMC molecules are calculated using the projection-based WFT-in-DFT embedding method,[45,46] which allows for a WFT-level treatment of the electrons in the active region and a DFT-level treatment of the electrons in surrounding molecules (see Fig. 2.1). This method describes

all subsystem interactions at the level of DFT,[45] while avoiding the need for an optimized effective potential step,[64–66] and without requiring the introduction of link-atoms or similar approximations that are typical of other widely used embedding approaches.[67–77] It has been previously established that this method can be accurately and efficiently applied to large systems via systematic truncation of the atomic orbital (AO) basis set representation of the electrons treated at the WFT level,[46] and it has been shown that errors associated with the DFT-level treatment of interactions between electrons treated at different levels of theory can be systematically analyzed and controlled.[78] These features have enabled successful application of the method to a variety of systems, including water clusters,[45,46] polypeptides,[46] the components of several prototypical organic reactions,[78] and organometallic hydrogen evolution catalysts.[79]

The projection-based embedding procedure is implemented as follows: first, a DFT calculation is performed over the full system, yielding a set of occupied MOs, $\{\phi_i\}$. The set of atoms in the active region is used to partition the MOs into one set corresponding to subsystem A, $\{\phi_i\}_\mathrm{A}$, and another set corresponding to subsystem B, $\{\phi_i\}_\mathrm{B}$. If the atom upon which an MO in $\{\phi_i\}$ has the largest Mulliken population is in the active region, that MO is assigned to $\{\phi_i\}_\mathrm{A}$; otherwise, it is assigned to $\{\phi_i\}_\mathrm{B}$. The subsystem MOs $\{\phi_i\}_\mathrm{A}$ and $\{\phi_i\}_\mathrm{B}$ are then used to construct the corresponding density matrices, γ^A and γ^B.

The embedded electronic-structure calculation on subsystem A is performed using a modified Fock matrix, given by

$$\mathbf{f}^\mathrm{A} = \mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^\mathrm{A}, \gamma^\mathrm{B}] + \mathbf{g}[\gamma^\mathrm{A}_\mathrm{emb}], \tag{2.4}$$

where $\gamma^\mathrm{A}_\mathrm{emb}$ is the density matrix associated with the MO eigenfunctions of $\mathbf{f}^\mathrm{A}$, and $\mathbf{g}$ consists of the two-electron contributions to the Fock matrix. The embedded core

Hamiltonian is defined as

$$\mathbf{h}^{\mathrm{A\ in\ B}}[\gamma^{\mathrm{A}}, \gamma^{\mathrm{B}}] = \mathbf{h} + \mathbf{g}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}] - \mathbf{g}[\gamma^{\mathrm{A}}] + \mu \mathbf{P}^{\mathrm{B}}, \tag{2.5}$$

where $\mathbf{h}$ is the standard one-electron core Hamiltonian, and μ is a level-shift parameter. Elements of the projection operator $\mathbf{P}^{\mathrm{B}}$ are given by

$$P^{\mathrm{B}}_{\alpha\beta} \equiv \langle b_\alpha | \left\{ \sum_{i \in \mathrm{B}} |\phi_i\rangle\langle\phi_i| \right\} |b_\beta\rangle, \tag{2.6}$$

where the b_α are the AO basis functions and the summation spans the MOs in $\{\phi_i\}_{\mathrm{B}}$. In the limit of $\mu \to \infty$, the MOs of subsystem A are enforced to be mutually orthogonal with those of subsystem B.[80–87] For the case of DFT-in-DFT embedding, a DFT calculation on subsystem A, using the Fock matrix in Eq. 2.4, is iterated to self-consistency with respect to $\gamma^{\mathrm{A}}_{\mathrm{emb}}$. For the case of WFT-in-DFT embedding, a WFT calculation is performed on the electrons in subsystem A, replacing the standard one-electron core Hamiltonian with the embedding core Hamiltonian in Eq. 2.5. A more complete description of WFT-in-DFT embedding is provided in Ref. 45.

Truncation of the AO basis set representation of subsystem A significantly reduces the cost of the embedded WFT calculation.[46] For this reason, we define a set of border atoms that correspond to the atoms within a distance r_{border} of any of the atoms in the active region. The calculation on subsystem A is performed using only the basis functions that are centered on either an atom in the active region or a border atom. A given MO in subsystem B is included in the sum in Eq. 2.6 only if it sufficiently occupies the basis functions centered on the border atoms. Specifically, for this MO, only if the absolute value of the combined Mulliken population associated with this set of AO basis functions exceeds a threshold τ is it included in the sum; otherwise, interactions between that MO and the MOs in subsystem A are treated

using the approximate Thomas-Fermi (TF) functional.[88,89] For the case of $r_{\text{border}} = 0$, no MO projection is employed, and the embedding potential evaluation reduces to that of orbital-free DFT embedding,[66,90,91] with the non-additive kinetic energy computed using the TF functional. The energy of a WFT-in-DFT truncated embedding calculation is

$$\begin{aligned} E_{\text{WFT}}[\Psi^{\text{A}};\gamma^{\text{A}},\gamma^{\text{B}}] = \langle\Psi^{\text{A}}|\hat{H}^{\text{A in B}}[\gamma^{\text{A}},\gamma^{\text{B}}]|\Psi^{\text{A}}\rangle \\ + E_{\text{DFT}}[\gamma^{\text{A}}+\gamma^{\text{B}}] - E_{\text{DFT}}[\gamma^{\text{A}}] \\ - \text{tr}\left[\gamma^{\text{A}}(\mathbf{h}^{\text{A in B}}[\gamma^{\text{A}},\gamma^{\text{B}}] - \mathbf{h})\right] \\ + T_{\text{s}}[\gamma^{\text{A}}+\gamma^{\text{C}}] - T_{\text{s}}[\gamma^{\text{A}}] \\ - T_{\text{s}}[\gamma^{\text{C}}], \end{aligned} \tag{2.7}$$

where $|\Psi^{\text{A}}\rangle$ is the embedded wavefunction from the WFT-in-DFT embedding calculation, $\hat{H}^{\text{A in B}}[\gamma^{\text{A}},\gamma^{\text{B}}]$ is the Hamiltonian resulting from replacing the standard core Hamiltonian with the embedded core Hamiltonian, $E_{\text{DFT}}[\gamma]$ is the DFT-level energy, $T_{\text{s}}[\gamma]$ is the TF energy, and γ^{C} is the density of the MOs in subsystem B that are not included in the projector operator. With this energy expression, interactions within subsystem A are treated at the level of WFT, interactions within subsystem B are treated at the level of DFT, and all interactions between subsystems are treated at the level of DFT. A full discussion of AO truncation is provided in Ref. 46.

As indicated by the MM region in Section 2.2.1, some of the reported embedding calculations utilize a point-charge representation for solvent molecules that are distant from the active region; we refer to such calculations as WFT-in-DFT-in-MM embedding calculations. These calculations are performed in the same manner as the WFT-in-DFT calculations, except that the set of DFT MOs, $\{\phi_i\}$, is obtained from a DFT calculation that is embedded in the potential produced by the point charges (i.e., B3LYP-in-MM embedding), in the same manner as Ref. 92,93. The

point-charge potential is added to the Hamiltonian in Eq. 2.5, and the DFT-level energy $E_{\mathrm{DFT}}[\gamma^{\mathrm{A}} + \gamma^{\mathrm{B}}]$ in Eq. 2.7 includes contributions from electrostatic interactions with the point charges.

2.2.3.1 Embedding Calculation Details

In order to obtain the vertical IEs in step (2) of Section 2.2.1, two sets of embedding calculations are performed on each molecular configuration: one for which the active region has a charge of 0 (the reduced state), and one for which it has a charge of +1 (the oxidized state). All of the embedding calculations are performed at either the B3LYP-in-B3LYP-in-MM, CCSD-in-B3LYP-in-MM, or CCSD(T)-in-B3LYP-in-MM levels of theory, using a basis set truncation parameter $\tau = 0.1$ and a level shift parameter $\mu = 10^6$. Any molecule having an atom within a distance r_{DFT} of any atom in the active region is treated at the DFT level, as indicated by the blue circle in Fig. 2.1. Any other molecule having an atom within a distance r_{MM} of any atom in the active region is treated using a set of point charges, as indicated by the brown circle in Fig. 2.1; in the determination of this cutoff, dummy atoms from the MD force field are counted as atoms. All embedding calculations are performed using the Molpro software package.[94]

Except where noted in Section 2.3.1, all DFT and WFT calculations are performed using restricted open-shell versions of the methods, and all WFT calculations use reference MOs obtained from the corresponding B3LYP calculations. For example, the CCSD(T) reference MOs in the CCSD(T)-in-B3LYP-in-MM embedding calculations are obtained from the corresponding B3LYP-in-B3LYP-in-MM embedding calculations. Although it is more common to use RHF reference MOs for coupled-cluster calculations, it is not necessarily optimal to do so,[95,96] and our choice of reference MOs is found to improve the convergence of the calculations. Correct convergence of the B3LYP-level calculations is found to be non-trivial, particularly for calculations

in the oxidized state. To address this issue, all embedded B3LYP calculations are performed in several steps: first, a maximum of 20 iterations of the B3LYP calculation are performed using level shifts of -20.0 E_h and -10.0 E_h for the alpha-spin and beta-spin orbitals, respectively, followed by a maximum of 50 iterations using level shifts of -2.0 E_h and -1.0 E_h, followed by a maximum of 50 iterations using level shifts of -0.5 E_h and -0.25 E_h, followed by iteration until convergence using no level shifts. To improve the convergence of the B3LYP-in-B3LYP-in-MM calculations on the active region in the oxidized state, these calculations are initialized using the MOs for the active region obtained from the corresponding B3LYP-in-B3LYP-in-MM calculations in the reduced state. The basis set used for the active region is specified for each calculation (e.g., CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ indicates that the aug-cc-pVTZ basis set is used for the atoms in the active region), while all calculations employ the cc-pVDZ basis set for atoms in the DFT region. The set of MOs $\{\phi_i\}$ is obtained by performing a B3LYP-in-MM calculation, in which a B3LYP calculation on the union of the active region and the DFT region is embedded in the PC region; separate localization of the resulting core and valence MOs is then performed using the Boys method.[97] To avoid the tendency of DFT to overdelocalize electron holes (see Section 2.3.1), each embedding calculation that is performed in the oxidized state employs the B3LYP-in-MM MOs, $\{\phi_i\}$, associated with the same molecular configuration in the reduced state. Thus, when an embedding calculation is performed on subsystem A in the oxidized state, the electronic distribution of subsystem B corresponds to that of the reduced state. In Section 2.2.3.2, we describe an iterative procedure to account for electronic relaxation of the environment in the oxidized system.

2.2.3.2 Electronic Relaxation of the DFT Region with Respect to Oxidation of the Active Region

To account for electronic relaxation of the molecules in the DFT region in WFT-in-DFT-in-MM embedding calculations performed on the oxidized state of the solvent molecules, as mentioned in Section 2.2.3.1, an iterative protocol is employed. We now describe in detail the steps involved in this iterative protocol.

In the first step, monomolecular ($r_{\text{border}} = 0.0$ Å) density matrices are acquired for both subsystem A and subsystem B. This is accomplished by performing a monomolecular B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding calculation on the active region to acquire a monomolecular density matrix for subsystem A. Following this, a monomolecular B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding calculation is performed on the DFT region (i.e., the roles of the active region and the DFT region are swapped), embedded in the potential produced by the monomolecular density matrix of subsystem A.

In the second step, the active region is oxidized. A monomolecular B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding calculation is performed on the active region, embedded in the potential produced by the monomolecular density matrix for subsystem B from the first step. This embedding calculation is performed with a +1 charge for the active region, yielding the electronically unrelaxed energy of the oxidized state.

In the third step, the electronic density of each subsystem is iteratively permitted to relax with respect to the electronic density of the other subsystem. Specifically, a monomolecular B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding calculation is performed on the DFT region, embedded in the potential produced by the density matrix from the preceding embedding calculation on the active region. Then, a monomolecular B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding calculation is performed on the active region, embedded in the potential produced by the density matrix from the

preceding embedding calculation on the DFT region; the calculation on the active region is performed with a +1 charge. This third step of the electronic relaxation protocol is iterated until convergence is achieved. Upon convergence, the energy of the final embedding calculation on the active region corresponds to the energy of the oxidized state.

All of the above steps are performed using the same values of r_{DFT} and r_{MM} as step (2) of Section 2.2.1, but r_{border} is set to zero in order to ensure that charge does not leak between subsystems. The relaxation energy is simply the difference between the relaxed and unrelaxed energies of the oxidized state. The relaxation energy is then added as a correction to the WFT-in-DFT embedding energies obtained using the protocol in Section 2.2.3.1, resulting in an electronically-relaxed WFT-in-DFT embedding energy. In Fig. 2.12 of Sec. 2.5, benchmarking calculations are presented that both illustrate the accuracy of this protocol and demonstrate that a single iteration of the third step in the above protocol is sufficient to converge the calculations; as a result, all further calculations of the electronic relaxation energy are performed using a single iteration. All calculations in Section 2.3.2 and Section 2.3.3 employ this relaxation correction.

