UNIVERSITY OF CALIFORNIA
RIVERSIDE

Fragment-Based QM/MM Method for Modeling Molecular Crystals and Clusters

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

by

Kaushik Dhansukh Nanda

August 2013

Dissertation Committee:

Prof. Gregory J O Beran, Chairperson
Prof. Eric Chronister
Prof. Chia-en Chang

The Dissertation of Kaushik Dhansukh Nanda is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my advisor, Prof. Gregory Beran for his constant guidance, teachings and support. I absolutely admire his leadership qualities and management style which has helped in my development. I am also grateful to all my labmates, Kelly Theel, Yin Luo, Dr. Shuhao Wen, Yuanhang Huang, Joshua Hartman, Yonaton Heit and Dr. Ali Sebetci for many helpful and friendly discussions as well as productive collaborations. Finally, I would like to thank my parents, sister, extended family and friends for their invaluable support.

For the work illustrated in Chapter 2, the computer time was provided by the National Science Foundation Teragrid (TG-CHE090099). For the collaboration work shown in Chapter 3, I would like to thank our collaborators, Prof. Travis Fridgen, Elizabeth Gillis, Maria Demireva, and Prof. Evan. R. Williams. For this work in Chapter 3, we are grateful for the generous financial support of our work by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the U.S. National Science Foundation (Grant CHE-1012833 for ERW and CHE-1112568 for GB). The computational resources of both the Atlantic Computational Excellence Network (ACE-Net) and the Western Canada Research Grid (WestGrid) are gratefully acknowledged. EALG acknowledges the CGS-D granted by NSERC. Funding for the work shown in Chapter 4 from the National Science Foundation (CHE-1112568) and supercomputer time from the Teragrid (TG-CHE110064) are gratefully acknowledged. We would thank Prof. Bartolomeo Civalleri, Dr. Fook Tham and Dr. Shuhao Wen for helpful discussions. For the work shown in Chapter 6, Prof. Christoph Salzmann is gratefully acknowledged for providing us with the initial guess geometries of the various ice-XV polymorphs.

To my family and friends.

ABSTRACT OF THE DISSERTATION

Fragment-Based QM/MM Method for Modeling Molecular Crystals and Clusters

by

Kaushik Dhansukh Nanda

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, August 2013
Prof. Gregory J O Beran, Chairperson

Molecular aggregates like molecular crystals and clusters find important applications as pharmaceutical drugs, explosives, organic semi-conductors, materials for fuel storage, etc. These systems are dominated by a variety of intermolecular interactions of different strengths like hydrogen bonding, dispersion, electrostatics and induction. Traditional classical force field methods for studying the properties of these chemical systems lack the desirable accuracy for treatment of these different types of intermolcular interactions, while efficient treatment with electronic structure methods like second-order perturbative Moller-Plesset (MP2) and coupled cluster methods are unaffordable for these large chemical systems. Methods based on density functional theory (DFT) suffer from their inability to be systematically improvable. Hence, alternative methods are desirable for electronic structure quality predictions while being computationally affordable for these molecular crystals and clusters.

The Hybrid Many-Body Interaction (HMBI) method described in this dissertation has been developed for studying the properties of these molecular aggregates. In this method, the system is broken down into fragments and the most important short-range interactions are treated using highly accurate electronic structure methods while

the less important but more expensive interfragment interactions are treated using inexpensive classical force fields. Here, we demonstrate that the HMBI predictions are electronic structure quality while being computationally affordable. Moreover, these predictions can be systematically improved by use of more accurate electronic structure methods and force fields.

Here, the HMBI method has been employed in predicting the energetics and structure of molecular crystals and clusters. Some other capabilities of this method include prediction of the crystal structure in the presence of external stress, vibrational spectra, phonon dispersion curves, thermal properties like sublimation heats and specific heat capacities and elastic constants. We demonstrate that accurate HMBI predictions of these crystal properties allows for accurate identification and screening of different crystal polymorphs which is important in various applications of these materials.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Broad Overview of Computational Chemistry Method Development

Accurate property prediction for a variety of chemical systems which was impossible with experiments previously is now possible thanks to the improved analytical chemistry techniques and theoretical chemistry tools. In particular, understanding most experimental predictions with theory has now become a common practice in chemical research. Hence the role of sound theoretical methods has become crucial in assisting and sometimes even validating experiments. Quite obviously, theoretical method development in chemical research is an ever-growing field. Another hand-in-glove contribution to the rise of theoretical method development research in chemistry comes from the ever-improving computer technology which these theoretical chemistry tools can take full-fledged advantage of. This dissertation describes development of one such theoretical computational method to study a certain class of chemical systems, namely molecular clusters and crystals and to help predict their chemical properties.

Theoretical chemistry methods could be broadly classified into two categories, force field methods based on classical Newtonian mechanics and electronic structure methods based on quantum mechanics. Both these categories have their pros and cons. Force field methods are computationally inexpensive and hence can be easily used to study large chemical systems like biomolecules, crystals, etc. This is the primary reason for their sustained success. However, they often lack the desired accuracy. On the other hand, although electronic structure methods can be very accurate, their high computational cost renders them unaffordable for systems involving more than a few hundred electrons. This trade-off between computational cost and accuracy has led to the development of many hybrid methods which employ different levels of theory together for a particular system. This is usually achieved by breaking the system spatially into regions of more and less importance and then using high and low levels of theory respectively on these regions [2] or by using different levels of theory based according to the type of interaction of its fragments[3, 4, 5, 6, 7, 8, 9]. The fragment-based QM/MM Hybrid Many-Body Interaction (HMBI) method [9, 10, 11, 12, 13, 14, 15, 16] described in this dissertation falls in this latter category of hybrid computational chemistry methods. This method is targeted towards predicting properties of large chemical systems like molecular clusters and molecular crystals by using both electronic structure methods and force field methods in a symbiotic manner. The predicted properties include structural, thermodynamical, vibrational and mechano-elastic properties of these systems.

1.2 The System: Molecular Crystals and Clusters

Aggregates of molecules like molecular crystals and clusters play an important role in chemistry and find many important applications. For example, molecular

crystals find important applications as pharmaceutical drugs like aspirin and acetomophen, organic semiconductors like Rubrene [17] and explosives like RDX, FOX-7, etc. Other examples include interstitial ices which find significant interest in stellar chemistry for their role in assisting small molecule reactions on their surfaces. Naturally found clathrate-hydrates, made up of water host frameworks with trapped small organic molecules like methane and propane find interest as potential sources of fuel.

Molecular crystals and clusters could be categorized on the basis of the type of intermolecular forces dominating the packing. For example, crystals like ice or water clusters are dominated by strong hydrogen bonding while crystalline benzene is a dispersion-bound crystal. Due to the weak nature of the intermolecular forces, it is quite easy for these systems to deform and to exist in multiple polymorphic structures. For example, crystals like aspirin are known to exist as two polymorphs which lie within a few kJ/mol of each other in terms of lattice energies. Hence, while predicting the structure of these systems is often difficult experimentally, the knowledge of these polymorphs is very important for industrial applications. For example, a well-known HIV drug Ritonavir had to be pulled out of commercial market due to a post-process identification of another more stable polymorph which had poor bioavailability, costing the pharmaceutical company hundreds of millions of dollars [18, 19]. Potentially, theoretical tools could help identify the polymorphs and then predict their correct stability order and assist experiments for these applications. Hence, reliable theoretical chemistry methods should be able to correctly distinguish between these isoenergetic structures by using a balanced treatment of the different types of intramolecular and intermolecular forces like electrostatics, induction, dispersion, etc. Desirable theoretical methods should also be linear-scaling with system size so that they remain affordable when treating these large chemical systems. In the next section, we will briefly explore some of the compu-

tational chemistry methods widely used to study various chemical systems and also the approximations behind these methods.

1.3 Essential Pre-requisites

1.3.1 Born-Oppenheimer Approximation

According to the Born-Oppenheimer Approximation[20], since the electron is more than three orders of magnitude lighter than the lightest nuclei, the electrons move significantly faster than the nuclei. In other words, the electrons rearrange very quickly in response to a small perturbation in the nuclear arrangement. This means that for a particular nuclear arrangement, one can separately solve the electronic Schrodinger equation 1.5. Once this is done, one can perturb the nuclear arrangement and solve the new electronic Schrodinger equation and so on. Mathematically, this approximation is explained by the following set of equations, where $\mathbf{r}$ and $\mathbf{R}$ are the electronic and nuclear coordinates respectively. Schrodinger equation for the total system is given by

$$\mathbf{H_{total}}\psi_{\mathbf{total}}\left(\mathbf{r},\mathbf{R}\right) = \mathbf{E_{total}}\psi_{\mathbf{total}}\left(\mathbf{r},\mathbf{R}\right) \tag{1.1}$$

Born-Oppenheimer approximation allows the following decomposition of the total wave-function.

$$\psi_{\mathbf{total}}\left(\mathbf{r},\mathbf{R}\right) = \psi_{\mathbf{nuc}}\left(\mathbf{R}\right)\psi_{\mathbf{elec}}\left(\mathbf{r};\mathbf{R}\right) \tag{1.2}$$

The total Hamiltonian can now be given by the sum of the kinetic energy operator for the nuclei and the electronic Hamiltonian which depends parametrically on the nuclear positions.

$$\mathbf{H_{total}} = \mathbf{T_{nuc}} + \mathbf{H_{elec}} \tag{1.3}$$

where

$$\mathbf{H_{elec}} = \mathbf{T_{elec}} + \mathbf{V_{nuc-elec}} + \mathbf{V_{elec-elec}} + \mathbf{V_{nuc-nuc}} \tag{1.4}$$

The electronic energy is obtained by solving the electronic Schrodinger equation for a particular nuclear arrangement.

$$\mathbf{H_{elec}}\psi_{\mathbf{elec}} = \mathbf{E_{elec}}\psi_{\mathbf{elec}} \tag{1.5}$$

From equations 1.1, 1.2 and 1.3,

$$\left(\mathbf{T_{nuc}} + \mathbf{H_{elec}}\right)\psi_{\mathbf{elec}}\psi_{\mathbf{nuc}} = \mathbf{E_{total}}\psi_{\mathbf{elec}}\psi_{\mathbf{nuc}} \tag{1.6}$$

Using equations 1.5 and 1.6,

$$\left(\mathbf{T_{nuc}} + \mathbf{E_{elec}}\right)\psi_{\mathbf{elec}}\psi_{\mathbf{nuc}} = \mathbf{E_{total}}\psi_{\mathbf{elec}}\psi_{\mathbf{nuc}} \tag{1.7}$$

$$\left(\mathbf{T_{nuc}} + \mathbf{E_{elec}}\right)\psi_{\mathbf{nuc}} = \mathbf{E_{total}}\psi_{\mathbf{nuc}} \tag{1.8}$$

Clearly, the electronic energy acts as a potential in the nuclear Hamiltonian. Hence the set of electronic energies for different nuclei arrangements is called the Potential Energy Surface (PES) which can be visualized as a $3N$-dimensional surface on which the system with N-nuclei moves. The aim of most of computational chemistry methods is to map out the potential energy surface accurately. Many vibrational, thermal and elastic properties can be evaluated once the potential energy surface has been predicted. As a result, this approximation is the basic premise on which most computational chemistry methods are based upon.