Fig. 2.2 shows that the average contribution of the relaxation energy to the vertical IE of EC in the EC system is -0.51 eV. Although this contribution is fairly large, the standard deviation of the distribution is only 0.04 eV. Thus the electronic relaxation energy adds an approximately constant shift to the vertical IEs within the EC system.

It is important to note that although the above protocol allows for electronic relaxation, the relaxation process constrains the excess charge associated with the oxidized state to reside in the active region. It is worth noting that our relaxation protocol thus achieves an effect that is similar to that of constrained DFT.[98]

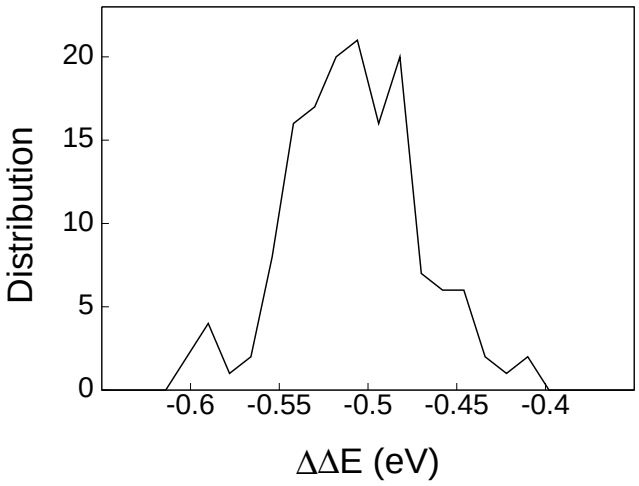

Figure 2.2: Contribution of the electronic relaxation of subsystem B to the vertical IE of subsystem A, calculated for EC molecules in the EC system, and using $r_{\mathrm{DFT}} = 4.0$ Å and $r_{\mathrm{MM}} = 50.0$ Å.

2.2.3.3 Effective Point Charges for the MM Region

Atoms in the MM region (indicated by the brown circle in Fig. 2.1) are represented using point charges. We now describe a protocol for determining these point charges such that they reproduce, on average, the B3LYP/cc-pVDZ molecular dipoles in the solvent environment. The following protocol is performed using the ensemble of sampled molecular configurations from the classical MD trajectories. The resulting point-charge values are reported in Table 2.3.

In the first step of the protocol, for each liquid configuration of the neat EC and neat DMC systems a B3LYP/cc-pVDZ calculation is performed on the active molecule in isolation from the solvent environment. From these B3LYP/cc-pVDZ calculations, we obtain the molecular dipole of the molecule and the Mulliken charge of each atom in the molecule. An initial set of point charges for the EC and DMC molecules is obtained by first averaging the Mulliken charges of atoms having the same bonding configurations and then scaling these charges by a multiplicative factor in order to reproduce the average magnitude of the molecular dipole.

In the second step, a set of B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding calcu-

lations is performed using the same liquid configurations of neat EC and neat DMC from the first step. The molecules in the MM region are represented using the sets of point charges obtained from the single-molecule calculations in the first step, and the embedding cutoffs are set to $r_{\text{border}} = 0.0$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å. A new set of point charges is obtained by scaling the averaged Mulliken charges of the molecule in the active region to reproduce the average magnitude of the molecular dipole. These charges are then used to perform another set of B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding calculations, in order to self-consistently determine the correct charge distribution and molecular dipole moment of the molecules in the condensed phase. After a total of three iterations of this self-consistent process, the average magnitude of the molecular dipoles of both EC and DMC are found to change by less than 0.01 Debye, indicating that the calculation of the MM point charges is converged.

The final point charges for EC and DMC molecules in the respective neat systems are provided in Table 2.3. In order to determine an appropriate set of point charges for EC and DMC molecules in the EC:DMC system, the same iterative process is applied, except that for the first iteration, the initial set of point charges are chosen to be the converged point charges of EC in the EC system and DMC in the DMC system. As shown in Table 2.3, after a single iteration, the point charges for EC and DMC in the mixture change very little relative to their values in the corresponding neat systems, so no further iterations are performed.

For the purpose of performing the WFT-in-DFT-in-MM embedding calculations, it is not necessary to determine a set of point charges for the solvent molecules in their oxidized states, because only the molecule in the active region is ever represented in the oxidized state. Nonetheless, for the analysis performed in Section 2.3.3.3, it is useful to have a point-charge representation of the oxidized solvent molecules. The point-charge representation of EC^+ is determined by performing a single set of

	EC		
Atom	EC	EC:DMC	EC^+
C′	0.35813	0.35212	0.42044
C	0.16731	0.16377	0.12942
O′	-0.34325	-0.34090	-0.05581
O	-0.31361	-0.30984	-0.14457
H	0.06943	0.07023	0.16641
	DMC		
Atom	DMC	EC:DMC	DMC^+
C′	0.64825	0.67547	0.41529
C	0.24169	0.25136	0.08120
O′	-0.63655	-0.66405	-0.08538
O	-0.51865	-0.53580	-0.12720
H	0.09037	0.09291	0.12702

Table 2.3: Point Charges Representing the EC and DMC Molecules

Top: Values of the point charges representing EC molecules in neat EC, EC molecules in the EC:DMC mixtures, and EC^+ cations in neat EC, with C′ and O′ representing the carbonyl carbon and carbonyl oxygen atom types, respectively. The average Mulliken point charges for EC molecules in the EC system are all scaled by a factor of 1.1811 to obtain the reported values, while the average Mulliken point charges for EC molecules in the EC:DMC mixture are scaled by a factor of 1.1819. Bottom: Values of the point charges representing DMC molecules in neat DMC, DMC molecules in the EC:DMC mixtures, and DMC^+ cations in neat DMC. The average Mulliken point charges for DMC molecules in the DMC system are all scaled by a factor of 2.0105 to obtain the reported values, while the average Mulliken point charges for DMC molecules in the EC:DMC mixture are scaled by a factor of 2.0620.

B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding calculations on the neat EC system in the oxidized state, using $r_{\text{border}} = 0.0$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å. The same approach was also used to determine the point-charge representation of DMC^+. Ensemble-averaged Mulliken charges for the EC and DMC molecules in the oxidized state are provided in Table 2.3. In order to avoid altering the +1 charge, the partial charges on the molecules in the oxidized systems are not scaled to self-consistently reproduce the molecular dipoles.

2.2.3.4 Convergence Tests for the Embedding Cutoffs

Numerical tests are performed to demonstrate the convergence of the embedding parameters r_{border}, r_{DFT}, and r_{MM}. All tests are performed using the ensemble of 151 liquid configurations of the neat EC system. Fig. 2.3(a) shows the effect of increasing r_{border} from 0.0 Å to 2.5 Å on the distribution of vertical IEs. Even at $r_{\text{border}} = 0.0$ Å, which corresponds to TF embedding with no projection, the truncation error is fairly small; this result is consistent with the fact that the atoms in the active region are not covalently bound to the atoms in the environment. Fig. 2.3(b) shows the convergence of the vertical IEs with respect to increasing r_{DFT}. The standard deviation of this distribution is 0.06 eV, indicating that the results are reasonably well converged at $r_{\text{DFT}} = 3.0$ Å. Fig. 2.3(c) shows the convergence of the vertical IEs with respect to the value of r_{MM}. The standard deviation of this distribution is 0.11 eV, indicating that the results are reasonably well converged at $r_{\text{DFT}} = 25.0$ Å. For the remainder of this paper, the more conservative cutoffs of $r_{\text{border}} = 2.5$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å are employed in the condensed-phase embedding calculations.

2.2.3.5 Convergence Tests for the Basis Set

To illustrate the difference between vertical IEs calculated at the B3LYP and CCSD(T) levels of theory, and to demonstrate convergence of the vertical IEs with respect to the size of the active-region basis set, we perform calculations on isolated EC and DMC molecules. Fig. 2.4 shows the results of CCSD(T)-in-B3LYP-in-MM and B3LYP-in-B3LYP-in-MM calculations on isolated molecules ($r_{\text{border}} = r_{\text{DFT}} = r_{\text{MM}} = 0.0$ Å) for configurations of EC in the EC system and DMC in the DMC system. Calculations are performed using the cc-pVDZ, aug-cc-pVDZ, and aug-cc-pVTZ basis sets for the active region. Fig. 2.4(a) and Fig. 2.4(b) show that at the B3LYP level of theory, the vertical IEs calculated using the aug-cc-pVDZ basis set are higher than those calculated using the cc-pVDZ basis set. Increasing the size of the basis set

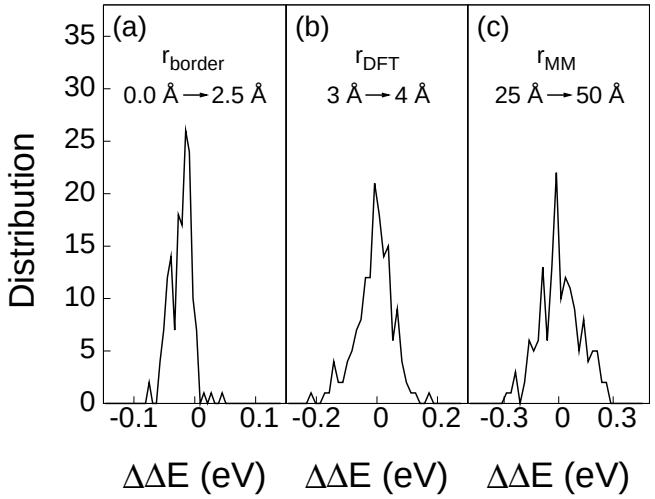

Figure 2.3: Demonstration of the convergence of each of the embedding cutoffs for CCSD(T)-in-B3LYP-in-MM/cc-pVDZ embedding calculations on configurations of the EC system. Except where noted for each part, the embedding cutoffs are set to $r_{\mathrm{MM}} = 50.0$ Å, $r_{\mathrm{DFT}} = 4.0$ Å, and $r_{\mathrm{border}} = 2.5$ Å. (a) Change in the vertical IEs upon increasing r_{border} from 0.0 Å to 2.5 Å. (b) Change in the vertical IEs upon increasing r_{DFT} from 3.0 Å to 4.0 Å. (c) Change in the vertical IEs upon increasing r_{MM} from 25.0 Å to 50.0 Å.

from aug-cc-pVDZ to aug-cc-pVTZ only negligibly increases the B3LYP vertical IEs. Fig. 2.4(c) and Fig. 2.4(d) show that the CCSD(T) results follow a similar trend, except that the difference between the results obtained using the aug-cc-pVDZ and aug-cc-pVTZ basis sets is slightly larger. This is consistent with the fact that DFT tends to converge more quickly with respect to basis set size than WFT methods.

For EC, the average vertical IE at the CCSD(T)/aug-cc-pVTZ level is 11.11 eV, which is consistent with the experimentally measured gas-phase vertical IEs in the range of 10.89-11.45 eV.[99–102] The average B3LYP/aug-cc-pVTZ vertical IE is 10.82 eV, which falls below this experimental range. For DMC, the average vertical IE at the CCSD(T)/aug-cc-pVTZ level is 11.02 eV, which is consistent with the experimentally measured IE of 11.00 eV.[101,102]

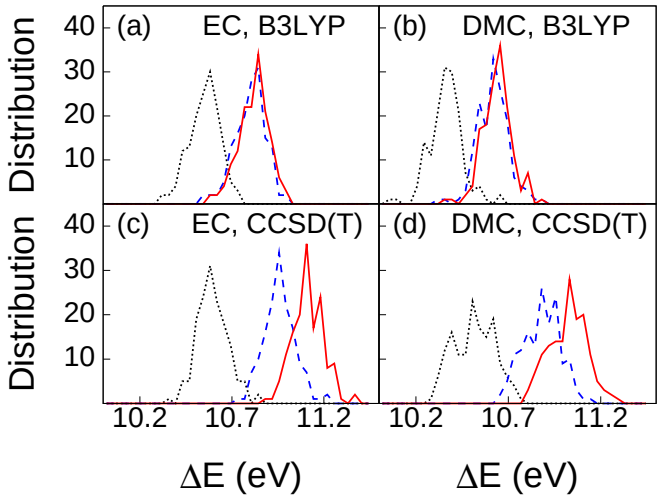

Figure 2.4: (a) The distribution of vertical IEs of configurations of isolated EC molecules, calculated at the B3LYP level of theory and using the cc-pVDZ basis set (black, dotted), the aug-cc-pVDZ basis set (blue, dashed), and the aug-cc-pVTZ basis set (red, solid). Similarly, the other panels show the distribution of vertical IEs for isolated molecules of (b) DMC at the B3LYP level of theory, (c) EC at the CCSD(T) level of theory, and (d) DMC at the CCSD(T) level of theory.

2.3 Results

2.3.1 WFT-in-DFT Corrects Over-Delocalization of the Electron Hole

An important reason for the use of WFT methods for the study of oxidation potentials is the propensity of DFT to over-delocalize electron holes.[103–107] In this section, it is confirmed that DFT functionals exhibit this problem when applied to carbonate solvents, and it is further shown that WFT-in-DFT embedding corrects this problem. Additionally, we show that even at the DFT-in-DFT level of theory, the projection-based embedding method allows for the enforcement of constraints to ensure localization of the electron hole.

Fig. 2.5(a) and Fig. 2.5(b) illustrate the electron hole produced by the B3LYP/aug-cc-pVDZ and HF/aug-cc-pVDZ levels of theory, respectively, for a configuration of two EC molecules in the oxidized state, $(EC_2)^+$. At the B3LYP level, the electron

hole is delocalized across both molecules, while at the HF level, it is localized on a single molecule. As a measure of the extent of localization of the electron hole, we define $\delta_{\max} \equiv \max\{\delta_A, \delta_B\}$, where δ_A and δ_B are the summed Mulliken charges of the atoms in molecules A and B, respectively. The blue squares in Fig. 2.5(c) show that for a series of configurations of $(EC_2)^+$, B3LYP yields values of $\delta_{\max}$ that are close to 0.5, indicating that the electron hole is significantly delocalized between the two molecules. These dimer configurations are obtained from configurations of the neat EC system taken 500 ps apart, using the molecule in the active region and a randomly selected molecule within 4.0 Å. In contrast to the B3LYP results, the black circles in Fig. 2.5(c) show that HF theory produces values of $\delta_{\max}$ that are always close to +1.0, indicating that the electron hole is well localized on a single molecule. The black pluses and blue crosses in Fig. 2.5(c) demonstrate that the unrestricted versions of these methods (UB3LYP and UHF, respectively) yield results that are nearly identical to the restricted open-shell versions of the methods. Shown by the hollow red diamonds, the M05 functional[108] is found to exhibit somewhat more localization than B3LYP, but a significant amount of charge is still shared between molecules. The red triangles show that the related M05-2X functional[109] typically produces highly localized electron holes, although in four of the sixteen test cases, the M05-2X holes are much more delocalized.

As another means of characterizing the extent of delocalization of the electron hole, we define the delocalization energy,

$$E_{\text{deloc}} \equiv \Delta E_{AB} - \min\{\Delta E_A, \Delta E_B\}. \tag{2.8}$$

Here A and B correspond to each of the molecules in the two-molecule systems described above, $\min\{x, y\}$ represents the minimum of x and y, ΔE_{AB} is the vertical IE of the full two-molecule system, and ΔE_X is the vertical IE obtained from an

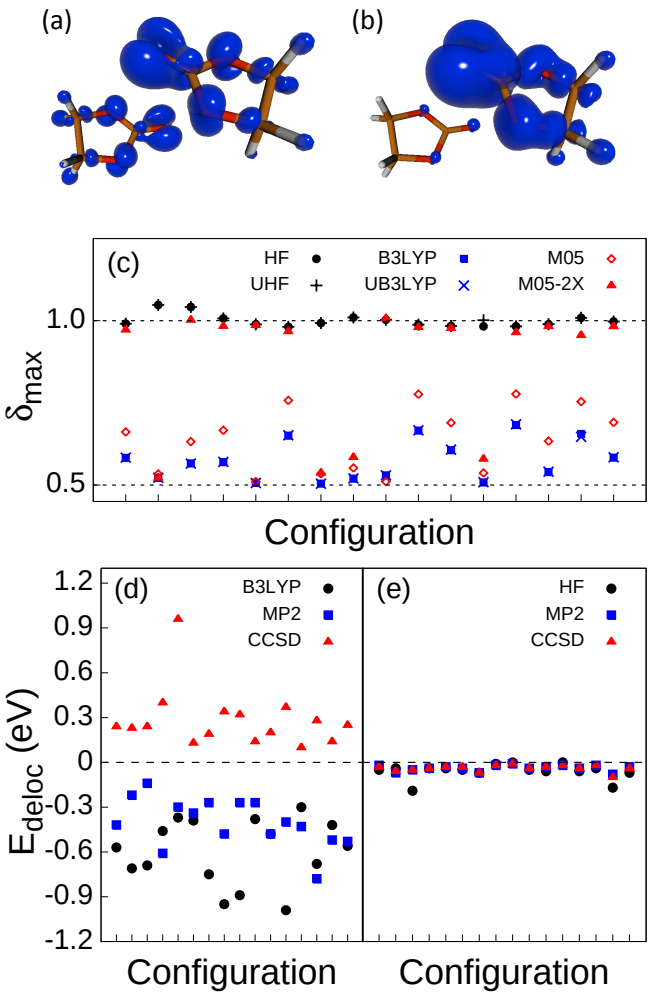

Figure 2.5: (a) Isosurface of the oxidized electron hole density of a two molecule EC system at $\rho = 0.005$ and calculated at the B3LYP/aug-cc-pVDZ level of theory. (b) Same as (a), except at the HF/aug-cc-pVDZ level of theory. (c) Largest combined Mulliken charge ($\delta_{\max}$) on either molecule in configurations of a two-molecule EC system, calculated at the HF (black, circles), unrestricted-HF (black, pluses), B3LYP (blue, squares), unrestricted-B3LYP (blue, crosses), M05 (red, hollow diamonds), and M05-2X (red, triangles) levels of theory. All calculations employed the aug-cc-pVDZ basis set. (d) The B3LYP delocalization energy (black, circles), the MP2 delocalization energy using B3LYP reference MOs (blue, squares), and the CCSD delocalization energy using B3LYP reference MOs (red, triangles). (e) The HF delocalization energy (black, circles), the MP2 delocalization energy using HF reference MOs (blue, squares), and the CCSD delocalization energy using HF reference MOs (red, triangles).

embedding calculation in which molecule X is the molecule in the active region. The aug-cc-pVDZ basis set is used for all atoms, and the embedding calculations use $r_{\text{border}} = \infty$ and do not include the electronic relaxation correction introduced in Section 2.2.3.2. The delocalization energy is defined such that a large, negative value indicates a strong energetic penalty associated with confinement of the electron hole on a single molecule.

The black circles in Fig. 2.5(d) show that the delocalization energy of the B3LYP calculations is both large and negative, consistent with the observation that B3LYP theory tends to produce highly delocalized holes; conversely, the black circles in Fig. 2.5(e) show that the delocalization energy of the HF calculations is small, consistent with the observation that HF theory tends to produce highly localized holes. The red triangles in Fig. 2.5(d) show the result of CCSD calculations that are performed using B3LYP reference MOs, while the red triangles in Fig. 2.5(e) show the result of CCSD calculations that are performed using HF reference MOs. When using B3LYP reference MOs, the delocalization energies at the CCSD level of theory are positive, implying that CCSD theory favors holes that are more localized than those of B3LYP theory.

From Fig. 2.5, it is thus apparent that the delocalized electron holes from the B3LYP level of theory are an artifact associated with the tendency of DFT functionals to over-delocalize charge,[103–107] and that this artifact is corrected at the CCSD level of theory. This conclusion is reinforced by the trend observed in Fig. 2.5(c), in which the amount of localization exhibited by each method increases with the amount of exact exchange incorporated into the method (i.e., in the order B3LYP < M05 < M05-2X < HF).[107] Fig. 2.5(d) also reveals that when using B3LYP reference MOs, the delocalization energies at the MP2 level of theory are negative, indicating that MP2 theory favors electron holes that are more delocalized than those of CCSD theory. The propensity of DFT to over-delocalize the electron hole emphasizes the importance

of treating the active region using high-level WFT methods, such as coupled-cluster theory.

2.3.2 Solvation Effects on Different Length Scales

We now examine the effect of the solvent environment on the vertical IEs of condensed-phase EC and DMC molecules. Fig. 2.6(a) shows the distribution of vertical IEs calculated from a set of CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ embedding calculations performed on configurations of EC molecules in the neat EC system (black) and DMC molecules in the neat DMC system (blue), using embedding cutoffs of $r_{\text{border}} = 2.5$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å. The average vertical IE of DMC in the condensed-phase DMC system is 0.85 eV lower than the average vertical IE of EC in the condensed-phase EC system, despite the observation in Fig. 2.4 that the average vertical IE of isolated EC and DMC molecules are within 0.1 eV one another. The large difference between the condensed-phase vertical IEs of the EC and DMC systems is a result of solvation effects, which are further investigated in Fig. 2.6(b) and Fig. 2.6(c).

Fig. 2.6(b) shows the degree to which inclusion of the DFT region contributes to the distributions in Fig. 2.6(a). Keeping $r_{\text{MM}} = 0.0$ Å for all calculations, the distributions show the change in vertical IEs resulting from increasing the value of r_{DFT} from 0.0 Å to 4.0 Å in embedding calculations for the vertical IE of EC molecules in the EC system (black) and for DMC molecules in the DMC system (blue). Inclusion of the DFT region shifts the average vertical IE of EC in the EC system by -0.1 eV, while shifting the average vertical IE of DMC in the DMC system by -0.6 eV. This accounts for a majority of the shift in the average vertical IEs of the condensed-phase EC and DMC systems.

Fig. 2.6(c) shows the degree to which inclusion of the MM region contributes to the distributions in Fig. 2.6(a). The distributions show the change in vertical

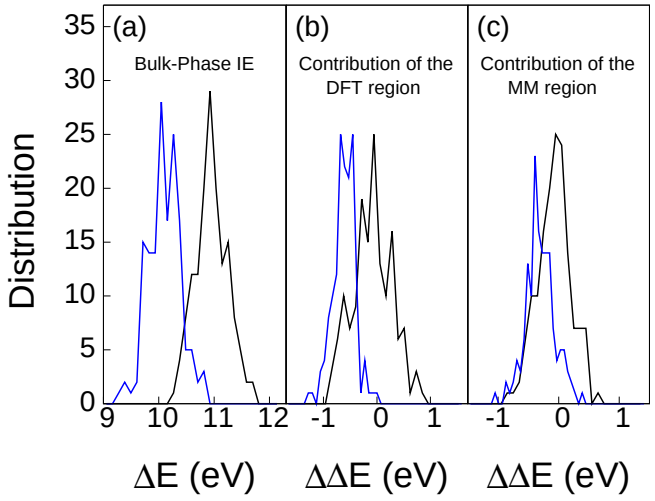

Figure 2.6: (a) Distribution of vertical IEs from a series of CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ embedding calculations on EC molecules in the EC system (black) and DMC molecules in the DMC system (blue), using embedding cutoffs of $r_{\mathrm{border}} = 2.5$ Å, $r_{\mathrm{DFT}} = 4.0$ Å, and $r_{\mathrm{MM}} = 50.0$ Å. (b) Contribution of the DFT region to the vertical IEs calculated in (a). The distributions are obtained from the difference between vertical IEs calculated using $r_{\mathrm{border}} = 2.5$ Å, $r_{\mathrm{DFT}} = 4.0$ Å, $r_{\mathrm{MM}} = 0.0$ Å and vertical IEs calculated using $r_{\mathrm{border}} = 0.0$ Å, $r_{\mathrm{DFT}} = 0.0$ Å, $r_{\mathrm{MM}} = 0.0$ Å. (c) Contribution of the MM region to the vertical IEs calculated in (a). The distributions are obtained from the difference between the vertical IEs calculated in part (a) and vertical IEs calculated using $r_{\mathrm{border}} = 2.5$ Å, $r_{\mathrm{DFT}} = 4.0$ Å, $r_{\mathrm{MM}} = 0.0$ Å.

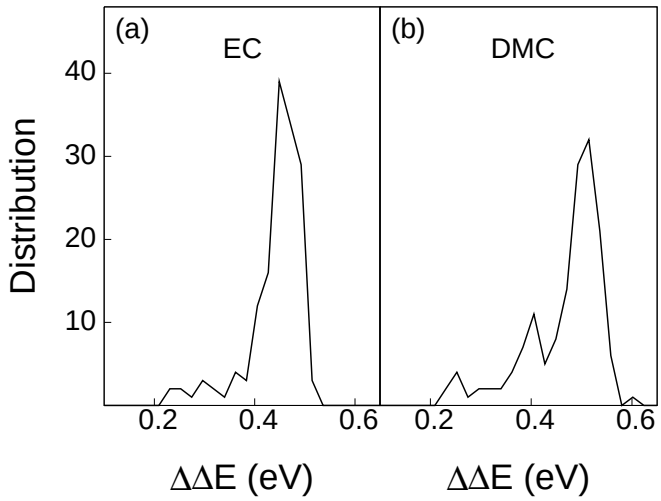

Figure 2.7: The difference between the CCSD(T)-in-B3LYP/aug-cc-pVTZ and B3LYP-in-B3LYP/aug-cc-pVTZ vertical IEs for (a) the EC in the EC system and (b) the DMC in the DMC system. The calculations are performed using embedding cutoffs of $r_{\text{border}} = 2.5$ Å, $r_{\text{MM}} = 50.0$ Å, and $r_{\text{DFT}} = 4.0$ Å.

IEs resulting from increasing the value of r_{pc} from 0.0 Å to 50.0 Å in embedding calculations for the vertical IE of EC molecules in the EC system (black) and for DMC molecules in the DMC system (blue). Inclusion of the MM region shifts the average vertical IE of EC in the EC system by -0.1 eV, while shifting the average vertical IE of DMC in the DMC system by -0.3 eV. Although these average shifts are somewhat smaller than those caused by inclusion of the DFT region, they are still substantial.