1.3.2 Force Field Methods

Instead of solving the electronic Schrodinger's equation to obtain the potential energy surface, force field methods employ empirical or semi-empirical or fitted parameters to define the potential energy of the chemical system[21]. The bond lengths, bond

angles, dihedral angles, charges, multipoles, etc. are parametrized for a wide range of geometries for a chemical system using high level computations or experimental data. In short, the potential energy of a chemical system with a defined geometry is a function of these force field parameters for the equilibrium geometry for the chemical system. Typically, in a polarizable force field, the energy of the system is broken down into component energies coming out of geometric parameters, electrostatics, induction and dispersion.

$$E = E_{bond} + E_{angle} + E_{dihedral} + E_{electrostatic} + E_{induction} + E_{vdw} \tag{1.9}$$

The E_{bond} and the E_{angle} terms could be approximated using harmonic potential around the equilibrium geometries, so that

$$E_{bond} = \sum_{all\ bonds} \frac{1}{2} k_{bond} \left(R - R_0\right)^2 \tag{1.10}$$

$$E_{angle} = \sum_{all\ angles} \frac{1}{2} k_{angle} \left(\theta - \theta_0\right)^2 \tag{1.11}$$

However, the harmonic potential does not correctly predict the behaviour far away from equilibrium and hence these terms are sometimes approximated by Morse potential which has the form

$$E_{bond} = D\left(1 - e^{\alpha(R-R_0)}\right) \tag{1.12}$$

The $E_{dihedral}$ term usually are defined using periodic functions consistent with the periodicity of that dihedral, ω.

$$E_{dihedral} = \sum_{\omega} \sum_{n=0..} C_n cos(n\omega) \tag{1.13}$$

The E_{vdw} terms are usually described using functional forms similar to Lennard-Jones Potential given by

$$E_{vdw} = \sum_{i}^{atoms} \sum_{j}^{atoms} \left(\frac{C_1^{ij}}{R_{ij}^{12}} - \frac{C_2^{ij}}{R_{ij}^{6}} \right) \tag{1.14}$$

Permanent charges and multipoles are usually placed on atoms to describe the electrostatics. The $E_{electrostatic}$ term considers all the interactions like charge-charge, charge-multipole and multipole-multipole using functional forms similar to the Coulomb's interaction for charge-charge interaction.

$$E_{charge-charge} = \sum_{i} \sum_{j>i} \frac{Q_i Q_j}{\epsilon R_{ij}} \tag{1.15}$$

The induction terms usually require self-consistent solution of all the induced multipoles on each atom. Once the induced multipoles are known, the induction energy can be computed similarly to the electrostatic energy. Accurate treatment of electrostatics and induction energies is especially important for large chemical systems consisting of polar molecules. For example, in crystalline ice, the induction energy consists about 10-20% of the electrostatics.

The role of force field methods in studying large systems like reactions in solution, biomolecules like enzymes, DNA, etc. and crystals cannot be overemphasized. The quality of force fields can be improved by better parametrization or by getting rid of the parametrization altogether[12]. For example, ab-initio force fields which compute the force field parameters like distributed multipole moments and polarizabilties on the fly have become popular. To summarize, improved force field methods are still deemed to provide good balance between accuracy and cost for studying large chemical systems.

1.3.3 Electronic Structure Methods

The aim of the electronic structure methods is to accurately map out the potential energy surface of the chemical system using Quantum Mechanics. As already described, under the Born-Oppenheimer approximation, the problem of computing the electronic structure reduces to the problem of solving the electronic Hamiltonian in

the Schrodinger equation. However, except for the simplest one-electron Hydrogen atom-type system, this problem cannot be solved exactly due to the complex nature of electron-electron interaction. Hence, the Hamiltonian or the wavefunction needs to be approximated and different approaches exist with varying success in terms of accuracy and reliability.

Hartree-Fock theory is perhaps the most basic premise upon which electronic structure methods that approximate the wavefunction are built upon. It is a mean-field theory in which the wavefunction is described as a Slater determinant made up of "N" one-electron spin orbitals. The ground state wavefunction is the one which variationally minimizes the electronic energy. The use of a Slater determinant implicitly includes the Fermi correlation between the electrons arising due to the Pauli exclusion priciple. However, the Coulomb correlation is only included in a mean-field sense and is mostly missed by this theory. Most contemporary electronic structure methods are targeted towards getting this missing correlation between electrons. Even though the electronic correlation effects account for typically about 1% of the electronic energy, due to the need for accurate predictions the cost of recovering this bit is generally orders of magnitude higher than the cost expended on running a Hartree-Fock calculation itself.

In electronic structure methods, the correlation energy is defined as the contribution of the electronic energy the Hartree-Fock method fails to model (show equation). Various approaches have been developed to get more accurate predictions. For example, given the small correlation contribution to the electronic energy, one can use perturbation methods and treat the post-HF corrections to the energy and wavefunction as small perturbations. The Moller-Plesset perturbation theory (MP2)[22] is one such method in which the first order correction to the Hamiltonian leads to second order correction in energy. Even though quite successful, these methods could sometimes be erroneous

in that higher order perturbation terms could be important. Often successive perturbation term show oscillatory trend which could make these approaches non-convergent and unreliable for some system. Another marked feature for MP2 is its overestimation of correlation effects for systems showing significant pi-pi stacking, for example, MP2 with benzene. The basis set convergence issues of MP2 energies make these methods unreliable for dispersion-bound systems, though some methods like explicitly correlated MP2 methods[23, 24, 25, 26], spin-component-scaled MP2[27, 28, 29, 30, 31], etc. do address this issue to some extent. These methods scale as N^5 with system size which renders them useless for the treatment of large systems. Periodic MP2 methods have been recently implemented and have been used to study some small molecule crystals[32, 33] but these methods also suffer from the same drawbacks as non-periodic MP2.

The benchmark of electronic structure methods is usually considered to be coupled cluster methods[34] like CCSD(T)[35, 36]. In these methods, the excited state multi-electron Slater determinants are formed using the exponential coupled cluster operator on the ground state Hartree-Fock Slater determinant. These methods show better basis set convergence of the electronic energy compared to MP2. Unfortunately, CCSD(T) scales as N^7 with system size and is usually unaffordable for systems with over hundred electrons. Moreover, gradients and periodic implementation of coupled cluster methods are computationally very expensive to run, which makes them inapplicable for structure predictions and studying periodic systems like crystals.

Density functional theory (DFT)[37] developed on the theorems of Hohenburg, Kohn and Sham is probably the method of choice for studying small molecules and chemical reactions in the gas phase. This method is based on the theorem which says that the electronic energy of the system is a functional of its electronic density. However the exact form of this functional is unknown and hence approximate energy function-

als are used to varying success. Periodic DFT has also been implemented based on the Bloch's theorem. DFT methods however miss a huge chunk of the correlation energy. Many post-DFT empirical or semi-empirical dispersion-corrections[38, 39] have been implemented quite succesfully to capture the correlation energy but few problem cases have also been known. Also, dispersion corrections which work well with some functionals might not work well with others. A huge disadvantage of DFT methods lies in their inability to systematically improve their calculations. Examples have been known where a functionals and dispersion correction pair which works for a particular system has been shown to give erroneous results for another similar system. For example, benzene can be accurately treated with B3LYP-D but this functional overestimates the lattice parameter along the pi-stacking direction for diiodobenzene. Moreover, for some systems like glycine, the differences in the predicted energetics among different functionals exceeds the energy differences between competing structures, which makes it difficult to choose a reliable density functional and dispersion correction pair for a particular system. Also, these methods are not variational. Despite these problems, due to the relatively low cost of DFT methods compared to other wavefunction-based methods like MP2 and CCSD(T) and their good performance in general, DFT methods will have a dominant role to play in quantum chemistry. Nevertheless, alternative methods that are systematically improvable are desired.

1.3.4 Hybrid Methods

Hybrid methods are computational chemistry methods which target affordable treatment of large chemical systems for which electronic structure methods are otherwise computationally inaffordable. The main idea behind these methods is to use accurate electronic structure methods for the most important regions or interactions within

the chemical system while using lower-accuracy methods for less significant regions or interactions, thereby keeping the computational cost in check while being relatively accurate in its predictions. Popular hybrid methods like ONIOM[2] (Our own N-layered Integrated molecular Orbital and molecular Mechanics) decompose non-homogeneous system spatially into regions of varying signifance and then using different levels of theory for these different regions. The electronic energy for an ONIOM calculation is then given by

$$E_{large}^{ONIOM} = E_{large}^{Low} + E_{small}^{high} - E_{small}^{low} \tag{1.16}$$

This partitioning scheme works well for studying critical localized interactions in the system but is inapplicable for studying systems where these interactions are spatially homogeneous throughout the system as in the condensed-phase systems like molecular crystals. Clearly, for such latter systems alternative methods which partition the system based on the type of interactions are desired.

Fragment-based methods[3, 4, 5, 6, 7, 8, 9] based on many-body interaction (MBI) expansion 1.17 of the system on the other hand fragment the system spatially but then treat the various intra-fragment (1-body) and inter-fragment (2-body, 3-body and higher order) interactions with different levels of theory.

$$E_{full} = \sum_{i}^{fragments} E_i + \sum_{ij}^{dimers} \Delta^2 E_{ij} + \sum_{ijk}^{timers} \Delta^3 E_{ijk} + \cdots \tag{1.17}$$

where

$$\Delta^2 E_{ij} = E_{ij} - E_i - E_j \tag{1.18}$$

and

$$\Delta^3 E_{ijk} = E_{ijk} - \Delta^2 E_{ij} - \Delta^2 E_{jk} - \Delta^2 E_{ik} - E_i - E_j - E_k \tag{1.19}$$

and so on. In such fragment-based methods, usually the most important intra-fragment and pairwise interaction terms are treated with more accurate electronic structure meth-

ods like DFT, MP2, and CCSD(T), mainly because the cost for the treatment of trimers and higher order interaction terms increases significantly even for moderate sized fragments. For systems like molecular clusters of water molecules or ice crystals, the many-body induction contribution is significant. Similarly, for systems like benzene crystals the 3-body dispersion interactions play a significant role. Hence, the many-body terms cannot be neglected. The various fragment-based methods differ in the way they treat these many-body terms. For example, methods based on the Fragment molecular Orbital[3, 4] (FMO) approach like FMO2 incorporate the many-body electrostatics into the effective 1-body and 2-body terms using self-consistent electrostatic embedding. Similar methods like the Binary Interaction method[5] instead approximate the embedding potential with atomic point charges. On the other hand, the many-body interactions could be treated with lower level of theory like periodic HF or periodic DFT while treating the 1-body and 2-body terms with MP2 or coupled cluster methods.

Some of the features of these fragment-based methods are as follows. Predictions with these methods can be highly accurate, similar to electronic structure methods in quality. These can be made linear-scaling with system size thereby making them affordable for treatment of large systems. Calculations with these methods are usually easy to parallelize onto multiple processors. More importantly, predictions with these methods can be systematically improved by use of highly accurate electronic structure methods on the fragments or by explicitly calculating the higher order terms like 3-body with higher level of theory. The derivatives of the electronic energy with respect to the various degrees of freedom which are important for structure prediction are usually straightforward, though the use of embedding potential or charges complicates these terms.

The topic of this dissertation, the Hybrid Many-Body Interaction (HMBI)

method which will be formally introduced in the next chapter is another fragment-based method that treats the important 1-body and short-range 2-body interaction terms with high level electronic structure methods while approximating the long-range 2-body interaction terms and higher order terms in the many-body interaction expansion with inexpensive polarizable force fields.

1.4 Outline of the Dissertation

In the next chapter, the basic energy formalism of the HMBI method for molecular clusters and molecular crystals will be showcased and how the spatial truncation scheme can make it linear-scaling with system size will be discussed. Also, we explore some of the advantages of this method over other methods like the periodic density functional theory (pDFT) and force field methods. Finally, we show how HMBI can accurately predict lattice energies on test set crystals and then systematically improve these results. Results in section 2.2 have been adapted from "Predicting organic crystal lattice energies with chemical accuracy", *J. Phys. Chem. Lett.*, 1:34803487, 2010.

The text of Chapter 3, in part or in full, is a reprint of the material as is appears in "Structures and energetics of electrosprayed uracil$_n$Ca^{2+} clusters (n=14-4) in the gas phase", *Phys. Chem. Chem. Phys.*, 14:3304-3315, 2012. Prof. G. J. O. Beran listed in that publication directed and supervised the research with HMBI which forms the basis for this chapter. Here, we explore how the HMBI method can help explain the oscillatory dissociation energy trend of the (uracil)$_n$Ca^{2+} (n=4,..,14) clusters observed with temperature dependent blackbody infrared radiative dissociation (BIRD) experiments. This study was done in collaboration with Elizabeth A.L. Gillis and Prof. Travis D. Fridgen at Department of Chemistry, Memorial University of Newfoundland,

who performed the BIRD experiments and Maria Demireva and Prof. Evan. R. Williams at Department of Chemistry, University of California, Berkeley, who performed the master equation modeling of the kinetics.