Fig. 2.7 illustrates the effect of describing the active region at the CCSD(T) level of theory, rather than using B3LYP, in the condensed-phase IE calculations. The distributions show the difference between IEs calculated using CCSD(T)-in-B3LYP-in-MM embedding and IEs calculated using B3LYP-in-B3LYP-in-MM embedding for both EC in the EC system (Fig. 2.7(a)) and DMC in the DMC system (Fig. 2.7(b)). In the condensed-phase, B3LYP-in-B3LYP-in-MM embedding on average underestimates the vertical IEs by more than 0.4 eV relative to CCSD(T)-in-B3LYP-in-MM embedding.

2.3.3 Solvent Response to Oxidation

2.3.3.1 EC and DMC Solvation Obeys Linear Response

Fig. 2.8 examines the statistics of EC solvation from the results of CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ embedding calculations on neat EC, using $r_{\text{border}} = 2.5$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å. The dotted black line in Fig. 2.8(a) shows the distribution of vertical IEs for a neutral EC molecule in the condensed phase, taken over the ensemble of configurations associated with the neutral EC system; this distribution is identical to that presented for EC in Fig. 2.6(a). The dotted blue line in Fig. 2.8(a) shows the distribution of vertical IEs for a neutral EC molecule in the condensed phase, taken over the ensemble of configurations associated with the EC^+ system, where the vertical IE calculation is performed on the particular EC molecule that carried a positive charge in the generation of the EC^+ ensemble. The solid curves show Gaussian fits to each of these distributions, which have a standard deviation of 0.29 eV and 0.27 eV, respectively. These results are consistent with the application of linear response theory for the statistical distribution of solvent configurations, which predicts that the two distributions should exhibit Gaussian statistics having the same standard deviation.[49–53] Because linear response theory holds for this system, it is possible to use Eq. 2.1 and Eq. 2.2 to construct Marcus-type free-energy parabolas corresponding to the half-reaction associated with oxidation of an EC molecule, as is depicted in Fig. 2.8(b).[49–53]

Similarly, Fig. 2.9 shows the corresponding vertical IE distributions for neat DMC, EC in the EC:DMC mixture, and DMC in the EC:DMC mixture. As shown in Table 2.4, the standard deviation of each of the reduced distributions is within approximately 10% of the corresponding oxidized distribution. This small difference in standard deviations is consistent with that of other systems that have been identified as obeying linear response theory.[49–53] We thus conclude that, neglecting any

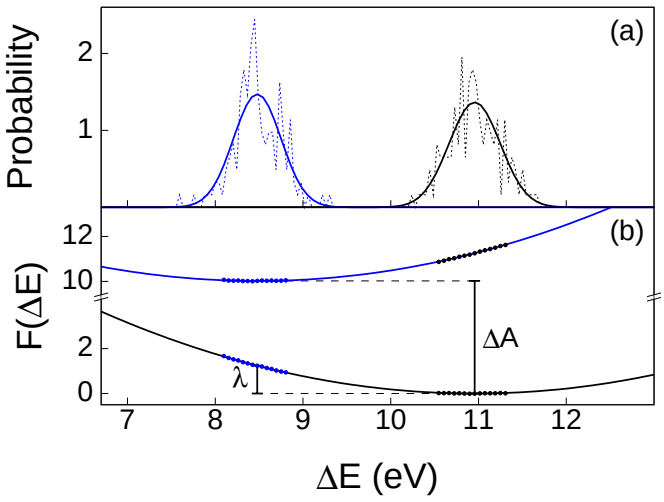

Figure 2.8: (a) Equilibrium probability distributions, $P_{\mathrm{M}}(\Delta E)$, of the vertical IE of EC molecules, ΔE, calculated using CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ embedding. "M" corresponds either to the reduced EC system (R, black) or the oxidized EC^+ system (O, blue). The distributions have similar standard deviations, implying that the linear response approximation is accurate for this system. The best fit Gaussian distributions, $g_{\mathrm{M}}(\Delta E)$, are indicated in solid lines. (b) Diabatic free energy profiles constructed from the equilibrium distributions shown in (a). The solid lines indicate the parabolas obtained from ΔA and λ, which are determined by applying Eq. 2.2 and Eq. 2.1. The sets of data points near the minimum of each parabola (i.e., the upper left and lower right sets of data points) were obtained by applying $F_{\mathrm{M}}(\Delta E) = -\beta^{-1}\ln\left(P_{\mathrm{M}}(\Delta E)\right) + F_{\mathrm{M}}^{\mathrm{ref}}$, where $F_{\mathrm{M}}^{\mathrm{ref}} = \beta^{-1}\ln\left(g_{\mathrm{M}}^{\mathrm{max}}\right) + \delta_{\mathrm{MO}}\Delta A$, and $g_{\mathrm{M}}^{\mathrm{max}}$ is the maximum of the Gaussian fit in (a). The linear free energy relation $F_{\mathrm{O}}(\Delta E) - F_{\mathrm{R}}(\Delta E) = \Delta E$ is then applied to these data points in order to obtain the other sets of data points (i.e., the lower left and upper right sets of data points).

subsequent oxidative decomposition reactions, the oxidation of EC and DMC solvents likewise obey linear response theory.

Molecule	System	σ, Neutral	σ, Oxidized
EC	EC	0.29(2)	0.27(2)
DMC	DMC	0.29(2)	0.28(1)
EC	EC:DMC	0.33(1)	0.31(1)
DMC	EC:DMC	0.29(2)	0.33(2)

Table 2.4: Standard Deviations of the Vertical IE Distributions.

Each standard deviation, σ, corresponds to one of the vertical IE distributions in Fig. 2.8(a) or Fig. 2.9. The statistical uncertainty associated with the last reported digit in these quantities, reported in parentheses, is estimated by first calculating the standard deviations of five sets of 30 vertical IEs, and then calculating the standard error of the mean of these values. All values are reported in eV.

The preceding demonstration of linear response justifies the use of Eq. 2.1, Eq. 2.2, and Eq. 2.3 to calculate the reorganization energies and oxidation potentials for each system, which are provided in Table 2.5. The oxidation potential calculated at the CCSD(T)-in-B3LYP-in-MM level is consistently 0.4-0.5 V higher than the oxidation potential calculated at the B3LYP-in-B3LYP-in-MM level. The CCSD(T)-in-B3LYP-in-MM and B3LYP-in-B3LYP-in-MM reorganization energies are in much closer agreement, with all of the reorganization energies being approximately in the range 1.2-1.3 eV.

2.3.3.2 Failure of Conventional Dielectric Continuum Models for the Treatment of DMC

It is informative to compare the reorganization energies in Table 2.5 with the results of various dielectric continuum models, starting first with a very simple description. For a spherical ion exchanging an electron with an electrode infinitely far away, the reorganization energy is[110]

$$\lambda = \frac{e^2}{8\pi\epsilon_o r}\left(\frac{1}{\epsilon_{\text{optical}}} - \frac{1}{\epsilon_{\text{static}}}\right), \tag{2.9}$$

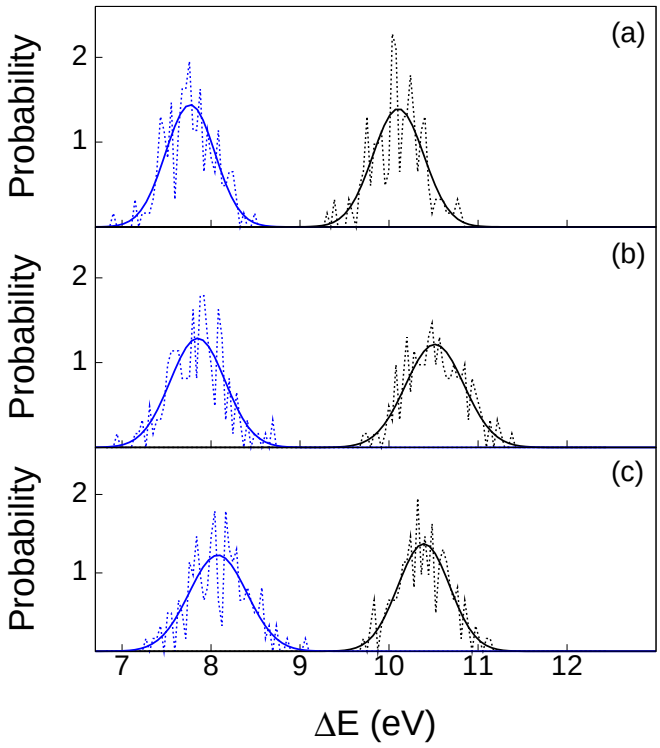

Figure 2.9: Vertical IE probability distributions calculated using CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ embedding. (a) Vertical IE probability distribution of DMC molecules in the DMC system (black) and DMC molecules in the DMC^+ system (blue). (b) Vertical IE probability distribution of EC molecules in the EC:DMC system (black) and EC molecules in the EC^+:DMC system (blue). (c) Vertical IE probability distribution of DMC molecules in the EC:DMC system (black) and DMC molecules in the EC:DMC^+ system (blue).

B3LYP-in-B3LYP-in-MM					
Molecule	System	$\langle \Delta E \rangle_0$	$\langle \Delta E \rangle_{+1}$	$\mathscr{E}$	λ
EC	EC	10.51(2)	8.03(2)	7.87(2)	1.24(2)
DMC	DMC	9.64(2)	7.31(2)	7.07(2)	1.17(2)
EC	EC:DMC	10.07(3)	7.40(3)	7.33(2)	1.34(2)
DMC	EC:DMC	9.93(2)	7.64(2)	7.39(2)	1.15(2)
CCSD(T)-in-B3LYP-in-MM					
Molecule	System	$\langle \Delta E \rangle_0$	$\langle \Delta E \rangle_{+1}$	$\mathscr{E}$	λ
EC	EC	10.95(2)	8.48(2)	8.32(2)	1.24(2)
DMC	DMC	10.10(2)	7.77(2)	7.54(2)	1.17(2)
EC	EC:DMC	10.51(3)	7.85(3)	7.78(2)	1.33(2)
DMC	EC:DMC	10.39(2)	8.08(3)	7.83(2)	1.16(2)

Table 2.5: Embedding Results for the Oxidation Potentials and Reorganization Energies

The results are obtained using $r_{\text{border}} = 2.5$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å and the aug-cc-pVTZ basis set. The top results correspond to B3LYP-in-B3LYP-in-MM embedding, while the lower results correspond to CCSD(T)-in-B3LYP-in-MM embedding. The average vertical IE of each molecule in the indicated system, $\langle \Delta E \rangle_0$, is provided. Similarly, the average vertical IE of each molecule in the corresponding oxidized system, $\langle \Delta E \rangle_{+1}$, is provided. The value in parentheses indicates the statistical uncertainty in the last reported digit. The intrinsic oxidation potentials, $\mathscr{E}$, and solvent reorganization energies, λ, are also provided, using Eq. 2.1 and Eq. 2.3. The oxidation potentials are reported in V relative to Li^+/Li, while all other values are reported in eV.

where r is the radius of the ion and $\epsilon_{\text{optical}}$ and ϵ_{static} are the optical and static dielectric constants, respectively. As an estimate of the radius of EC and DMC, we use the radius of spheres having a volume equal to the inverse number density of neat EC and neat DMC, yielding radii of 2.99 Å and 3.29 Å, respectively. The value of $\epsilon_{\text{optical}}$ corresponds to the square of the index of refraction,[111] which is 1.4158 for EC at 40°C and 1.369 for DMC.[112–114] The experimental values of ϵ_{static} are 89.78 for EC and 3.107 for DMC.[2] Using these values, Eq. 2.9 predicts a reorganization energy of 1.17 eV for neat EC, in good agreement with the embedding results in Table 2.5, and a reorganization energy of 0.46 eV for neat DMC, which is substantially lower than the reorganization energy obtaining using embedding.

For additional comparison with the reorganization energies in Table 2.5, a set of

B3LYP/aug-cc-pVTZ calculations are performed using two continuum models: the integral-equation-formalism polarizable continuum model (IEF-PCM)[115,116] and the solvation model density (SMD) model.[117] The calculations are performed in the oxidized state, using B3LYP/aug-cc-pVTZ optimized geometries of the molecules in the corresponding solvated reduced state. The calculations on each molecule use the Gaussian09 default parameters for a generic solvent,[118] except that the values of the static and optical dielectric constants are replaced by the corresponding experimental values for the neat solvent. The solvent reorganization energy is

$$\lambda = G_{\mathrm{neq}} - G_{\mathrm{eq}}, \tag{2.10}$$

where the subscript "eq" indicates that the solvent continuum is allowed to fully relax to the electronic distribution of the oxidized molecule, and the subscript "neq" indicates that the free energy calculation is performed using a nonequilibrium solvation dielectric continuum associated with the reaction field of the neutral system.[103] Both zero-point energy and intramolecular vibrational contributions to the oxidation potential of EC and DMC molecules are neglected, since they are expected to be small.[10,31] Table 2.6 provides the results of these calculations, which are nearly identical to the results from Eq. 2.9. For all three dielectric continuum models considered here, it is found that the reorganization energy of DMC is substantially underestimated in comparison to the embedding calculations that include explicit solvation.