The text of Chapter 4, in part or in full, is a reprint of the material as is appears in "Prediction of organic molecular crystal geometries from MP2-level fragment quantum mechanical/molecular mechanical calculations", *J. Chem. Phys.*, 137:174106, 2012. Prof. G. J. O. Beran listed in that publication directed and supervised the research which forms the basis for this chapter. Chapter 4 focusses on the full crystal structure prediction of molecular crystals. The formalism for the HMBI lattice gradients is also presented. Performance of the HMBI method on test set molecular crystals is compared against that of force field and periodic DFT methods. It is shown that the HMBI optimized nuclear coordinates and lattice parameters lie within a few percent and the HMBI optimized lattice energies lie within a few kJ/mol of the experimental values.

In Chapter 5, we explore how lattice dynamics can be incorporated with HMBI. We also discuss how HMBI with lattice dynamics can accurately predict vibrations and thermal and elastic properties of test set molecular crystals. The fact that lattice dynamics calculations are quite inexpensive with HMBI will be emphasized.

HMBI stability order predictions on the various polymorphs of ice XV consistent with the experiments but predicted incorrectly with DFT calculations previously will be discussed in Chapter 6. Systematic improvement of the lattice energies of these polymorphs is discussed and energy breakdowns that help explain the stability order trend will be ellaborated.

Chapter 2

Hybrid Many-Body Interaction Method

2.1 Formalism

The Hybrid Many-Body Interaction (HMBI) method is a fragment-based quantum mechanical (QM)/ molecular mechanical (MM) method founded on the many-body interaction expansion discussed in the previous chapter. The main goal with this method is predict properties of aggregates of molecules like molecular crystals and clusters with electronic-structure accuracy while being computationally affordable. The many-body interaction expansion is given by

$$\begin{aligned} E^{MBI} &= E_{1-Body} + E_{2-Body} + E_{3-Body} + \cdots \\ &= E_{1-Body} + E_{2-Body} + E_{Many-Body} \end{aligned} \tag{2.1}$$

In the HMBI method, the 1-body and short-range 2-body interactions are treated with highly accurate electronic structure methods like MP2[22] and CCSD(T)[35, 36] while less important but the non-negligible long-range 2-body interactions and all higher order

many-body interactions are treated using classical polarizable force fields.

$$E^{HMBI} = E^{QM}_{1-Body} + E^{QM}_{short-range\ 2-Body} + E^{MM}_{long-range\ 2-Body} + E^{MM}_{Many-Body} \tag{2.2}$$

Using the many-body interaction expansion, the $E^{MM}_{Many-Body}$ can rewritten as

$$E^{MM}_{Many-Body} = E^{MM}_{full} - E^{MM}_{1-Body} - E^{MM}_{short-range\ 2-body} - E^{MM}_{long-range\ 2-body} \tag{2.3}$$

Plugging the above expression into the HMBI expression, we get

$$\begin{aligned} E^{HMBI} &= E^{QM}_{1-Body} + E^{QM}_{short-range\ 2-Body} + E^{MM}_{long-range\ 2-Body} \\ &\quad + E^{MM}_{full} - E^{MM}_{1-Body} - E^{MM}_{short-range\ 2-body} - E^{MM}_{long-range\ 2-body} \\ &= E^{MM}_{full} + \left(E^{QM}_{1-Body} - E^{MM}_{1-Body}\right) + \left(E^{QM}_{short-range\ 2-Body} - E^{MM}_{short-range\ 2-body}\right) \\ &= E^{MM}_{full} + \sum_{i}^{monomers} \left(E^{QM}_{i} - E^{MM}_{i}\right) + \sum_{ij}^{short-range\ dimers} \left(\Delta^2 E^{QM}_{ij} - \Delta^2 E^{MM}_{ij}\right) \end{aligned} \tag{2.4}$$

To partition the pairwise interactions of the fragments into short-range and long-range, we introduce a spatial truncation damping function[40], d_{ij}, employing two spatial cut-offs, c_1 and c_0 and the shortest distance R between the monomers defining the dimer.

$$d_{ij}(R) = \begin{cases} 1 & \text{if } R \leq r_1 \\ \frac{1}{1+e^{2|c_1-c_0|/(c_1-R)-|c_1-c_0|/(R-c_0)}} & \text{if } c_1 < R < c_0 \\ 0 & \text{if } R \geq c_0 \end{cases} \tag{2.5}$$

This means that if the monomers are within a cutoff distance c_1, then the pairwise interaction should be treated with QM. If the shortest distance between the monomers is in between c_1 and c_0, then the pairwise interaction needs to be interpolated as a linear combination of quantum and classical interactions. If the shortest distance is greater than c_0, then the interaction is supposed to be treated classically. Such a damping function ensures that there are no discontinuities on the potential energy surface and that

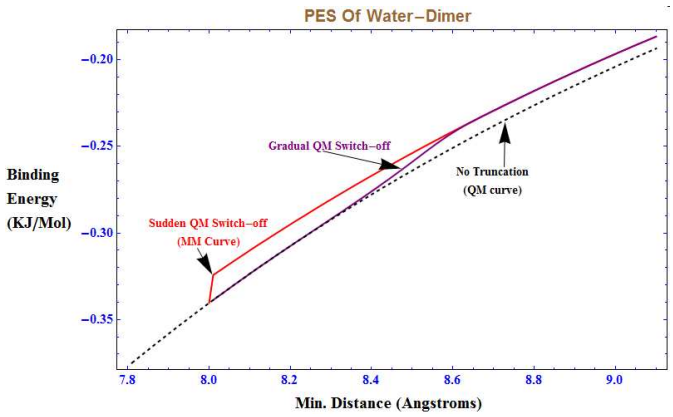


Figure 2.1: Gradual switch-off from the QM PES to the MM PES due to the spatial damping function as a function of shortest intermolecular distance for a water dimer

the transition from the short-range quantum treatment to long-range classical treatment is smooth (figure 2.1). The damping function introduces errors which are less than 0.1 kJ/mol/monomer. The HMBI energy expression for a general system (periodic or non-periodic) becomes

$$E^{HMBI} = E^{MM}_{PBC} + \sum_{i} \left(E^{QM}_{i} - E^{MM}_{i} \right) + \sum_{ij(0)} d_{ij(0)} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(0)} \right) + \frac{1}{2} \sum_{i} \sum_{j(n_v)}^{images} d_{ij(n_v)} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(n_v)} \right) \tag{2.6}$$

where the last summation should not be included for a non-periodic system. Here the first summation is over all the monomers within the central unit cell, the second summation is over all the dimers within the central unit cell and last summation is over all the short-range dimers that can be formed for the monomers within the central unit cell (index = 0) and the monomers in the image unit cells (index = n_v). The factor of 1/2 is included to avoid double counting.

The benefits of using the force fields are many-fold. Firstly, these polarizable force fields self-consistently capture the induction effects which are important espe-

cially for aggregates with polar molecules or hydrogen bonding. Secondly, for periodic implementations, the inexpensive force fields replace the expensive periodic electronic structure methods. The periodic electronic structure methods are expensive due to the slow 1/r decay of the electrostatic interactions which can now be replaced by trivial Ewald sums. Thirdly, one can use any off-the-shelf polarizable force fields or one can get force field parameters on-the-fly in an ab-initio fashion. Finally, the use of force field for treating the long-range pairwise interactions is a good approximation as the electronic energies with the force field and the electronic structure tend to converge to similar values in the long-range. This reduces the computational cost significantly.

The HMBI implementation has the following advantages over the electronic structure and other fragment-based methods for studying large molecular aggregates. Firstly, one could use very expensive, accurate methods like coupled cluster methods for the most important interactions. The spatial partitioning of the pairwise (2-body) interaction terms into short-range and long-range in fact makes the method linear scaling with system size. Secondly, no embedding potential is used which makes the calculations of the gradients and hessians of the electronic energy straightforward, which is important for structure prediction. The one disadvantage of the HMBI scheme is that it is not applicable when the fragments are covalently bonded to each other, for example, large biomolecules like DNA and enzymes.

The following discussion covers the previous performance of HMBI in calculating the lattice energies for a test set of crystals. The systematic improvement of the predictions is also emphasized. HMBI can predict lattice energies within chemical accuracy was confirmed. This work acted as a prelude to the work presented in the subsequent chapters.

2.2 Previous Performance of HMBI: Accurate Prediction of Lattice Energies of Molecular Crystals

In this study adapted from Ref. [11], the HMBI method was used to predict the lattice energies of five crystals, namely, hexagonal (I_h) ice, formamide, acetamide, benzene and imidazole. This set of crystals was selected because of the different types of intermolecular forces present in each. For example, the hydrogen bonding or induction dominates the many-body effects in ice, acetamide and formamide, $\pi-\pi$ stacking dispersion interactions dominate the benzene crystal while imidazole displays both dispersion and hydrogen bonding interactions. Thereby, the premise that a reliable theoretical method should be able to give accurate and balanced treatment of different kinds of interactions in the chemical systems was tested for HMBI with this test set of molecular crystals. Moreover, the availability of the experimental data on sublimation heats and the force field parameters with the polarizable AMOEBA[41] force field employed here, were other reasons for this choice of molecular crystals.

In this study, the experimental crystal structures served as initial guess structures. Then the nuclear degrees of freedom were optimized with the lattice parameters frozen at the experimental values. For these geometry optimizations, the Resolution-of-the-Identity second order Moller-Plesset perturbation (RI-MP2)[42, 43, 44] level of theory was employed with the Dunning aug-cc-pVDZ basis set[45, 46] for the quantum treatment while the AMOEBA force field was employed for the classical treatment. The quantum calculations were done using Q-Chem[47] while the classical calculations were done using the Tinker package[48]. Geometry optimizations were done using the open-source DL-FIND optimizer[49]. Using these optimized structures, the HMBI single points were improved by first extrapolating to the complete basis set (CBS) limit using

dual-basis[50] RI-MP2 level of theory and the Dunning aug-cc-pVTZ and aug-cc-pVQZ basis sets. Then, single point energies were calculated for these optimized structures using the more accurate but expensive CCSD(T) level of theory with the PSI3 package[51]. For ice, aug-cc-pVTZ basis set was used, for formamide and imidazole, aug-cc-pVDZ basis set was used and for benzene and acetamide, 6-31+G* basis set was used with CCSD(T). For the respective crystal and basis set, the level of theory correction was estimated using the formula, $\Delta^{CCSD(T)} = E^{CCSD(T)} - E^{MP2}$. These level of theory corrections especially improve the predictions when many-body dispersion effects are strong in the crystal because MP2 lacks description for the many-body dispersion effects. Finally, a crude correction for the error introduced due to frozen lattice parameters during the geometry optimization was estimated using the following algorithm. First, the lattice parameters were isotropically changed by 1%. Second, the nuclear coordinates were optimized with planewave DFT with PBE functional using the Dacapo software. Third, the potential energy surface was scanned for each set of nuclear and lattice parameters by finding the single points with HMBI DB-RI-MP2/aug-cc-pVDZ:AMOEBA scheme. Finally, cubic splines were used to identify the optimal structure and to estimate these corrections in the lattice energies.

Comparison of lattice energies with experiments is not possible as experiments are done at finite temperatures while the HMBI calculations are done at 0 K. However, the sublimation heats, ΔH_{sub} of these crystals were turned into the lattice energies $U_{lattice}$ using the formula

$$U_{lattice} = \Delta H_{sub}\,(T) + 2RT \tag{2.7}$$

Meanwhile, the experimental zero point energies were estimated using the IR frequencies.

Table 2.1 clearly shows how the lattice energy predictions with HMBI can

Table 2.1: HMBI-Predicted Crystal Lattice Energies (kJ/mol)

QM level[a]	ice	form--amide	acet--amide	imid--azole	benzene
force-field contribution[b]	13.9	-5.4	3.0	4.9	-2.0
DB-RI-MP2/aug-cc-pVDZ	53.2	67.5	74.1	95.3	62.2
DB-RI-MP2/aug-cc-pVTZ	57.1	72.3	78.5	99.1	62.1
DB-RI-MP2/aug-cc-pVQZ	58.7	74.1	80.3	99.7	62.3
DB-RI-MP2/CBS	60.3	76.0	81.7	101.7	63.1
error vs expt.	1	-6	-4	15	11
$\Delta^{CCSD(T)}$[c]	0.4	1.8	-0.1	-14.2	-10.4
DB-RI-MP2/CBS + $\Delta^{CCSD(T)}$	60.6	77.8	81.6	87.5	52.7
error vs expt.	2	-4	-4	-4	1
est. lattice param. relax., $\Delta E^{relax}_{lattice}$	0.0	0.1	n/a	0.3	3.7
best estimate[d]	60.6	77.9	81.6	87.8	56.4
error vs expt.	2	-4	-4	-3	4
experiment[e]	59	82 ± 3	86 ± 2	91 ± 4	52 ± 3

[a]Counterpoise corrected. [b]Classical polarization contribution. It is already included in the reported lattice energies. [c]Post-MP2 correction, $\Delta^{CCSD(T)} = E^{CCSD(T)}_{lattice} - E^{MP2}_{lattice}$, using the basis sets described in the text. [d]Best estimate = $E^{DB-RI-MP2/CBS} + \Delta^{CCSD(T)} + \Delta E^{relax}_{lattice}$. [e]Reported errors are the standard deviation among the set of extrapolated 0 K lattice energies. Actual experimental errors may be larger.

be systematically improved by first extrapolating to complete basis set limit and then improving the level of theory correction and finally accounting for the relaxation of the structure. These results compare favorably with other high-level benchmark calculations on these crystals. For these five crystals, the predicted lattice energies are within a kcal/mol of the experiments. These errors arise due to the inherent errors with AMOEBA force field and the use of small basis set with CCSD(T) while calculating the level of theory corrections. AMOEBA lacks the many-body dispersion treatment which is significant in dispersion dominated crystals like benzene and imidazole. Moreover, it also overestimate the many-body effects in ice and underestimates them for formamide.

2.3 Summary

In this chapter, the basic energy formalism of the HMBI method was presented. Some advantages of such a fragment-based QM/MM method over conventional methods like periodic DFT for treatment of molecular aggregates have been emphasized. Apart from being trivially parallelizable, this method does not partition these homogenous systems spatially, rather the most important interactions are dealt with highly accurate electronic structure methods while the less important long-range interactions are treated with polarizable force fields. More importantly, it was shown that the main advantage of this method is that the lattice energy predictions can be systematically improved by use of better force fields and levels of theory. Highly accurate benchmark lattice energies predicted with this method were within a kcal/mol of the experiments.

The work shown in this chapter acts as the starting point on which the subsequent chapters in this dissertation are based upon. In the next chapter, the success of this method in identifying the correct gas-phase dissociation energies of the $\text{Uracil}_n\text{Ca}^{2+}$

(n=4-14) clusters will be demonstrated. In a further attempt to get rid of the rather crude relaxation correction in this chapter and allow for complete crystal structure prediction, chapter 4 will illustrate the nuclear and lattice gradients implementation for HMBI and show some benchmark results. Furthermore, chapter 5 will illustrate how HMBI can predict measurable crystal properties like sublimation heats at finite temperature for direct comparison against experimental data rather than the crude comparisons of the experimental and predicted lattice energies at 0 K, which has been used here.

Chapter 3

Structures and Energetics of Electrosprayed Uracil$_n$Ca^{2+} Clusters (n = 14-4) in the Gas Phase

3.1 Outline

Clusters of uracil (U) about a calcium dication, U$_n$Ca^{2+} (n = 14-4), have been studied in the gas phase by both experimental and theoretical methods. Temperature dependent blackbody infrared radiative dissociation (BIRD) experiments were performed on U$_n$Ca^{2+} clusters with n = 14-5 and the observed Arrhenius parameters are reported here. Master equation modeling of the BIRD kinetics data was carried out to determine threshold dissociation energies. Initial geometry calculations were performed using the B3LYP density functional and 3-21G(d) basis set. A sample of ten conformations

per cluster was obtained through a simulated annealing study. These structures were optimized using B3LYP/6-31G(d) level of theory. Fragment-based hybrid many body interaction (HMBI) MP2/6-311++G(2df,2p)/Amoeba calculations were performed on representative conformations to determine theoretical binding energies. Results were examined in relation to cluster size (n). A significant increase in the energy required to remove uracil from U_6Ca^{2+} when compared to larger clusters supports previous reports that the calcium ion is coordinated by six uracil molecules in the formation of an inner shell. For clusters larger than n = 6, an odd-even alternation in threshold dissociation energies was observed, suggesting that the outer shell uracil molecules bind as dimers to the inner core. Proposed binding schemes are presented. Multiple structures of U_5Ca^{2+} are suggested as being present in the gas phase where the fifth uracil may be either part of the first or second solvation shell.

3.2 Introduction

Adenine, guanine, cytosine, thymine, and uracil nucleobases constitute one of the major components of DNA/RNA molecules. These bases exhibit a natural tendency to bind together through non-covalent interactions such as hydrogen bonding. Base pairing plays a crucial role in the structure and thus the function of DNA. While this aggregation has obvious implications in biological systems, it has also gained interest in the field of supramolecular chemistry where physical manipulation of molecules is limited by the small scale in which work can be performed[52]. The construction of assemblies including grid, closed, or ribbon structures relies on molecular recognition and self-assembly processes such as those which occur between nucelobases[53, 54]. While hydrogen bonding is an important aspect in developing these different structures, metal

ions often initiate and direct self-assembly of the bases and stabilize complexes through ion-dipole interactions.

The self-assembly of nucleobases into larger structures also has biological relevance in terms of telomere research. As telomere DNA ends are characterized by tracts of guanine residues, a quartet of bases (G-quartet) is known to form and has become the focus of considerable interest due to its relevance in cancer research[54, 55, 56, 57, 58, 59, 60, 61]. While not to the same extent as the G-quartet, the observation of a U-quartet[62] has opened the door for similar tetrameric structures to be examined alongside the well-established guanine structure[55, 63, 64, 65, 66, 67]. This has led to a broader examination of the role the metal cation plays in terms of both size and charge in the formation of nucleobase clusters. In work by Koch et al.[64], clusters of nucleobases with alkali metal ions not only demonstrated the stability of the $[U_4Na]^+$ quartet but also that other ions stabilize clusters of various sizes, e.g. $[U_5K]^+$. This is consistent with recent work where uracil and thymine were found to form quintet structures in the presence of K^+, Rb^+, and Cs^+ cations[68].

One feature of the reported clustering of nucleobases with metal ions was the observation of larger clusters of up to fourteen bases per ion[64, 69]. Larger clusters composed of uracil were more recently examined by Zins et al. where structures were suggested based on collision induced dissociation (CID) experiments and ab initio (HF/STO-3G) and DFT (BLYP/3-21G(d)) calculations[70]. The structures involving the doubly charged calcium cation were more closely examined[69], and it was suggested that the complexes consisted of an inner core of six uracil molecules about Ca^{2+}, with larger clusters being the result of uracil binding with the inner core through hydrogen bonding interactions.

The aim of this study was to further examine the larger clusters of uracil with

Ca^{2+} using Fourier transform ion cyclotron resonance tandem mass spectrometry (FT-ICR MS/MS) to more definitively characterize their structures. To accomplish this, black- body infrared radiative dissociation (BIRD)[71, 72] and infrared multi-photon dissociation (IRMPD) are used. Temperature-dependent BIRD experiments are performed to determine Arrhenius parameters and extract threshold dissociation energies using master equation modeling. These are then interpreted on the basis of cluster ion structure and size. IRMPD is used to study the dissociation kinetics for several uracil clusters in order to obtain information about the potential presence of multiple conformers[73]. Electronic structure calculations are also performed to determine possible cluster structures and to rationalize the observed kinetics.

3.3 Methods

3.3.1 Experimental

All experiments were performed at Memorial University of Newfoundland using an ApexQe Bruker FT-ICR MS. Uracil and calcium chloride were purchased from Sigma-Aldrich and BDH chemicals, respectively, and used without further purification. A solution containing 2.5 mM uracil in a 50/50 mixture of 18 MO Millipore water and acetonitrile, to which a few drops of 0.01 M $CaCl_2$ (aq) were added, was electrosprayed at a flow rate of 112 mL h^{-1} to generate U_nCa^{2+} complexes with n = 4-14. Heating of the ICR cell was achieved through the use of a heating jacket which was placed around the flight tube, extending from the ICR cell to the gate valve. Temperature calibration was performed by placing a J-type thermocouple in the center of the ICR cell and measuring the internal and external temperatures after heating. This was performed over a two week period. BIRD studies were performed by isolating a single complex in the ICR cell

and monitoring the change in abundance of precursor and fragment ions as a function of time for a given temperature. This was repeated for each cluster with the exception of $U_{13}Ca^{2+}$ which could not be obtained with appreciable intensity for isolation. In this case, kinetics data were measured for the dissociation of $U_{13}Ca^{2+}$ produced from the dissociation of $U_{14}Ca^{2+}$. Measurements were obtained over a 293-380 K temperature range.

The BIRD rate constants (k) were determined by fitting the observed first-order kinetics to

$$[I_m]_t = [I_m]_0 e^{-kt} \tag{3.1}$$

where $[I_m]_0$ is the normalized intensity of cluster m at $t = 0$, and $[I_m]_t$ is the intensity of this same cluster at reaction delay time, t. The BIRD rate constants were then plotted as a function of inverse temperature (i.e. Arrhenius plot) and the Arrhenius parameters for dissociation were obtained using

$$lnk = lnA - \frac{E_a}{k_B T} \tag{3.2}$$

From a linear fit to eqn (2), the observed pre-exponential factor (A) and activation energy, E_a, were obtained from the intercept and slope, respectively.

The IRMPD kinetics experiments were performed on clusters with n = 5, 6, 8, and 10. Clusters were isolated in the ICR cell and were subsequently irradiated with a 25 W CO2 laser to induce dissociation. The change in intensity of reactant and products was monitored as a function of irradiation time. The laser power was set at 20% and sixteen scans were averaged. These IRMPD experiments were performed at room temperature.

3.3.2 Computational

3.3.2.1 Geometry and Binding Energies.

Simulated annealing and HMBI calculations were performed at UC Riverside. Initial geometry optimizations and frequency calculations were performed using the B3LYP density functional and 3-21G(d) basis set with the Gaussian 09 suite of programs[71, 72]. Given the wide range of possible conformations in these U_nCa^{2+} clusters, an additional sample of 10 or more con- formations for clusters with n = 4-9 were generated by simulated annealing using the Amoeba force field[41] and the Tinker software package[48]. These structures were subse- quently optimized with B3LYP/6-31G(d) using the Q-Chem 3.1 electronic structure package[47].

For each cluster size, representative low-energy conformations were selected for higher-level single-point energy calculations using the recently developed hybrid many-body interaction (HMBI) model[9, 10, 11, 12]. Briefly, the HMBI model is a fragment approach that decomposes U_nCa^{2+} clusters into a system of interacting mono- mers, i.e. individual uracil molecules and the calcium cation. The monomer energies and the pairwise interaction energies of these individual monomers are computed using quantum mechanics (QM), while the less important but non-negligible 3-body and higher order interaction energies are computed using a polarizable molecular mechanics (MM) force field.

For this study, the QM terms were calculated using MP2 level of theory and 6-311++G(2df,2p) basis set using Q-Chem. MP2 theory was selected because it captures potentially important van der Waals dispersion interactions in these pi-stacking systems. All QM pairwise interaction energies were corrected for basis set superposition error (BSSE) using the standard Counterpoise correction. The MM energies were com-

puted using the Amoeba polarizable force field as implemented in the Tinker software package. Conventional MP2 calculations on the largest cluster at this level of theory would involve over 4400 basis functions and would be utterly cost prohibitive with conventional algorithms. The HMBI approach drastically reduces the computational cost while providing results that are very close to fully QM calculations[10, 12]. All reported electronic energies were calculated at this level of theory.

3.3.2.2 Master Equation Modeling.

Master equation modeling of the BIRD data was performed at Berkeley. Arrhenius parameters (E_a, A) obtained from temperature dependent BIRD experiments are not necessarily equal to those in the limit of high pressure (E_a^∞, A^∞). These values become equal when the rate of energy exchange with the blackbody field (i.e. absorption and emission rates) of the ion population is significantly faster than the dissociation rate. In this rapid energy exchange (REX) limit, the internal energy distribution of the ion population is characterized by a Boltzmann at the temperature of the black- body field, and threshold dissociation energies, E_0, can easily be obtained from the measured activation energies[72]. However, for ions with dissociation rates similar to or faster than the rate of energy exchange with the blackbody field, E_0 values must be obtained by master equation modeling[74, 75, 76] or under some conditions using a truncated Boltzmann model[74]. A more detailed discussion of all these cases is given elsewhere[71, 72, 76]. Master equation modeling was performed on all clusters.