The oxidation potentials calculated using the implicit solvent models also disagree with the results of the embedding calculations. Using the IEF-PCM solvation model, the oxidation potential of neat EC is 7.39 V, while the oxidation potential of neat DMC is 7.94. Similarly, the oxidation potentials for neat EC and neat DMC in the SMD solvation model are 7.06 V and 7.78 V, respectively. This is qualitatively different from the results in Table 2.5, where the oxidation potential of neat EC is

found to be higher than that of neat DMC.

Molecule	Model	λ_{ref}	λ
EC	Eq. 2.9	1.24	1.17
EC	IEF-PCM		1.17
EC	SMD		1.19
DMC	Eq. 2.9	1.17	0.46
DMC	IEF-PCM		0.48
DMC	SMD		0.48

Table 2.6: Dielectric Continuum Results for the Reorganization Energies

The reorganization energies, λ, for the oxidation of EC and DMC molecules are calculated using various dielectric continuum models. The calculations in IEF-PCM and SMD solvation are run using the B3LYP/aug-cc-pVTZ level of theory and the Gaussian09 defaults for a generic solvent, except that the values of ϵ_0 and ϵ_∞ are set to the corresponding experimental values for the indicated molecule. Also shown for each molecule is λ_{ref}, the B3LYP-in-B3LYP-in-MM reorganization energy of the neat liquid from Table 2.5. All values are reported in eV.

It is clear that these continuum models fail to correctly account for the relevant intermolecular interactions between DMC molecules. In Section 2.3.3.3, we demonstrate that this failure is a consequence of the large permanent quadrupole moment of DMC molecules.

2.3.3.3 Importance of Quadrupolar Interactions in DMC Solvation

To investigate the inaccuracy of the dielectric continuum models for describing DMC solvation, we first investigate the nature of the intermolecular interactions that determine the relative vertical IEs of each system. Fig. 2.10 uses a simple point charge model to confirm that the vertical IEs are primarily modulated by electrostatic interactions in both the neat EC and neat DMC systems. In this point charge model, all atoms are represented using the corresponding charge from Table 2.3, allowing calculation of the electrostatic interaction energy between the atoms in the oxidized molecule and all other atoms, E_{pc}. The change in E_{pc} upon switching the atomic charges from the neutral to the oxidized systems, ΔE_{pc}, is then calculated for each configuration

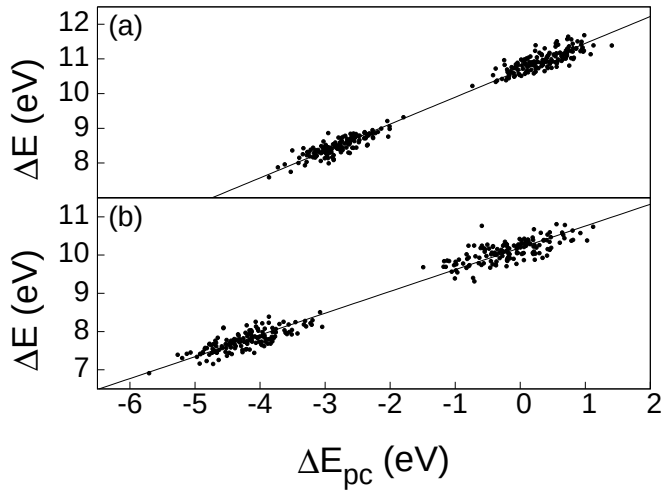

Figure 2.10: Examination of the accuracy of the electrostatic model for estimation of the vertical IE energy of EC and DMC molecules. (a) The vertical IE of EC in the EC and EC^+ systems, calculated using CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ embedding ($r_{\text{border}} = 2.5$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å), is plotted against ΔE_{pc}. The best-fit line is given by the equation $\Delta E = 0.78\Delta E_{\text{pc}} + 10.67$, with an R^2 value of 0.98. Because of the bimodal nature of this distribution, a more informative measure of the quality of this fit is the R^2 value of the line when only one mode (i.e., the data for either the EC or EC^+ system) is incorporated into the calculation of R^2. This yields R^2 values of 0.59 for the EC system and 0.69 for the EC^+ system. (b) The corresponding result for DMC in the DMC and DMC^+ systems. The best-fit line is given by the equation $\Delta E = 0.57\Delta E_{\text{pc}} + 10.19$, with an R^2 value of 0.97. The corresponding R^2 values for the individual DMC and DMC^+ systems are 0.42 and 0.57, respectively.

of the liquid. Fig. 2.10 shows the correlation between ΔE_{pc} and the corresponding CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ embedding calculations; the two clusters of data points in each panel correspond to the configurations associated with the oxidized and neutral systems. Despite its simplicity, the point-charge model correlates directly with the embedding results. This implies that the primary factor influencing the relative vertical IEs of each solvent molecule is the extent to which ionization of the molecule is stabilized by electrostatic interactions with the surrounding solvent environment.

Further examination of these electrostatic interactions yields an explanation for the failure of dielectric continuum representations of DMC. It is possible to deter-

mine the relative contribution of dipole and quadrupole intermolecular interactions to ΔE_{pc} by representing all molecules, aside from the molecule to be oxidized, using a truncated multipole expansion. We define $E_{\text{n-pole}}$ as the electrostatic interaction energy between the point charges associated with the molecule to be oxidized and the truncated multipole representation of all the other molecules. Fig. 2.11(a) compares each value of ΔE_{pc} for the neat EC system with the corresponding value of $\Delta E_{\text{n-pole}}$, in which the multipole expansion is truncated at either the dipole or quadrupole level. In the limit in which $\Delta E_{\text{n-pole}}$ is calculated using all terms in the multipole expansion, ΔE_{pc} and $\Delta E_{\text{n-pole}}$ would agree exactly, and all points would reside on the black line. Even when the multipole expansion is truncated at the level of dipoles, the points in Fig. 2.11(a) agree well with the black line, indicating that intermolecular interactions between EC molecules are dominated by dipole effects, with only a small contribution from higher-order terms in the multipole expansion. Fig. 2.11(b) shows that this is not the case for DMC; instead, an accurate representation of the intermolecular interactions between DMC molecules requires the incorporation of quadrupole effects. Section 2.6 confirms that this conclusion is robust with respect to the parameterization of the DMC point charges.

The dielectric constant of a material reflects the strength of the dipole interactions of the material.[119,120] Nonetheless, in other explicit solvent models, it has been observed that although the solvation properties of a material can be significantly enhanced by a large molecular quadrupole moment, increasing the magnitude of the quadrupole moment often reduces the dielectric constant.[121–124] Correctly accounting for these quadrupole interactions in implicit solvent models is challenging, and involves the introduction of parameters that report on microscopic-scale properties.[125] For these reasons, dielectric continuum models that do not incorporate the relevant molecular-level details of solvent interactions underestimate the reorganization energy of DMC, as observed in Table 2.6.

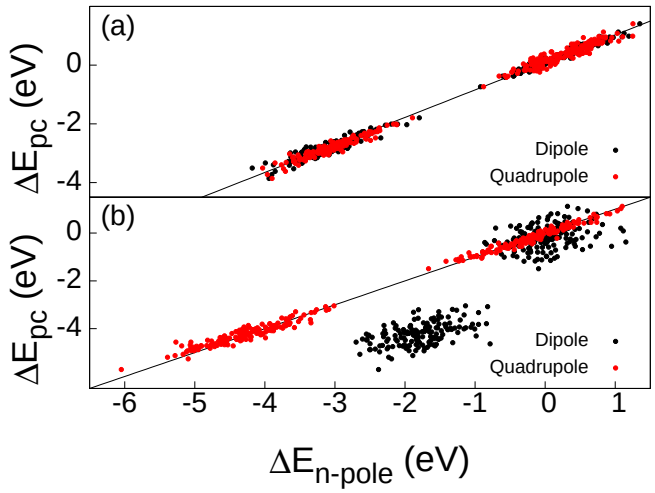

Figure 2.11: The change in electrostatic energy upon oxidation, obtained using a simple point-charge model for each atom (ΔE_{pc}), is compared against the change in electrostatic energy upon oxidation, obtained by truncating the point-charges of each molecule as a multipole expansion truncated at either the dipole or quadrupole levels ($\Delta E_{\text{n-pole}}$). The dipoles and quadrupoles are calculated relative to the center of mass of each molecule. Values are provided for (a) EC in both the EC and EC^+ systems and for (b) DMC in both the DMC and DMC^+ systems. The lines correspond to the equation $\Delta E_{\text{pc}} = \Delta E_{\text{n-pole}}$, and deviations from these lines indicate truncation error in the multipole expansion.

The important role of quadrupolar interactions in DMC solutions provides useful insight into the nature of ion solvation in Li-ion batteries. It has long been expected that because EC has a much larger dipole moment than DMC, the solvation shell of ions in EC:DMC mixtures should exhibit a strong preference for neighboring EC molecules instead of DMC molecules.[126] For example, this intuition is consistent with the Born model of ion solvation,[127–129]

$$E_{\text{solvation}} = -\frac{Z^2 e^2}{8\pi\epsilon_o r}\left(1 - \frac{1}{\epsilon_{\text{static}}}\right), \tag{2.11}$$

where $E_{\text{solvation}}$ is the solvation energy of an ion having charge Z and radius r. Eq. 2.11 predicts that the solvation energy of an ion in EC should be higher than that of an ion in DMC, leading to the conclusion that in EC:DMC mixtures, Li^+ cations should exhibit strong preference for EC solvation. Studies of this issue have led to mixed conclusions; whereas the results of electrospray ionization mass spectrometry[130–132] and ^{13}C NMR spectroscopy[133,134] have identified only a negligible amount of DMC in coordination with Li^+ cations, gas-phase QM calculations,[55] liquid-phase MD simulations,[55] Raman spectroscopy,[135] and ^{17}O NMR spectroscopy[136] have indicated that DMC participates significantly in the coordination of Li^+ cations. Although previously regarded as counter-intuitive, the possibility that DMC participates non-negligibly in the coordination of Li^+ cations can be understood in terms of the results in Fig. 2.11. Despite having a small dipole and small dielectric constant, DMC molecules exhibit large quadrupolar interactions that can favorably solvate ions. Models that do not account for quadrupolar interactions are thus likely to underestimate the favorability of ion solvation by DMC.

2.4 Conclusions

We have employed high-accuracy WFT-in-DFT embedding methods to analyze the oxidative stability of liquid-phase mixtures of EC and DMC. DFT calculations with the B3LYP functional are found to systematically underestimate the gas-phase vertical IEs by about 0.4-0.5 eV, and many DFT functionals are found to over-delocalize the electron hole in the oxidized electrolyte solutions. To avoid these failures, we apply the newly-developed projection-based embedding method to accurately calculate the CCSD(T)-level vertical IEs, adiabatic oxidation potentials, and reorganization energies of EC and DMC molecules in explicit solvent.

By performing embedding calculations over ensembles of liquid configurations obtained from classical MD simulations, we obtain equilibrium distributions of vertical IEs. These distributions reveal that the reorganization of the solvent is consistent with the predictions of linear response theory, allowing for calculation of the oxidation potentials and the solvent reorganization energies. The solvent reorganization energies of neat EC and neat DMC are both found to be 1.2 eV. Although simple dielectric continuum models accurately reproduce the reorganization energy of EC, they significantly underestimate the reorganization energy of DMC, which we demonstrate to be a result of the important role of quadrupolar interactions in DMC solvation. In addition to explaining the large solvent reorganization energy of DMC, this observation provides insight into previously identified anomalies, such as the possibility that DMC strongly coordinates with Li^+ cations in EC:DMC mixtures.

2.5 Appendix: Benchmarking the Electronic Relaxation Calculations

Fig. 2.12 demonstrates both the convergence and accuracy of the electronic relaxation protocol described in Section 2.2.3.2. Results are provided for the errors, relative to full B3LYP calculations, of a series of IEs calculated using supermolecular ($r_{\text{border}} = \infty$) B3LYP-in-B3LYP/cc-pVDZ embedding. Inclusion of the electronic relaxation energy is observed to reduce the errors by approximately an order of magnitude, with a single iteration of the third step of the electronic relaxation protocol being sufficient to converge the energies. As a result, all relaxation calculations in the main text are performed using a single iteration of the relaxation protocol.

The data are calculated for twelve geometry configurations, obtained by modifying a single configuration from the EC ensemble. Each configuration includes the molecule in the active region and one of the molecules in the DFT region, with the molecule in the active region being replaced by a calcium atom at its center-of-mass, and with the active region corresponding to the calcium atom. This replacement was performed because the low IE of calcium ensures that the electron hole of the oxidized system is well localized on subsystem A, even at the B3LYP level of theory. If this analysis were performed using EC dimers instead of the calcium-EC system, it would be difficult to isolate errors caused by over-delocalization of the B3LYP electron hole (see Section 2.3.1 of the main text) from errors in the relaxation energy.