A matrix formalism[76] of the master equation was used. Briefly, the matrix contains the detailed rate constants for absorption, emission, and dissociation which give the energy transfer probability from one energy state to another. The master equation model was used to simulate BIRD for U_nCa^{2+} (n = 5-14) by reacting a Boltzmann

population at a given temperature and allowing this internal energy distribution to reach a steady-state, i.e. where linear BIRD kinetics would be observed. The modeled first order BIRD rate constant was obtained from the slope of the graph $ln([M]_t/[M]_0)$ vs. reaction time, where $[M]_t$ is the population remaining at reaction time t, and $[M]_0$ is the initial population. BIRD rate constants modeled for the highest and lowest temperatures in each experiment were used to construct Arrhenius plots. A range of E_0 values for U_nCa^{2+} with n = 5-14 were then extracted by fitting the experimental data with modeled Arrhenius data. A good fit to the experimental Arrhenius data needed to fulfill two requirements: (1) the modeled E_a value was within one standard deviation of the experimental Ea value and (2) the modeled BIRD rate constants at the highest and lowest temperatures of the Arrhenius plot were within a factor of two of the experimental values.

To account for uncertainties in the modeling parameters, transition state frequencies were varied to produce pre-exponential factors corresponding to a neutral to loose transition state, $log(A) \approx 14.5$ and 19.5, respectively. The integrated IR intensities are important when considering the rate of energy exchange between the ion and the blackbody radiation field. In this study, the integrated IR intensities were multiplied by 0.5 and 2 to account for any uncertainties in these calculated values. An extensive search for E_0 was performed over the 0.5-1.5 eV range (depending on cluster size) in 0.1 eV increments. For each set of E_0 and $log(A)$ parameters, a value of E_a was obtained from the modeled Arrhenius plot and compared to the experimental value; the squared difference between the calculated and experimental E_a was plotted as a function of both E_0 and $log(A)$. The calculated BIRD rate constants at the highest and lowest temperatures were compared to their respective experimental values by recording the sum of squared differences as a function of E_0 and $log(A)$. The range of E_0 values was

determined from the overlap of the two graphs satisfying both requirements for fitting the experimental Ea value and BIRD rate constants.

3.4 Results and Discussion

3.4.1 BIRD

Upon electrospray, 4-14 uracil molecules were consistently observed to complex with Ca^{2+}. However, complexes containing 2-3 uracil molecules could also be formed by changing experimental conditions including concentration and electrospray settings. A typical mass spectrum is shown in figure 3.1. There are variations in the relative intensities of the cluster ions, which suggest the possibility of magic number clusters, i.e. complexes which are notably abundant in relation to other ions present. For a given cluster m, a magic number cluster can be defined as having the ratio of intensities $I_m^2/(I_{m+1} \times I_{m-1})$ greater than one[77]. Using this definition, clusters with even numbers of uracil molecules (n = 14, 12, 10, 8, and 6) could consistently be assigned as magic number clusters.

The clusters were observed to dissociate through the sequential loss of a neutral uracil such that

$$U_nCa^{2+} \xrightarrow{k_n} U_{n-1}Ca^{2+} \xrightarrow{k_{n-1}} U_{n-2}Ca^{2+} \xrightarrow{k_{n-2}} \cdots U_4Ca^{2+} \tag{3.3}$$

This was observed for all complexes with the exception of U_4Ca^{2+} which dissociated into protonated uracil (UH+) and the deprotonated trimer, $[U_3\text{-H}]Ca^{+}$. BIRD studies on U_4Ca^{2+} were not feasible because the timescale for unimolecular dissociation activated by ambient blackbody radiation (i.e. BIRD) was far too long even at the highest temperatures. As a result, BIRD kinetics studies were not performed on the U_4Ca^{2+}

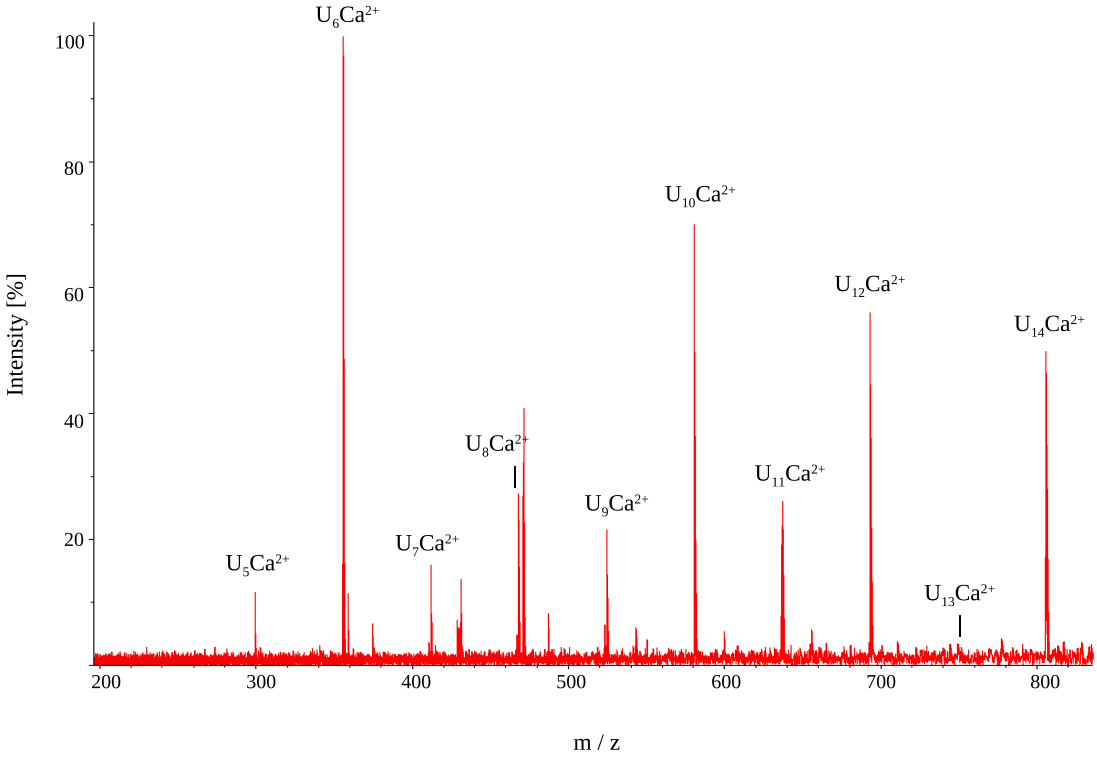


Figure 3.1: Representative mass spectrum of U_nCa^{2+} with n = 5-14

complex. U_5Ca^{2+} and U_6Ca^{2+} were also somewhat affected by this due to significantly slower dissociation rates at the lowest temperatures. Therefore, BIRD kinetics data were obtained at temperatures higher than 303 K and 315 K for U_5Ca^{2+} and U_6Ca^{2+}, respectively. It should also be noted that the high rate of dissociation for $U_{13}Ca^{2+}$ hindered attempts to isolate the cluster with sufficient intensity for reliable BIRD experiments. Kinetics data for $U_{13}Ca^{2+}$ were obtained by first dissociating $U_{14}Ca^{2+}$ into $U_{13}Ca^{2+}$, which subsequently underwent BIRD to form $U_{12}Ca^{2+}$. For simplicity, it is assumed that the structure of $U_{13}Ca^{2+}$ obtained from the dissociation of $U_{14}Ca^{2+}$ is similar to an isolated $U_{13}Ca^{2+}$ cluster. While this can not be confirmed, it is supported by both forms of $U_{13}Ca^{2+}$ undergoing very quick dissociation when compared to the other sized clusters. Rate constants are listed in table 3.1. Sample BIRD kinetics plots for $U_{14}Ca^{2+}$ over the temperature range 307-380 K are presented in figure 3.2.

Arrhenius plots for all clusters are presented in figure 3.3. The linear fits to the Arrhenius data for clusters with n = 6-14 are very good with R^2 values from linear regression ranging from 0.994 to 0.976. However, the data points for U_5Ca^{2+} have considerable scatter resulting in an $R^2 = 0.858$. Experimental activation energies, E_a^{obs}, and pre-exponential factors, A^{obs}, obtained from the Arrhenius plots are listed in table 3.2.

To investigate the relationship between dissociation energy and possible structural features, the E_a^{obs} values and the corresponding E_0 values obtained from master equation modeling are plotted as a function of cluster size (n) and are shown in figure 3.4. There is a significant decrease in the activation and threshold dissociation energy for clusters containing more than six uracil molecules compared to the values

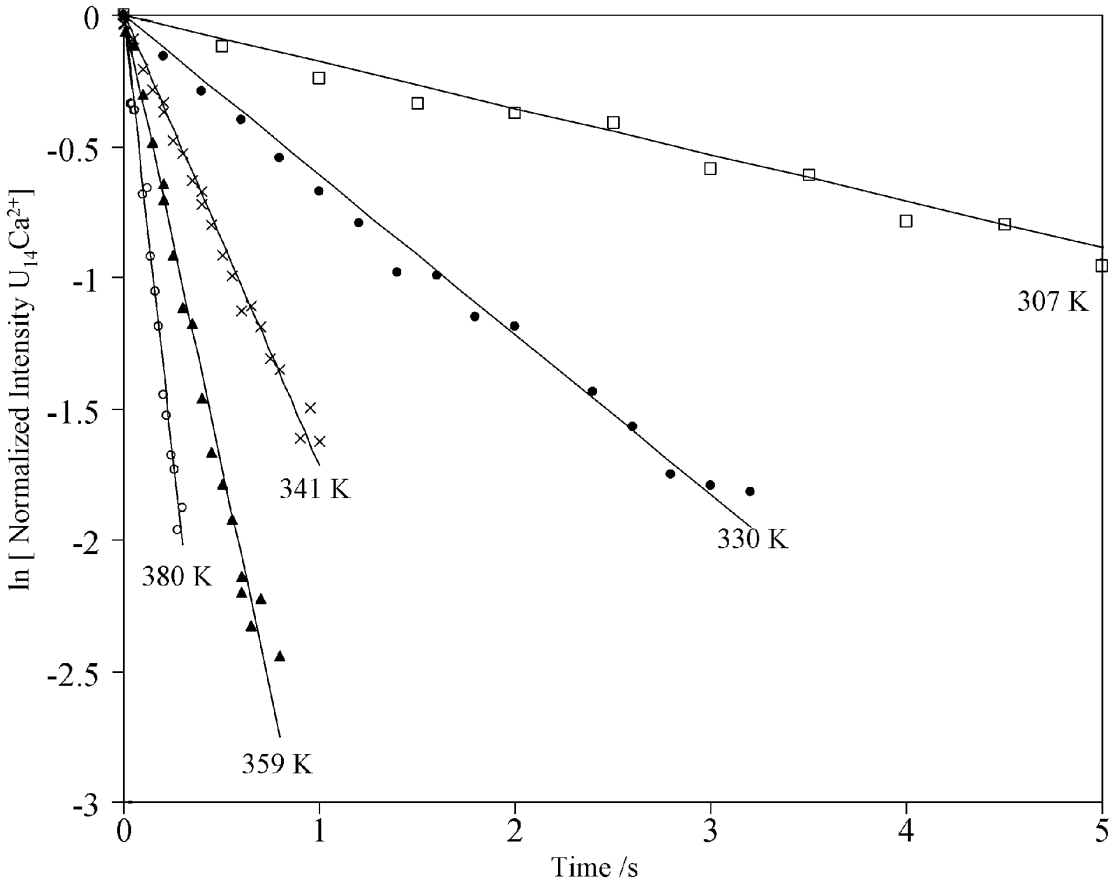


Figure 3.2: Experimental BIRD dissociation kinetics for $U_{14}Ca^{2+}$ at temperatures indicated.

	k/s^{-1}									
T/K	$U_{14}Ca^{2+}$	$U_{13}Ca^{2+}$	$U_{12}Ca^{2+}$	$U_{11}Ca^{2+}$	$U_{10}Ca^{2+}$	$U_{9}Ca^{2+}$	$U_{8}Ca^{2+}$	$U_{7}Ca^{2+}$	$U_{6}Ca^{2+}$	$U_{5}Ca^{2+}$
293	-	1.681	0.0613	0.356	0.0956	0.195	0.0894	0.131	-	-
295	0.1142	-	-	-	-	-	-	-	-	-
303	-	-	-	-	-	-	-	-	-	0.0063
307	0.181	1.875	0.161	0.542	0.162	0.361	0.354	0.389	-	-
315	-	-	-	-	-	-	-	-	0.0013	0.0148
319	0.462	3.761	0.382	0.953	0.384	0.704	0.530	0.665	-	-
330	0.619	4.260	0.542	1.187	0.585	0.992	0.818	0.925	0.007	-
336	-	-	-	-	-	-	-	-	0.015	-
341	1.655	8.748	1.351	2.342	1.400	1.798	1.364	1.408	0.027	0.017
353	-	-	-	-	-	-	-	-	-	0.027
359	3.484	13.328	2.540	3.994	3.575	4.150	2.616	2.794	0.134	0.099
374	4.813	20.888	3.851	6.025	4.180	4.524	3.560	4.643	0.610	0.196
380	6.656	21.681	3.649	7.516	3.958	5.459	5.771	6.933	1.681	0.163

Table 3.1: Observed BIRD rate constants for the dissociation of U_nCa^{2+} complexes into $U_{n-1}Ca^{2+}$ and neutral uracil

Table 3.2: Experimentally observed activation energy values (E_a^{obs}) and pre-exponential factors (A^{obs}), theoretical binding energies (ΔE), and threshold dissociation energies (E_0) for clusters U_nCa^{2+} with n=14-5. Energy values are listed in kJ mol^{-1}.