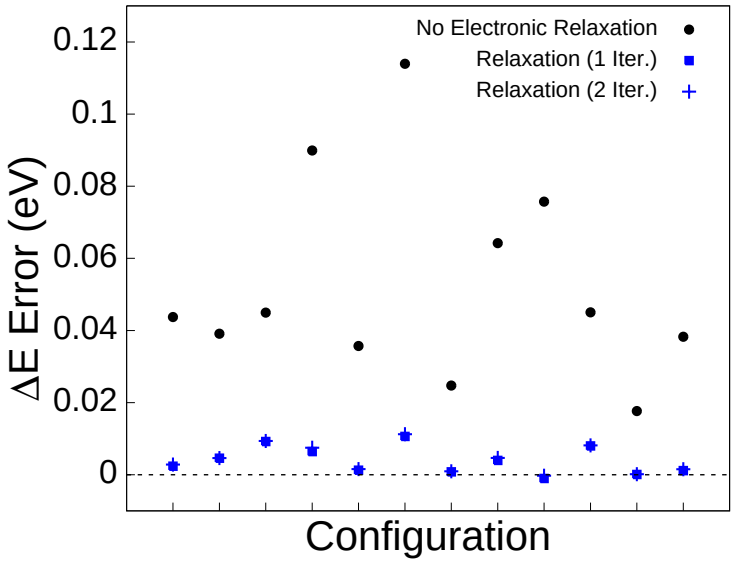

Figure 2.12: Demonstration of the accuracy of the electronic relaxation protocol described in Section 2.2.3.2 of the main text. Errors in the vertical IEs of a series of supermolecular B3LYP-in-B3LYP/cc-pVDZ embedding calculations are reported relative to the corresponding full B3LYP calculations. The calculations are performed for the case in which the electronic density of subsystem B is not permitted to relax with respect to the oxidation of subsystem A (black, circles), the case in which the electronic density of subsystem B is permitted to relax in a single iteration (blue, squares), and the case in which the electronic density of subsystem B is permitted to relax in two iterations (blue, pluses).

2.6 Appendix: Robustness of the Results with Respect to the Representation of the MM Region

In Section 2.3.3.3, we observed that implicit solvent models substantially underestimate the reorganization energy of DMC, due to the neglect of important quadrupolar interactions. In this section, we demonstrate that this conclusion is robust with respect to the point-charge representation of the MM region.

Table 2.7 provides the average magnitudes of the molecular dipole and quadrupole moments of both EC and DMC. These results are obtained using B3LYP-in-B3LYP-in-MM/cc-pVDZ embedding on 16 configurations of the neat EC system and 16 configurations of the neat DMC system, taken at 500 ps intervals. The embedding cutoffs are set to $r_{\text{border}} = 0.0$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å. The dipole and quadrupole moments of the molecule in the active region ($\vec{\mathbf{D}}$ and $\overleftrightarrow{\mathbf{Q}}$, respectively) are calculated relative to its center-of-mass, and the magnitudes of these moments are obtained using the expressions $D \equiv \left(\vec{\mathbf{D}} \cdot \vec{\mathbf{D}}\right)^{1/2}$ and $Q \equiv \left(\frac{2}{3}\overleftrightarrow{\mathbf{Q}} : \overleftrightarrow{\mathbf{Q}}\right)^{1/2}$, respectively.[137,138]

Molecule	$\langle D \rangle_{\text{DFT}}$	$\langle Q \rangle_{\text{DFT}}$	$\langle D \rangle_{\text{MM}}$	$\langle Q \rangle_{\text{MM}}$
EC	6.5(1)	3.2(1)	6.55(3)	3.2(1)
DMC	0.75(5)	11.8(1)	0.56(7)	29.4(1)

Table 2.7: Magnitude of the Dipole and Quadrupole Moments of EC and DMC

The average magnitude of the dipole and quadrupole moments of EC and DMC is obtained using both B3LYP-in-B3LYP-in-MM embedding (DFT), and using the point charge (MM) model described in Section 2.2.3.3. The dipole moments are reported in Debye, while the quadrupole moments are reported in Debye · Å.

Table 2.7 also provides the average magnitudes of the molecular dipole and quadrupole moments of EC and DMC, obtained using the MM point-charge representation described in Section 2.2.3.3. The molecular dipoles produced by the MM representation accurately reflect those obtained at the DFT level of theory, as expected due to the

way in which the point charges were determined in Section 2.2.3.3. For EC, the magnitudes of the B3LYP-level and MM quadrupoles agree very well, but for DMC, the MM representation overestimates the quadrupole moment by a factor of 2.487. We now investigate this issue to confirm that it does not impact our conclusions from Section 2.3.3.3.

We first confirm that the reported conclusions regarding the oxidation potential and reorganization energy of neat DMC calculated at the CCSD(T)-in-B3LYP-in-MM level are not sensitive to the magnitude of the quadrupoles in the MM region. In order to examine the sensitivity of the reorganization energy and oxidation potential of neat DMC to the magnitude of the DMC quadrupoles, we perform another set of CCSD(T)-in-B3LYP-in-MM/aug-cc-pVTZ embedding calculations on DMC in both the DMC and DMC^+ systems. This set of calculations utilizes a point-charge representation in which the charges for each atom in the MM region is reduced by a factor of 2.487 relative to the values reported in Table 2.3 of the main text, and the embedding cutoffs are set to $r_{\text{border}} = 2.5$ Å, $r_{\text{DFT}} = 4.0$ Å, and $r_{\text{MM}} = 50.0$ Å. The resulting value of $\langle \Delta E \rangle_0$ is 10.28 eV, while the value of $\langle \Delta E \rangle_{+1}$ is 8.34 eV. Using these average vertical IEs and Eq. 2.3 results in an oxidation potential of 7.91 V. While this value is 0.37 V higher than that reported in Table 2.5, this difference does not affect the conclusion that the oxidation potential of neat EC is higher than that of neat DMC. Similarly, using the above ensemble-averaged vertical IEs and Eq. 2.1 results in a solvent reorganization energy of 0.97 eV. The close agreement between this value and the neat DMC reorganization energy of 1.17 eV reported in Table 2.5 of the main text indicates that our CCSD(T)-in-B3LYP-in-MM reorganization energies are robust with respect to the point-charge representation of the MM region. The somewhat better robustness of the reorganization energies with respect to the parameterization of the point charges is largely a consequence of cancelation of errors in Eq. 2.1.

Secondly, we reconfirm the conclusion from Section 2.3.3.3 of the main text that quadrupolar intermolecular interactions are important for correctly describing DMC solvation, even if the magnitude of these quadrupolar interactions is reduced by a factor of 2.487. For a series of configurations for the EC^+ system, Fig. 2.13(a) shows $\Delta E_{\text{n}_{\text{th}}\text{-pole}}$, which we define as the contribution of the n_{th} order terms in the multipole expansion of $\Delta E_{\text{n-pole}}$; $\Delta E_{\text{n}_{\text{th}}\text{-pole}}$ results for both the dipole and quadrupole interactions are plotted. As was observed in Fig. 2.11 in the main text, the quadrupolar interactions (blue, squares) in neat EC are negligible compared to the dipolar interactions (black, circles). By contrast, Fig. 2.13(b) demonstrates the quadrupolar interactions in neat DMC, calculated using the point charges from Section 2.2.3.3, are approximately as large as the corresponding dipolar interactions. Even when the magnitudes of the quadrupole moments in neat DMC are reduced by a factor of 2.487, the resulting quadrupolar interactions (red, triangles) remain comparable to the dipolar interactions. We thus conclude that for any reasonable parameterization of the point charge representation of the MM region, the solvation properties of DMC will be strongly affected by quadrupolar interactions.

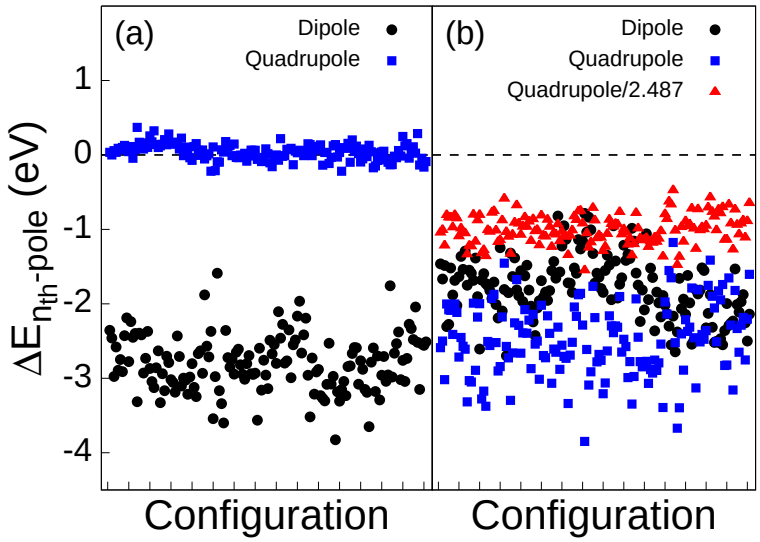

Figure 2.13: Demonstration that the conclusion that DMC quadrupolar interactions are significant is robust with respect to the parameterization of point charges. (a) The dipolar (black, circles) and quadrupolar (blue, squares) contributions to $\Delta E_{\text{n-pole}}$, $\Delta E_{\text{n}_{\text{th}}\text{-pole}}$, for each configuration of EC^+. (b) As (a), except that the dipolar and quadrupolar contributions are calculated for configurations of DMC^+. Also provided is the contribution of the DMC quadrupoles when the quadrupoles are reduced by a factor of 2.487 (red, triangles).

Bibliography

[1] Whittingham, M. S. Lithium Batteries and Cathode Materials. *Chem. Rev.* **2004**, *104*, 4271-4301

[2] Xu, K. Nonaqueous Liquid Electrolytes for Lithium-Based Rechargeable Batteries. *Chem. Rev.* **2004**, *104*, 4303-4417

[3] Yang, L.; Ravdel, B.; Lucht, B. L. Electrolyte Reactions with the Surface of High Voltage LiNi0.5Mn1.5O4 Cathodes for Lithium-Ion Batteries. *Electrochem. Solid-State Lett.* **2010**, *13*, A95-A97

[4] Goodenough, J. B.; Kim, Y. Challenges for Rechargeable Li Batteries. *Chem. Mater.* **2010**, *22*, 587-603

[5] von Cresce, A.; Xu, K. Electrolyte Additive in Support of 5 V Li Ion Chemistry Batteries and Energy Storage. *J. Electrochem. Soc.* **2011**, *158*, A337-A342

[6] Hu, M.; Pang, X.; Zhou, Z. Recent Progress in High-Voltage Lithium Ion Batteries. *J. Power Sources.* **2013**, *237*, 229-242

[7] Imhof, R.; Novák, P. Oxidative Electrolyte Solvent Degradation in Lithium-Ion Batteries - An *In Situ* Differential Electrochemical Mass Spectrometry Investigation. *J. Electrochem. Soc.* **1999**, *146*, 1702-1706

[8] Xing, L. D.; Li, W. S.; Wang, C. Y.; Gu, F. L.; Xu, M. Q.; Tan, C. L.; Yi, J. Theoretical Investigations on Oxidative Stability of Solvents and Oxidative

Decomposition Mechanism of Ethylene Carbonate for Lithium Ion Battery Use. *J. Phys. Chem. B.* **2009**, *113*, 16596-16602

[9] Arora, P.; White, R. E.; Doyle, M. Capacity Fade Mechanisms and Side Reactions in Lithium-Ion Batteries. *J. Electrochem. Soc.* **1998**, *145*, 3647-3667

[10] Borodin, O.; Behl, W.; Jow, T. R. Oxidative Stability and Initial Decomposition Reactions of Carbonate, Sulfone, and Alkyl Phosphate-Based Electrolytes. *J. Phys. Chem. C.* **2013**, *117*, 8661-8682

[11] Wang, Y.; Xing, L.; Borodin, O.; Huang, W.; Xu, M.; Li, X.; Li, W. Quantum Chemistry Study of the Oxidation-Induced Stability and Decomposition of Propylene Carbonate-Containing Complexes. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6560-6567

[12] Xing, L.; Borodin, O. Oxidation Induced Decomposition of Ethylene Carbonate from DFT calculations - Importance of Explicitly Treating Surrounding Solvent. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12838-12843

[13] Ein-Eli, Y.; Howard, W. F.; Lu, S. H.; Mukerjee, S.; McBreen, J.; Vaughey, J. T.; Thackeray, M. M. LiMn2-xCuxO4 Spinels (0.1 ¡= x ¡= 0.5): A New Class of 5 V Cathode Materials for Li Batteries - I. Electrochemical, Structural, and Spectroscopic Studies. *J. Electrochem. Soc.* **1998**, *145*, 1238-1244

[14] Wolfenstine, J.; Allen, J. LiNiPO4-LiCoPO4 Solid Solutions as Cathodes. *J. Power Sources.* **2004**, *136*, 150-153

[15] Wang, D.; Xiao, J.; Xu, W.; Nie, Z.; Chongmin, W.; Graff, G.; Zhang, J.-G. Preparation and Electrochemical Investigation of Li2CoPO4F Cathode Material for Lithium-Ion Batteries. *J. Power Sources.* **2011**, *196*, 2241-2245