					[b]ΔE (Theoretical)	
n	E_a^{obs}	$LogA^{obs}$	[a]E_0	[a]$LogA$	(1)	(2)
14	46.0	7.2	84 ± 8	14.3-19.3	60.5	-
13	29.9	5.5	75 ± 13	14.2-19.2	41.7	-
12	44.9	6.9	87 ± 11	14.3-19.4	73.8	-
11	33.3	5.4	80 ± 9	14.2-19.3	62.7	-
10	43.8	6.7	87 ± 12	14.3-19.4	77.5	-
9	36.7	5.8	80 ± 9	14.2-19.4	66.1	85.0
8	40.2	6.3	82 ± 9	14.3-19.4	104.7	121.1
7	39.1	6.2	83 pm 11	14.3-19.6	73.6	113.4
6	104.5	14.4	110 ± 6	14.8-19.7	132.9	145.8
5i[c]	-	-	98 ± 11	14.6-20.1	203.9	178
5ii[c]	40.8	4.8	84 ± 8	14.3-19.3	57	

[a]From master equation modeling.

[b]HMBI MP2/6-311++D(2df,2p)/AMOEBA energies based on (1) structures in figure 5, 6, and S2 and (2) lowest energy structures from simulated annealing.

[c]5i and 5ii refer to structures presented in figure 3.5.

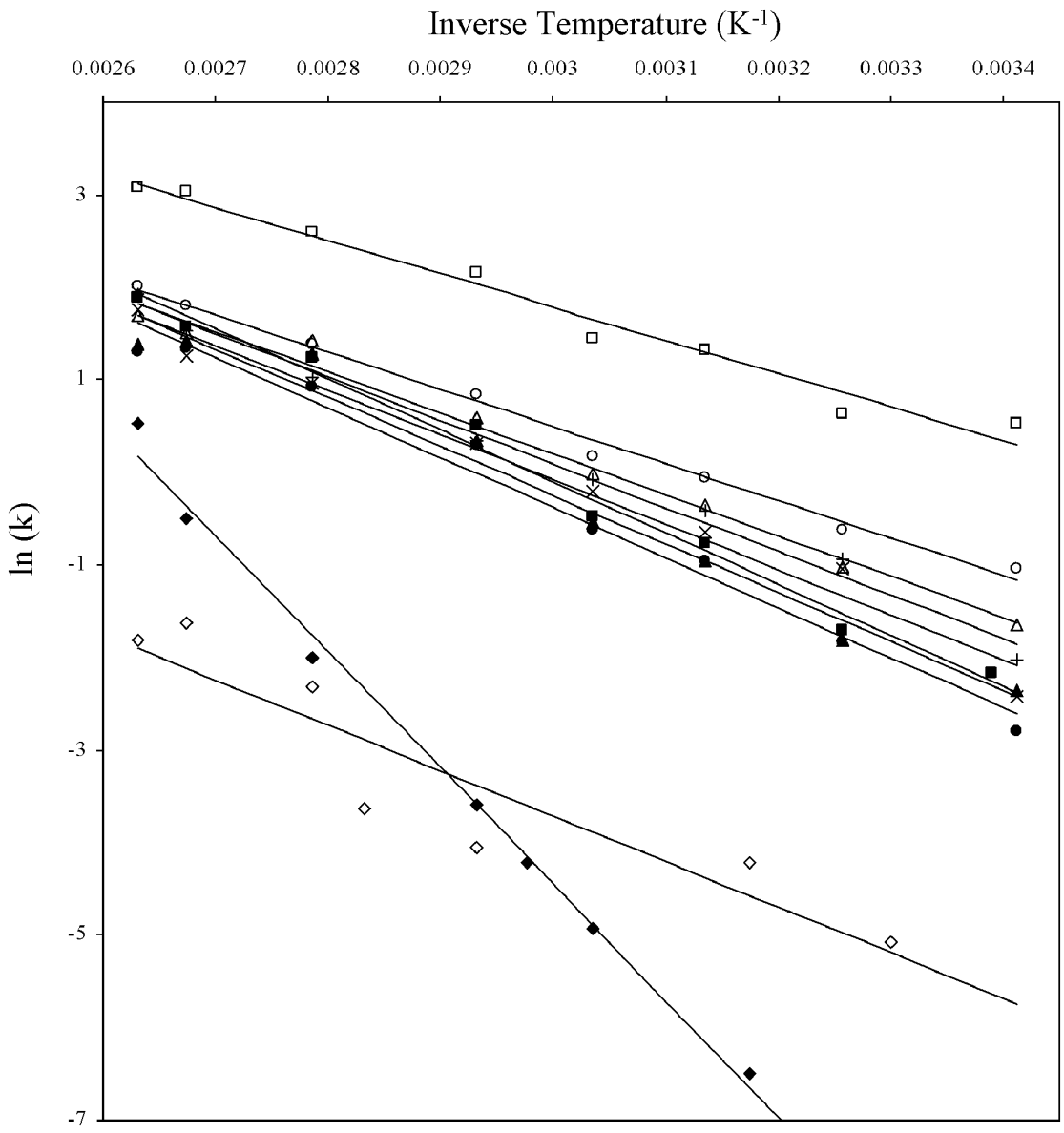


Figure 3.3: Arrhenius plots for U_nCa^{2+} for n=: 14 (filled square), 13 (clear square), 12 (•), 11 (◯), 10 (filled △), 9 (△), 8 (×), 7 (+), 6 (filled ⋄), and 5 (○). Activation energies and log(A) factors are obtained from the slope and intercept of these graphs, respectively. As can be seen, all plots are linear with increased scatter when n=5.

for U_6Ca^{2+} (figure 3.4, table 3.2). For example, the differences in the activation and threshold dissociation energies between U_7Ca^{2+} and U_6Ca^{2+} are 52 and 27 kJ mol^{-1}, respectively. This decrease in the threshold dissociation energy for n $\geq$ 7 suggests that

the calcium ion is coordinated by six uracil molecules in the formation of an inner core while additional uracil molecules bind to this center through hydrogen bonding interactions. These results are consistent with those from a previous study on calcium bound uracil clusters[69].

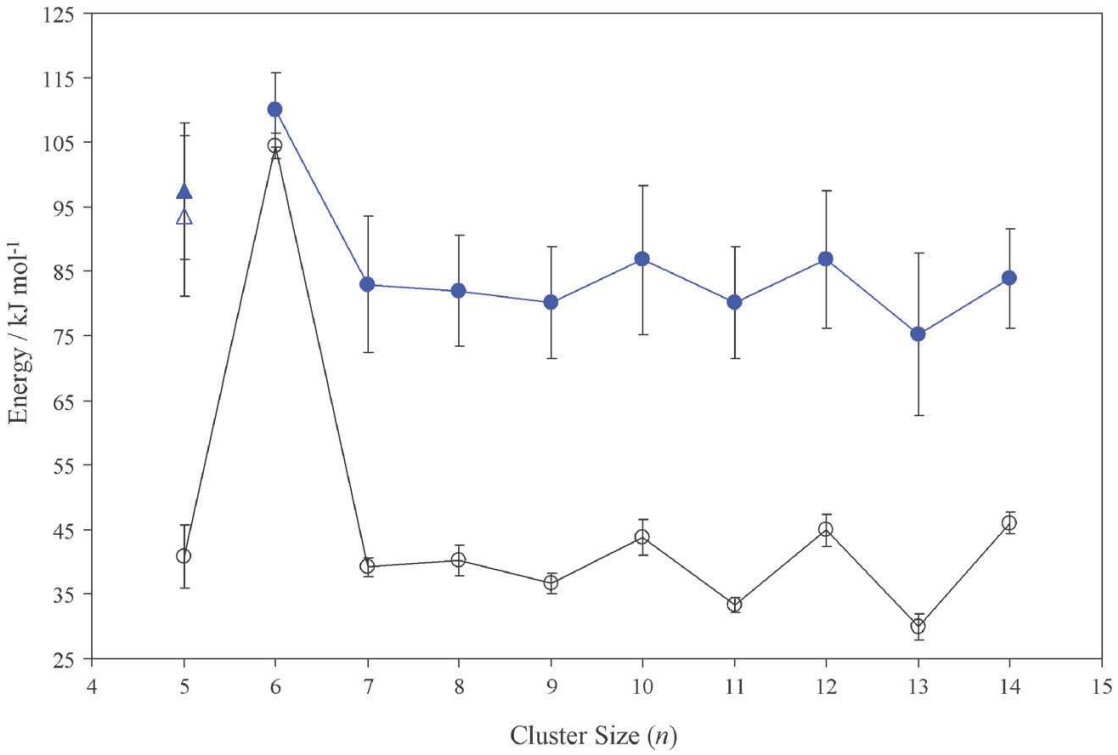


Figure 3.4: Summary of binding energies as a function of cluster size (n) including experimental activation energies (black curve) and master equation modeled threshold dissociation energies (blue curve). The average threshold dissociation energies obtained for n = 5 using frequencies and integrated IR intensities for structures 5i and 5ii (Figure 3.5) are indicated by filled and open triangles, respectively. Corresponding energy values are summarized in table 3.2.

The data in figure 3.4 also indicate that for clusters larger than U_6Ca^{2+}, there is an odd-even alternation in the E_a^{obs} values. This effect is less noticeable for the

modeled E_0 values. The activation energy for the clusters with an even number of uracil molecules is slightly greater than that for clusters with odd number of uracil molecules, with the activation energies ranging from ∼30-39 kJ mol^{-1} and ∼40-46 kJ mol^{-1} for odd and even numbered clusters, respectively. This trend is consistent with a photoionization time-of-flight mass spectrometry study of molecular uracil clusters of up to 30 units, where distinct alternation in cluster intensities between odd and even numbered clusters was observed[78]. This led the authors to suggest the formation of dimers within the clusters. The present BIRD results can also be interpreted that the outer-shell uracil molecules are bound as dimers to the inner core, which is in contrast to the previously suggested structures wherein the outer shell uracil molecules bind in a symmetric fashion about the center[69]. The structures of the U_nCa^{2+} clusters will be discussed in more detail below.

Finally, the measured BIRD activation energy for U_5Ca^{2+} is significantly lower than that for U_6Ca^{2+}, by 60 kJ mol^{-1}. This is quite interesting because previous studies on the larger clusters (n $\geq$ 6), as well as the present results, are interpreted based on a core of six uracils being bound to Ca^{2+}. The reasons for significant decrease in stability of the n = 5 cluster will be explored in more detail below.

3.4.2 Threshold Dissociation Energies From Master Equation modeling

Master equation modeling of U_nCa^{2+} with n = 7-14 resulted in good fits within the specified criteria with a range of E_0 values obtained. These are reported in table 3.2 and plotted as a function of cluster size in figure 3.4 (filled circles, triangles). The calculated E_0 values are consistently higher than the experimental activation energies, indicating that the dissociation kinetics for these clusters are not in the rapid energy

exchange (REX) limit. This is also apparent from the relatively fast BIRD rates. The master equation modeled threshold dissociation energies retain the odd-even alternation that is observed in the experimental activation energies. The trend for the threshold dissociation energy values is not as obvious as for the BIRD activation energies because of the wide range of possible E_0 values extracted using master equation modeling and indicated by the error bars in figure 3.4.