[16] Demeaux, J.; De Vito, E.; Lemordant, D.; Le Digabel, M.; Galiano, H.; Caillon-Caravanier, M.; Claude-Montigny, B. On the Limited Performances of Sulfone Electrolytes Towards the LiNi0.4Mn1.6O4 Spinel. *Phys. Chem. Chem. Phys.* **2013**, *15*, 20900-20910

[17] Xu, K.; Angell, C. A. Sulfone-Based Electrolytes for Lithium-Ion Batteries. *J. Electrochem. Soc.* **2002**, *149*, A920-A926

[18] Armand, M.; Endres, F.; MacFarlane, D. R.; Ohno, H.; Scrosati, B. Ionic-Liquid Materials for the Electrochemical Challenges of the Future. *Nat. Mater.* **2009**, *8*, 621-629

[19] Ong, S. P.; Andreussi, O.; Wu, Y. B.; Marzari, N.; Ceder, G. Electrochemical Windows of Room-Temperature Ionic Liquids from Molecular Dynamics and Density Functional Theory Calculations. *Chem. Mater.* **2011**, *23*, 2979-2986

[20] Tian, Y.-H.; Goff, G. S.; Runde, W. H.; Batista, E. R. Exploring Electrochemical Windows of Room-Temperature Ionic Liquids: A Computational Study. *J. Phys. Chem. B.* **2012**, *116*, 11943-11952

[21] Xing, L.; Borodin, O.; Smith, G. D.; Li, W. Density Functional Theory Study of the Role of Anions on the Oxidative Decomposition Reaction of Propylene Carbonate. *J. Phys. Chem.* **2011**, *115*, 13896-13905

[22] Borodin, O.; Jow, T. R. Quantum Chemistry Studies of the Oxidative Stability of Carbonate, Sulfone and Sulfonate-Based Electrolytes Doped with BF4 -, PF6 - Anions. *ECS Trans.* **2011**, *33*, 77-84

[23] Leung, K. First-Principles Modeling of the Initial Stages of Organic Solvent Decomposition on LixMn2O4(100) Surfaces. *J. Phys. Chem. C.* **2012**, *116*, 9852-9861

[24] Kanamura, K.; Toriyama, S.; Shiraishi, S.; Takehara, Z. Studies on Electrochemical Oxidation of Nonaqueous Electrolytes Using In Situ FTIR Spectroscopy. *J. Electrochem. Soc.* **1995**, *142*, 1383-1389

[25] Kanamura, K. Anodic Oxidation of Nonaqueous Electrolytes on Cathode Materials and Current Collectors for Rechargeable Lithium Batteries. *J. Power Sources.* **1999**, *81*, 123-129

[26] Ue, M.; Murakami, A.; Nakamura S.; Anodic Stability of Several Anions Examined by *Ab Initio* Molecular Orbital and Density Functional Theories. *J. Electrochem. Soc.* **2002**, *149*, A1572-A1577

[27] Borodin, O.: Molecular Modeling of Electrolytes. In *Electrolytes for Lithium and Lithium-Ion Batteries*; Jow, T. R., Xu, K., Borodin, O., Ue, M., Eds.; Springer: New York; 2014

[28] Bhatt, M. D.; Cho, M.; Cho, K. Interaction of Li+ ions with Ethylene Carbonate (EC): Density Functional Theory Calculations. *Applied Surface Science.* **2010**, *257*, 1463-1468

[29] Leung, K.; Tenney, C. M. Toward First Principles Prediction of Voltage Dependences of Electrolyte/Electrolyte Interfacial Processes in Lithium Ion Batteries. *J. Phys. Chem. C.* **2013**, *117*, 24224-24235

[30] Leung, K. Electronic Structure Modeling of Electrochemical Reactions at Electrode/Electrolyte Interfaces in Lithium Ion Batteries. *J. Phys. Chem. C.* **2013**, *117*, 1539-1547

[31] Johansson, P. Intrinsic Anion Oxidation Potentials. *J. Phys. Chem. A.* **2006**, *110*, 12077-12080

[32] Johansson, P. Intrinsic Anion Oxidation Potentials. *J. Phys. Chem. A*. **2007**, *111*, 1378-1379

[33] Assary, R. S.; Curtiss, L. A.; Redfern, P. C.; Zhang, Z.; Amine, K. Computational Studies of Polysiloxanes: Oxidation Potentials and Decomposition Reactions. *J. Phys. Chem. C.* **2011**, *115*, 12216-12223

[34] Zhang, X.; Pugh, J. K.; Ross, P. N. Computation of Thermodynamic Oxidation Potentials of Organic Solvents Using Density Functional Theory. *J. Electrochem. Soc.* **2001**, *148*, E183-E188

[35] Fu, Y.; Liu, L.; Yu, H.-Z.; Wang, Y.-M.; Guo, Q.-X. Quantum-Chemical Predictions of Absolute Standard Redox Potentials of Diverse Organic Molecules and Free Radicals in Acetonitrile. *J. Am. Chem. Soc.* **2005**, *127*, 7227-7234

[36] Zhang, Z.; Lu, J.; Assary, R. S.; Du, P.; Wang, H.-H.; Sun, Y.-K.; Qin, Y.; Lau, K. C.; Greeley, J.; Redfern. P. C.; Iddir, H.; Curtiss, L. A.; Amine, K. Increased Stability Toward Oxygen Reduction Products for Lithium-Air Batteries with Oligoether-Functionalized Silane Electrolytes. *J. Phys. Chem. C.* **2011**, *115*, 25535-25542

[37] Kang, S.; Park, M. H.; Lee, H.; Han, Y.-K. A Joint Experimental and Theoretical Determination of the Structures of Oxidized and Reduced Molecules. *Electrochem. Commun.* **2012**, *23*, 83-86

[38] Li, T.; Xing, L.; Li, W.; Peng, B.; Xu. M.; Gu. F.; Hu, S. Theoretic Calculation for Understanding the Oxidation Process of 1,4-Dimethoxybenzene-Based Compounds as Redox Shuttles for Overcharge Protection of Lithium Ion Batteries. *J. Phys. Chem. A*. **2011**, *115*, 4988-4994

[39] Wang, R. L.; Buhrmester, C.; Dahn, J. R. Calculations of Oxidation Potentials

of Redox Shuttle Additives for Li-ion Cells. *J. Electrochem. Soc.* **2006**, *153*, A445-A449

[40] Leung, K.; Qi, Y.; Zavadil, K. R.; Jung, Y. S.; Dillon, A. C.; Cavanagh, A. S.; Lee, S.-H.; George, S. M. Using Atomic Layer Deposition to Hinder Solvent Decomposition in Lithium Ion Batteries: First-Principles Modeling and Experimental Studies. *J. Am. Chem. Soc.* **2011**, *133*, 14741-14754

[41] Assary, R. S.; Curtiss, L. A.; Moore, J. S. Toward a Molecular Understanding of Energetics in Li-S Batteries Using Nonaqueous Electrolytes: A High-Level Quantum Chemical Study. *J. Phys. Chem. C.* **2014**, *118*, 11545-11558

[42] Das, U.; Lau, K. C.; Redfern, P. C.; Curtiss, L. A. Structure and Stability of Lithium Superoxide Clusters and Relevance to Li-O-2 Batteries. *J. Phys. Chem. Lett.* **2014**, *5*, 813-819

[43] Hammer, N. I.; Hinde, R. J.; Compton, R. N.; Diri, K.; Jordan, K. D.; Radisic, D.; Stokes, S. T.; Bowen, K. H. Dipole-Bound Anions of Highly Polar Molecules: Ethylene Carbonate and Vinylene Carbonate. *J. Chem. Phys.* **2004**, *120*, 685-690

[44] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479-483

[45] Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F., III A Simple, Exact Density-Functional-Theory Embedding Scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564-2568

[46] Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller, T. F., III Accurate Basis Set Truncation for Wavefunction Embedding *J. Chem. Phys.* **2013**, *139*, 024103

[47] Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652

[48] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *Ab Initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623-11627

[49] Tateyama, Y.; Blumberger, J.; Sprik, M.; Tavernelli, I. Density-Functional Molecular-Dynamics Study of the Redox Reactions of Two Anionic, Aqueous Transition-Metal Complexes. *J. Chem. Phys.* **2005**, *122*, 234505

[50] Blumberger, J.; Sprik, M. *Ab Initio* Molecular Dynamics Simulation of the Aqueous Ru2+/Ru3+ Redox Reaction: The Marcus Perspective. *J. Phys. Chem. B.* **2005**, *109*, 6793-6804

[51] Blumberger, J.; Tavernelli, I.; Klein, M. L.; Sprik, M. Diabatic Free Energy Curves and Coordination Fluctuations for the Aqueous Ag+/Ag2+ Redox Couple: A Biased Born-Oppenheimer Molecular Dynamics Investigation. *J. Chem. Phys.* **2006**, *124*, 064507

[52] Tateyama, Y.; Blumberger, J.; Ohno, T.; Sprik, M. Free Energy Calculation of Water Addition Coupled to Reduction of Aqueous RuO4-. *J. Chem. Phys.* **2007**, *126*, 204506

[53] VandeVondele, J.; Ayala, R.; Sulpizi, M.; and Sprik, M. Redox Free Energies and One-Electron Energy Levels in Density Functional Theory Based *Ab Initio* Molecular Dynamics. *J. Electroanal. Chem.* **2007**, *607*, 113-120

[54] Borodin, O.; Smith, G. D. Development of Many-body Polarizable Force Fields for Li-Battery Components: 1. Ether, Alkane, and Carbonate-Based Solvents. *J. Phys. Chem. B.* **2006**, *110*, 6279-6292

[55] Borodin, O.; Smith, G. D. Quantum Chemistry and Molecular Dynamics Simulation Study of Dimethyl Carbonate: Ethylene Carbonate Electrolytes Doped with LiPF6. *J. Phys. Chem. B.* **2009**, *113*, 1763-1776

[56] Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341

[57] Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618-622

[58] Dunning, T. H., Jr. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007-1023

[59] Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796-6806

[60] Nosé, S.: Molecular-Dynamics Simulations at Constant Temperature and Pressure. In *Computer Simulation in Materials Science*; Meyer, M., Pontikis, V., Eds.; Kluwer Academic Publishers: Netherlands; 1991

[61] Martyna, G. J.; Tuckerman, M.; Tobias, D. J.; Klein, M. L. Explicit Reversible Integrators for Extended Systems Dynamics. *Mol. Phys.* **1996**, *87*, 1117-1157

[62] Steinhauser, O. Reaction Field Simulation of Water. *Mol. Phys.* **1982**, *45*, 335-348

[63] Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chem. Phys.* **1981**, *59*, 341-350

[64] Goodpaster, J. D.; Ananth, N.; Manby, F. R.; Miller, T. F., III Exact Non-additive Kinetic Potentials for Embedded Density Functional Theory. *J. Chem. Phys.* **2010**, *133*, 084103

[65] Goodpaster, J. D.; Barnes, T. A.; Miller, T. F., III Embedded Density Functional Theory for Covalently Bonded and Strongly Interacting Subsystems. *J. Chem. Phys.* **2011**, *134*, 164108

[66] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III Density Functional Theory Embedding for Correlated Wavefunctions: Improved Methods for Open-Shell Systems and Transition Metal Complexes. *J. Chem. Phys.* **2012**, *137*, 224113

[67] Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227-249

[68] Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. A Generalized Hybrid Orbital (GHO) Method For the Treatment of Boundary Atoms in Combined QM/MM Calculations. *J. Phys. Chem. A.* **1998**, *102*, 4714-4721

[69] Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where are We, and Where do We Go From Here? *Theor. Chem. Acc.* **2007**, *117*, 185-199

[70] Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198-1229

[71] Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A New ONIOM Implementation in Gaussian98. Part I. The Calculation of Energies, Gradients, Vibrational Frequencies and Electric Field Derivatives. *THEOCHEM.* **1999**, *461-462*, 1-21

[72] F. Maseras, F.; K. Morokuma, K. IMOMM: A New Integrated *Ab Initio* + Molecular Mechanics Geometry Optimization Scheme of Equilibrium Structures and Transition States. *J. Comp. Chem.* **1995**, *16*, 1170-1179

[73] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: An Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701-706

[74] Fedorov, D. G.; Kitaura, K. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A.* **2007**, *111*, 6904-6914

[75] Arora, P.; Li, W.; Piecuch, P.; Evans, J. W.; Albao, M.; Gordon, M. S. Diffusion of Atomic Oxygen on the Si(100) Surface. *J. Phys. Chem. C.* **2010**, *114*, 12649-12658

[76] Pruitt, S. R.; Addicoat, M. A.; Collins, M. A.; Gordon, M. S. The Fragment Molecular Orbital and Systematic Molecular Fragmentation Methods Applied to Water Clusters. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7752-7764

[77] Brorsen, K. R.; Minezawa, N.; Xu, F.; Windus, T. L.; Gordon, M. S. Fragment Molecular Orbital Molecular Dynamics with the Fully Analytic Energy Gradient. *J. Chem. Theory Comput.* **2012**, *8*, 5008-5012