In the case of U_6Ca^{2+} and U_5Ca^{2+}, acceptable fits to the experimental Arrhenius data were not obtained. For U_6Ca^{2+}, fits could be obtained by multiplying the integrated IR intensities by a factor of 9 and the range of threshold dissociation energies obtained are listed in table 3.2 and shown in figure 3.4. The main difficulty with obtaining a good fit to the data was to reproduce the BIRD rate constant at the highest temperature (380 K) of the experiment. The simulated BIRD rate constant was constantly lower than the experimental value. However, a multiplication factor of 9 increased the absorption and emission rates significantly, allowing the modeled BIRD kinetics for U_6Ca^{2+} to approach the REX limit (i.e. $k_{1,rad} > kdissoc$) even at 380 K. This much higher multiplication factor used to fit the Arrhenius data for n = 6 indicates that the calculated radiative rates are underestimated for this cluster. A similar result was found for $Mg^{2+}(H_2O)_6$, where higher than normal multiplying factors were required to fit the Arrhenius data for the conformer present at the lower temperatures[79]. The experimental Arrhenius parameters (E_a = 104.5 kJ mol^{-1}; $log(A)$ = 14.4) are similar to the threshold dissociation energy (110 kJ mol^{-1}) and $log(A)$ values (14.8-19) used in the modeling, indicating that the dissociation kinetics for U_6Ca^{2+} are close to the REX limit. The threshold dissociation energy (110 kJ mol^{-1}) obtained for U_6Ca^{2+} is higher than the values for clusters with n = 7-14 (ranging from 75.3 to 86.8 kJ mol^{-1}). This difference in binding energies between n = 6 and the larger clusters is consistent with a

structure where all 6 uracil molecules are bound to the calcium cation to form the first solvation shell.

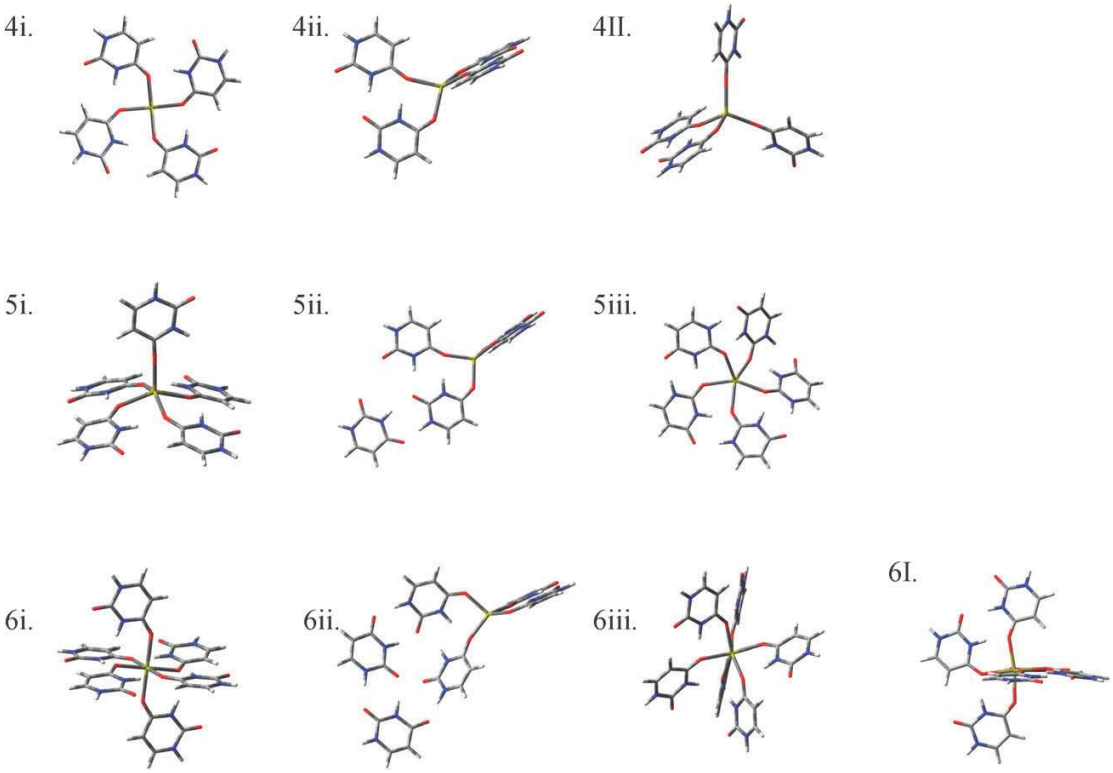


Figure 3.5: Proposed structures for U_4Ca^{2+}, U_5Ca^{2+}, and U_6Ca^{2+}.

Master equation modeling of U_5Ca^{2+} was performed using frequencies and integrated IR intensities for both a structure with 5 uracil molecules directly bound to the calcium cation (structure 5i in figure 3.5) and a structure with 4 uracil molecules directly bound to the calcium cation with the fifth uracil bound to one of the uracil molecules in the inner core (structure 5ii in figure 3.5). However, a good fit to the experimental data using scaling factors 0.5-2 for the integrated IR intensities was not obtained. Using multiplying factors of 0.25 and 0.15 produced good fits to the data for structures 5i and 5ii, respectively. The origin of the significantly lower multiplying factors required to fit the data for n = 5 compared to those for the larger clusters is unknown, although

the much higher uncertainty in the Arrhenius data for this cluster may contribute substantially to this discrepancy. The resulting range of threshold dissociation energies are presented in table 3.2 and figure 3.4. The threshold dissociation energies obtained for U_5Ca^{2+} are lower than those for U_6Ca^{2+}, which is in contrast to the theoretical binding energies with all five uracils bound to Ca^{2+} (vide infra). The lower E_0 value obtained for U_5Ca^{2+} from master equation modeling is, however, consistent with the presence of a less stable conformer (possibly structure 5ii) that is observed predominantly in the BIRD experiments due to its faster dissociation rate.

3.4.3 Structure and Dissociation Energy

Zins et al. have previously proposed structures for U_nCa^{2+} with n = 6-14 based on ab initio/DFT calculations and CID experiments.18 These structures are characterized by a first solvation shell containing six uracil molecules about the calcium ion with strong hydrogen bonding between uracil molecules. Larger clusters are then composed of uracil molecules binding to the inner core. These results are consistent with the trends in activation energies and threshold dissociation energies obtained from the BIRD experiments and master equation modeling performed in this work. Furthermore, the BIRD data is characterized by an odd-even alternation in the observed activation energies. To investigate the relative stabilities of clusters with odd and even numbers of uracil molecules in more detail, higher level computational methods were used to calculate possible structures for U_nCa^{2+} with n = 4-14.

It will be useful to begin by discussing some of the binding motifs identified from electronic structure calculations. For the smaller clusters, U_nCa^{2+} with n = 4-6, chemical intuition was initially used to identify some possible geometries while a

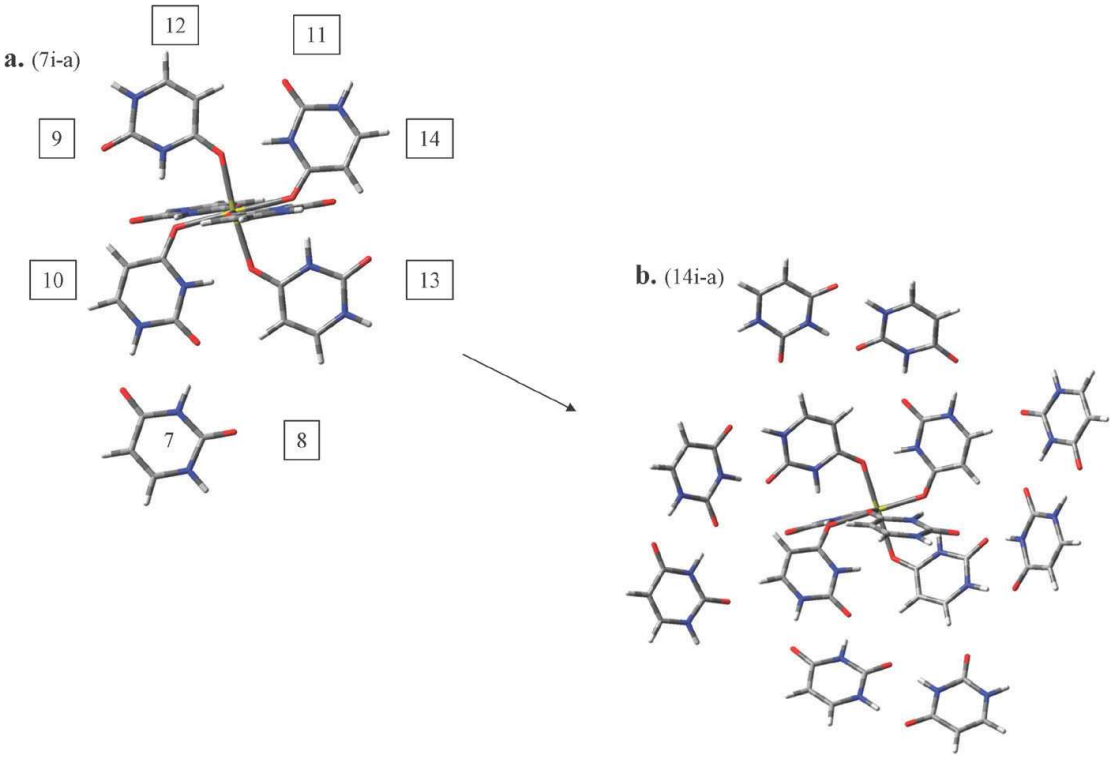


Figure 3.6: Representative structures for a. $U_7Ca^{2+}$and b. $U_{14}Ca^{2+}$. Numbers in boxes surrounding the U_7Ca^{2+} structure indicate the location and order of binding for clusters with 7-14 uracil molecules about the central U_6Ca^{2+} inner shell. Structures with n = 8-13 are presented in supplementary information S2.

series of structures similar to those proposed by Zins et al.18 were initially examined for the larger clusters (i.e. n = 7-14). These larger structures all involve an octahedral U_6Ca^{2+} cluster core. Additional uracil molecules contribute to a second solvent shell, forming concave arms around four of the six octahedral uracils, as shown in figure 3.6. Concern about whether other, lower-energy structures existed led us to perform a series of simulated annealing runs on the clusters with n $\leq$ 9. The results of this study showed a variety of new structures with substantially lower energies, including a new packing motif for clusters with n = 6-9. The energetics of selected structures are presented in table 3.3, and the relative binding energies are plotted in figure 3.7. The calculations

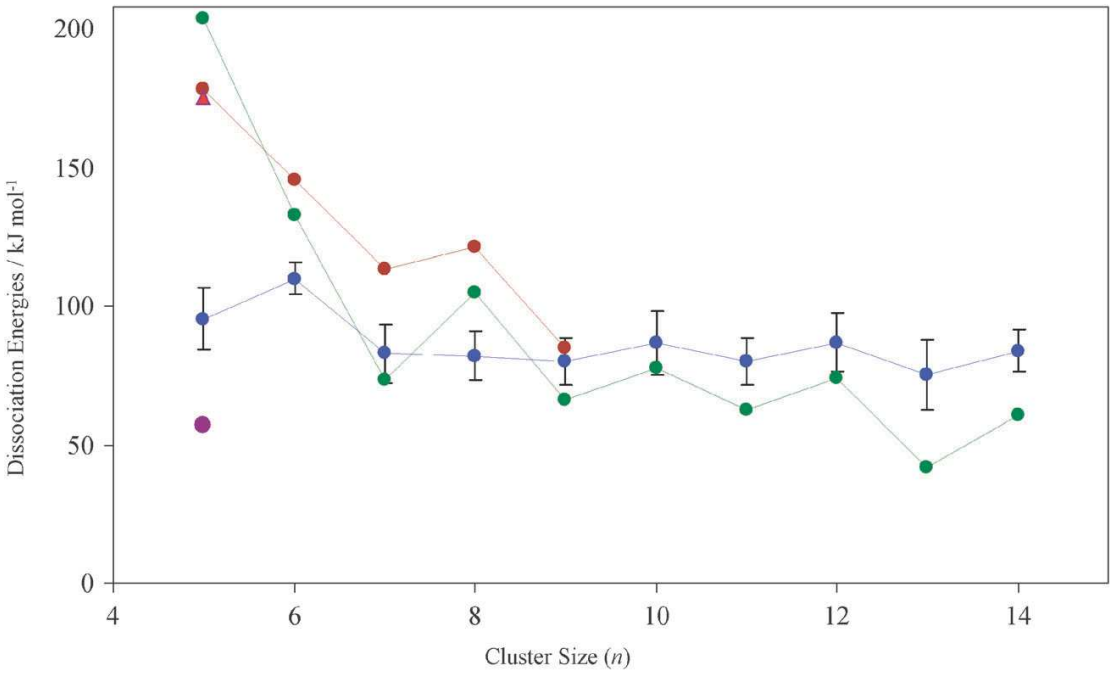


Figure 3.7: HMBI MP2/6-311++G(2df,2p)/Amoeba binding energies as a function of cluster size (n) obtained from initial geometries based on an octahedral 6i core (green) and lowest energy structures from simulated annealing (brown). Binding energies for structures 5ii (purple filled circle) and 5iii (red filled triangle) are also presented. Corresponding values are presented in table 3.2.

predict the odd-even oscillatory binding energies in the larger U_nCa^{2+} clusters (n = 7-14), as was observed experimentally. The binding energy trends will be discussed in greater detail below, first focusing on results from the initial conformations and then the more in-depth results of the simulated annealing study.