[78] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III Accurate and Systematically Improvable Density Functional Theory Embedding for Correlated Wavefunctions. *J. Chem. Phys.* **2014**, *140*, 18A507

[79] Huo, P.; Uyeda, C.; Goodpaster, J. D.; Peters, J. C.; Miller, T. F., III *in preparation*

[80] Lykos, P. G.; Parr, R. G. On the Pi-Electron Approximation and Its Possible Refinement. *J. Chem. Phys.* **1956**, *24*, 1166-1173

[81] Phillips, J. C.; Kleinman, L. New Method for Calculating Wave Functions in Crystals and Molecules. *Phys. Rev.* **1959**, *116*, 287-294

[82] Stoll, H.; Paulus, B.; Fulde, P. On the Accuracy of Correlation-Energy Expansions in Terms of Local Increments. *J. Chem. Phys.* **2005**, *123*, 144108

[83] Mata, R. A.; Werner. H.-J.; Schütz, M. Correlation Regions within a Localized Molecular Orbital Approach. *J. Chem. Phys.* **2008**, *128*, 144106

[84] Henderson, T. M. Embedding Wave Function Theory in Density Functional Theory. *J. Chem. Phys.* **2006**, *125*, 014105

[85] Cantu, A. A.; Huzinaga, S. Theory of Separability of Many-Electron Systems. *J. Chem. Phys.* **1971**, **55**, 5543-5549

[86] Swerts, B.; Chibotaru, L. F.; Lindh, R.; Seijo, L.; Barandiaran, Z.; Clima S.; Pierloot, K.; Hendrickx, M. F. A. Embedding Fragment *Ab Initio* Model Potentials in CASSCF/CASPT2 Calculations of Doped Solids: Implementation and Applications. *J. Chem. Theory Comput.* **2008**, *4*, 586-594

[87] Pascual, J. L.; Barros, N.; Barandiaran, Z.; Seijo, L. Improved Embedding *Ab Initio* Model Potentials for Embedded Cluster Calculations. *J. Phys. Chem. A.* **2009**, *113*, 12454-12460

[88] Thomas, L. H. The Calculation of Atomic Fields. *Proc. Cambridge Philos. Soc.* **1927**, *23*, 542-548

[89] Fermi, E. A Statistical Method for Determining some Properties of the Atoms and its Application to the Theory of the Periodic Table of Elements. *Z. Phys.* **1928**, *48*, 73-79

[90] Wesolowski, T. A.; Warshel, A. Frozen Density-Functional Approach for *Ab-Initio* Calculations of Solvated Molecules. *J. Phys. Chem.* **1993**, *97*, 8050-8053

[91] Libisch, F.; Huang, C.; Carter, E. A. Embedded Correlated Wavefunction Schemes: Theory and Applications. *Acc. Chem. Res.* **2014**, *47*, 2768-2775

[92] Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. Application of the Electrostatically Embedded Many-Body Expansion to Microsolvation of Ammonia in Water Clusters. *J. Chem. Theory Comput.* **2008**, *4*, 683-688

[93] Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. Evaluation of the Electrostatically Embedded Many-Body Expansion and the Electrostatically Embedded Many-Body Expansion of the Correlation Energy by Application to Low-Lying Water Hexamers. *J. Chem. Theory Comput.* **2008**, *4*, 33-41

[94] Werner, H.-J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Shütz, M. *et al.*, MOLPRO, a package of *ab initio* programs, 2014, see www.molpro.net

[95] Watts, J. D.; Bartlett, R. J. The Coupled-Cluster Single, Double, and Triple Excitation Model for Open-Shell Single Reference Functions. *J. Chem. Phys.* **1990**, *93*, 6104-6105

[96] Crawford, T. D.; Kraka, E.; Stanton, J. F.; Cremer, D. Problematic p-Benzyne: Orbital Instabilities, Biradical Character, and Broken Symmetry. *J. Chem. Phys.* **2001**, *114*, 10638-10650

[97] Foster, J. M.; Boys, S. F. Canonical Configuration Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300-302

[98] Wu, Q.; Van Voorhis, T. Constrained Density Functional Theory and its Application in Long-Range Electron Transfer. *J. Chem. Theory Comput.* **2006**, *2*, 765-774

[99] Bain, A. D.; Frost, D. C. Studies of the Carbonyl Group in Some Five-membered Ring Compounds by Photoelectron Spectroscopy. *Can. J. Chem.* **1973**, *51*, 1245-1247

[100] Yokoyama, Y.; Jinno, M.; Watanabe, I.; Ikeda, S. Identification of Accidentally Degenerate Bands in UV Photoelectron Spectra of Ethylene Carbonate and Propylene Carbonate. *J. Electron Spectrosc.* **1974**, *5*, 1095-1101

[101] Meeks, J. L.; Arnett, J. F.; Larson, D.; McGlynn, S. P. Photoelectron Spectroscopy of Carbonyls. Ionization Assignments. *Chem. Phys. Lett.* **1975**, *30*, 190-194

[102] McGlynn, S. P.; Meeks, J. L. Photoelectron Spectra of Carbonyls, Carbonates, Oxalates and Esterification Effects. *J. Electron Spectrosc.* **1976**, *8*, 85-93

[103] Caricato, M.; Ingrosso, F.; Mennucci, B.; Sato, H. Electron Transfer in a Radical Ion Pair: Quantum Calculations of the Solvent Reorganization Energy. *J. Phys. Chem. B.* **2006**, *110*, 25115-25121

[104] Polo, V.; Kraka, E.; Cremer, D. Electron Correlation and the Self-Interaction Error of Density Functional Theory. *Mol. Phys.* **2002**, *100*, 1771-1790

[105] Dreuw, A.; Weisman, J. L.; Head-Gordon, M. Long-Range Charge-Transfer Excited States in Time-Dependent Density Functional Theory Require Non-Local Exchange. *J. Chem. Phys.* **2003**, *119*, 2943-2946

[106] Dreuw, A.; Head-Gordon, M. Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zincbacteriochlorin-Bacterlochlorin and Bacteriochlorophyll-Spheroidene Complexes. *J. Amer. Chem. Soc.* **2004**, *126*, 4007-4016

[107] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science.* **2008**, *321*, 792-794

[108] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Exchange-Correlation Functional with Broad Accuracy for Metallic and Nonmetallic Compounds, Kinetics, and Noncovalent Interactions. *J. Chem. Phys.* **2005**, *123*, 161103

[109] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory and Comput.* **2006**, *2*, 364-382

[110] Barbara, P. F.; Meyer, T. J.; Ratner, M. A. Contemporary Issues in Electron Transfer Research. *J. Phys. Chem.* **1996**, *100*, 13148-13168

[111] Hambleton, K. G.; Hilsum, C.; Holeman, B. R. Determination of the Effective Ionic Charge of Gallium Arsenide from Direct Measurements of the Dielectric Constant. *Proc. Phys. Soc.* **1961**, *77*, 1147-1148

[112] Peppel, W. J. Preparation and Properties of the Alkylene Carbonates. *Ind. Eng. Chem.* **1958**, *50*, 767-770

[113] Ping, Y. Y.; Lapidus, A. L. Oxidative Carbonylation of Organic Compounds by Carbon Monoxide and Transition-Metal Salt Catalysts. *Russ. Chem. Bull.* **1977**, *26*, 325-327

[114] Comelli, F.; Bigi, A.; Vitalini, D.; Rubini, K. Densities, Viscosities, Refractive Indices, and Heat Capacities of Poly(ethylene glycol-ran-propylene glycol) Plus Esters of Carbonic Acid at (293.15 and 313.15) K and at Atmospheric Pressure. *J. Chem. Eng. Data.* **2010**, *55*, 205-210

[115] Mennucci, B.; Cammi, R.; Tomasi, J. Excited States and Solvatochromic Shifts within a Nonequilibrium Solvation Approach: A New Formulation of the Integral Equation Formalism Method at the Self-Consistent Field, Configuration Interaction, and Multiconfiguration Self-Consistent Field Level. *J. Chem. Phys.* **1998**, *109*, 2798-2807

[116] Scalmani, G.; Frisch, M. J. Continuous Surface Charge Polarizable Continuum Models of Solvation. I. General Formalism. *J. Chem. Phys.* **2010**, *132*, 114110

[117] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B.* **2009**, *113*, 6378-6396

[118] Frisch M. J.; Trucks, G. W.; Schlegel, H. B.; *et al.*, Gaussian 09, Gaussian, Inc., Wallingford, CT, 2009

[119] Zwanzig, R. W. Influence of Molecular Electric Quadrupole Moments on Dielectric Constants. *J. Chem. Phys.* **1956**, *25*, 211-216

[120] Reynolds, L.; Gardecki, J. A.; Frankland, S. J. V.; Horng, M. L.; Maroncelli, M. Dipole Solvation in Nondipolar Solvents: Experimental Studies of Reorganization Energies and Solvation Dynamics. *J. Phys. Chem.* **1996**, *100*, 10337-10354

[121] Patey, G. N.; Valleau, J. P. Fluids of Spheres Containing Quadrupoles and Dipoles: A Study Using Perturbation Theory and Monte Carlo Computations. *J. Chem. Phys.* **1976**, *64*, 170-184

[122] Carnie, S. J.; Patey, G. N. Fluids of Polarizable Hard Spheres with Dipoles and Tetrahedral Quadrupoles Integral Equation Results with Application to Liquid Water. *Mol. Phys.* **1982**, *47*, 1129-1151

[123] Rick, S. W. A Reoptimization of the Five-Site Water Potential (TIP5P) for Use with Ewald Sums. *J. Chem. Phys.* **2004**, *120*, 6085-6093

[124] Milischuk, A.; Matyushov, D. V. On the Validity of Dielectric Continuum Models in Application to Solvation in Molecular Solvents. *J. Chem. Phys.* **2003**, *118*, 1859-1862

[125] Shiratori, K.; Morita, A. Theory of Quadrupole Contributions from Interface and Bulk in Second-Order Optical Processes. *Bull. Chem. Soc. Jpn.* **2012**, *85*, 1061-1076

[126] Joho, F.; Novák, P. SNIFTIRS Investigation of the Oxidative Decomposition of Organic-Carbonate-Based Electrolytes for Lithium-Ion Cells. *Electrochimica Acta.* **2000**, *45*, 3589-3599

[127] Cramer, C. J.; Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161-2200

[128] Lange, A. W.; Herbert, J. M. Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. I. An Improved Effective Coulomb Operator. *J. Chem. Theory Comput.* **2012**, *8*, 1999-2011

[129] Lange, A. W.; Herbert, J. M. Improving Generalized Born Models by Exploiting Connections to Polarizable Continuum Models. II. Corrections for Salt Effects. *J. Chem. Theory Comput.* **2012**, *8*, 4381-4392

[130] Matsuda, Y.; Fukushima, T.; Hashimoto, H.; Arakawa, R. Solvation of Lithium Ions in Mixed Organic Electrolyte Solutions by Electrospray Ionization Mass Spectroscopy. *J. Electrochem. Soc.* **2002**, *149*, A1045-A1048

[131] von Cresce, A.; Xu, K. Preferential Solvation of Li+ Directs Formation of

Interphase on Graphitic Anode. *Electrochem. Solid State Lett.* **2011**, *14*, A154-A156

[132] Xu, K.; Cresce, A. v. W. Li+-Solvation/Desolvation Dictates Interphasial Processes on Graphitic Anode in Li Ion Cells. *J. Mater. Res.* **2012**, *27*, 2327-2341

[133] Reddy, V. P.; Smart, M. C.; Chin, K. B.; Ratnakumar, B. V.; Surampudi, S.; Hu, J. B.; Yan, P.; Prakash, G. K. S. C-13 NMR Spectroscopic, CV, and Conductivity Studies of Propylene Carbonate-Based Electrolytes Containing Various Lithium Salts. *Electrochem. Solid-State Lett.* **2005**, *8*, A294-A298

[134] Yang, L.; Xiao, A.; Lucht, B. L. Investigation of Solvation in Lithium Ion Battery Electrolytes by NMR Spectroscopy. *J. Mol. Liq.* **2010**, *154*, 131-133

[135] Morita, M.; Asai, Y.; Yoshimoto, N.; Ishikawa, M. A Raman Spectroscopic Study of Organic Electrolyte Solutions Based on Binary Solvent Systems of Ethylene Carbonate with Low Viscosity Solvents which Dissolve Different Lithium Salts. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 3451-3456

[136] Bogle, X.; Vazquez, R.; Greenbaum, S.; Cresce, A. v. W. Understanding Li+-Solvent Interaction in Nonaqueous Carbonate Electrolytes with O-17 NMR. *J. Phys. Chem. Lett.* **2013**, *4*, 1664-1668

[137] Gubbins, K. E.; Gray, C. G.; Machado, J. R. S. Theory of Fluids of Non-Axial Quadrupolar Molecules. 1. Thermodynamic Properties. *Mol. Phys.* **1981**, *42*, 817-841

[138] Grubbins, K. E.; Gray, C. G. Theory of Fluids of Non-Axial Quadrupolar Molecules. 2. Structural and Spectroscopic Properties. *Mol. Phys.* **1981**, *42*, 843-860