3.4.3.1 U_nCa^{2+} (n = 4-6).

Proposed structures for U_nCa^{2+} (n = 4-6) are presented in figure 3.5. Consider first the U_4Ca^{2+} cluster where two stable structural motifs were identified. The lowest energy structure, 4II, adopts a distorted tetrahedral geometry with hydrogen bonding

Table 3.3: Relative HMBI MP2/6-311++G(2df,2p)/Amoeba energies of Clusters U_nCa^{2+} with n = 4-8 in kJ mol^{-1}

	N						
Structure	4	5	6		7	8	9
i	32.8	0.0	[a]65.9/0.7	**i-a**	105.6	122.1	140.9
				i-b	15.9	37.8	[b]32.6
ii	[a]23.9/0.0	120.7	149.3		184.2	193.7	
iii		95.0	0.0		0.0	0.0	0.0

[a]Values indicated are for sub-structures "i" and "I" in series, respectively.

[b]Structure not presented.

occurring between a pair of uracil molecules. Structure 4ii, which is higher in energy by ∼24 kJ mol^{-1}, has a similar tetrahedral based structure. An additional configuration was also identified, 4i, which has a square planar (slightly concave) geometry characteristic of quartet complexes with all four uracil molecules binding to the calcium cation and interacting with each other through hydrogen bonding. This geometry lies roughly 32.8 kJ mol^{-1} above the lowest-energy tetrahedral structure (4II). Experimental BIRD data for U_4Ca^{2+} could not be obtained.

The lower activation energy and threshold dissociation energies for U_5Ca^{2+} compared to those for U_6Ca^{2+}, would not be expected based on the lowest energy structure obtained for U_5Ca^{2+} (5i). IRMPD kinetics studies were performed to further understand the possible reasons for this discrepancy. IRMPD kinetics were obtained for clusters with n = 5, 6, 8, and 10, the results of which are presented in figure 3.8 and S1.w Both IRMPD and BIRD resulted in sequential loss of uracil molecules (figure 8iii). For all four clusters, the IRMPD kinetics were found to be bi-exponential. However, there is a marked difference between the amount of the second population of ions for the n = 5 cluster compared with the n = 6, 8, and 10 clusters. For the latter three clusters the

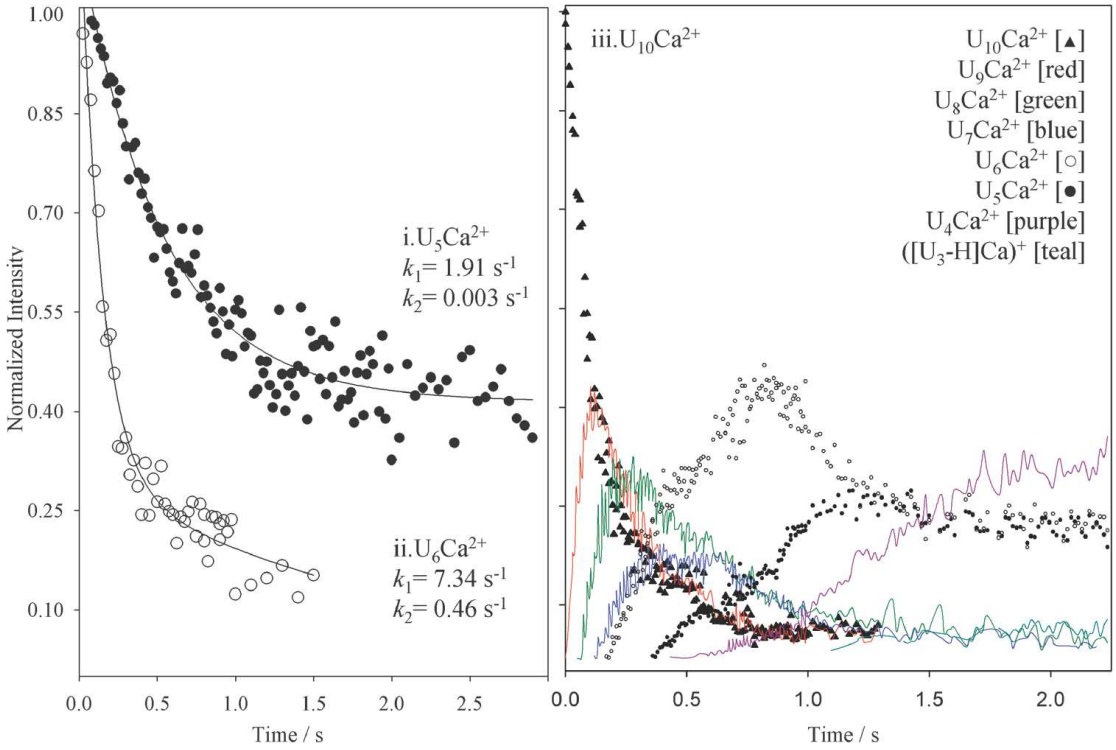


Figure 3.8: IRMPD kinetics of clusters i. U_5Ca^{2+}, ii. U_6Ca^{2+}, and iii. $U_{10}Ca^{2+}$. Curve (iii) also shows the dissociation of product ions through the dissociation of U_4Ca^{2+}. Curves and rate constants shown are the result of fitting the data to a bi-exponential decay function.

slower dissociation fraction are between 10-20%. The slower dissociating fractions for n = 6, 8 and 10 is likely to be a result of poor overlap of the ion population orbiting the ICR cell and the CO2 laser used for IRMPD. For U_5Ca^{2+}, however, the IRMPD kinetics data shows that about half[80] of the ion population has a significantly lower IRMPD rate constant suggesting the presence of at least two distinct non-interconverting populations of ions[73]. Bi-exponential curve-fitting of the kinetics results in rate constants $k_1 = 1.9$ s^{-1} and $k_2 = 0.003$ s^{-1} for the fast and slowly dissociating populations, respectively. The kinetics measured in the BIRD experiments for U_5Ca^{2+} represent dissociation of the more weakly bound ion population. The multiple conformers in the ion population

for U_5Ca^{2+} presumably contribute to the increased scatter observed in the Arrhenius plot for this ion.

Candidate structures for U_5Ca^{2+} to explain the bi-exponential IRMPD kinetics observed in the experiments are shown in figure 3.5. Forming a pentamer from the two tetramer structures discussed above (planar and tetrahedral) produces two very different structures. The square planar structure for U_4Ca^{2+} easily allows for a fifth uracil to bind directly to the calcium cation to produce the most favorable U_5Ca^{2+} structure, 5i. This structure adopts a square pyramidal arrangement with added stability from five strong uracil-calcium cation interactions. Alternatively, structure 5ii can be obtained from addition of a uracil to structure 4ii, placing the fifth uracil in a second solvent shell, hydrogen bonded to the tetrahedral core and interacting only relatively weakly with the calcium cation. This latter structure is some 120 kJ mol^{-1} less stable than the square pyramidal 5i. An additional structure was identified, 5iii (figure 3.5), that is similar to structure 5i as all five uracil molecules bind directly to the calcium cation. Similar to structures identified by Qiu et al[68] in a study of uracil as interacting with K^+, Rb^+, and Cs^+, structure 5iii is characterized by strong hydrogen bonding between uracil molecules and O2 oxygenCa^{2+} interactions. Energetically, the structure is favored over the tetrahedral configuration but is still $\sim$95 kJ mol^{-1} higher in energy than the energetically favored 5i square pyramidal structure.

The tetrahedral motif of structure 5ii places the fifth uracil in a second solvent shell where it is much easier to remove (57 kJ mol^{-1} instead of 178 and 175 kJ mol^{-1} for structures 5i and 5iii, respectively). It is unlikely that the observed fast kinetics corresponds to the strongly bound structures 5i or 5iii, however, the much weaker bound 5ii structure based on the tetrahedral core could explain the fast kinetics. While it is not entirely obvious what circumstances would lead to the formation of the tetrahedral

core given its much lower stability, it may form by addition of uracil to the lowest energy tetrahedral U_4Ca^{2+} structure. In this case, the geometry of the U_4Ca^{2+} structure would prevent the formation of structures 5i/5iii. We hypothesize, therefore, that a mixture of $U_{5i}Ca^{2+}$ and $U_{5ii}Ca^{2+}$ like structures (calcium coordinated by either four or five uracils) is present in the gas phase. Conformer 5i (and perhaps 5iii) is also likely present but due to its slower BIRD kinetics, does not contribute significantly to the observed BIRD data observed at shorter times. The presence of higher energy structures for U_5Ca^{2+} might be explained by considering their formation during the experimental process. For example, fragmentation of larger clusters such as U_6Ca^{2+} could occur in the source region of the mass spectrometer to form 5i like pentamers. Of course structures 5i (and 5iii) could also be formed during the electrospray process. It is also possible, however, that during the electrospray process, a uracil could condense with the thermodynamically more stable tetrahedral U_4Ca^{2+} to form 5ii; a cluster where the fifth uracil is part of the second solvation shell. It is hypothesized here that the tetrahedral based 5ii complex is that for which dissociation is observed in the BIRD experiments as the other complexes are too strongly bound.

Four possible structures are shown in figure 3.5 for U_6Ca^{2+}. Structure 6ii is similar to 5ii with an additional uracil molecule in the second solvent shell of the tetrahedral motif. Alternatively, one can form an octahedral-type structure 6i by adding the sixth uracil on the opposite side of the calcium as the fifth one in square pyramidal structure 5i. As expected, octahedral hexamer structures are much more stable than 6ii (by $\sim$83 and 150 kJ mol^{-1} for structures 6i and 6I, respectively) as a result of all six uracil molecules being allowed to bond directly to the calcium ion. Also identified is a very stable prism (or [2.2.2]propellane-like) motif (structure 6iii in figure 3.5) which lies only 0.7 kJ mol^{-1} below octahedral structure, 6I. These calculations obviously cannot

reliably be used to energetically rank such structures with similar energies, however, they at least suggest that both hexamer motifs (octahedral and prism) are equally possible. It should be noted that for all structures presented, with exception of 5iii the uracil monomer binds to the calcium cation through the O4 oxygen rather than the O2 oxygen. This is consistent with previous studies which have demonstrated that in the interaction of uracil with metal ions, uracil preferably binds to metal ions at the O4 oxygen[81, 82, 83, 84]. Specifically, Trujillo et al[85] found the binding of Ca^{2+} to the O4 oxygen in uracil was favored over the O2 oxygen by 41 kJ mol^{-1}. This is consistent with the Gibbs energies calculated in this work at the B3LYP/3-21g(d) level of theory for U_6Ca^{2+} which indicate that $O4Ca^{2+}$ binding is favored by $\sim$60 kJ mol^{-1} over binding to the O2 oxygen (i.e. $O2Ca^{2+}$).

The dissociation energies of the different U_6Ca^{2+} structures are relatively similar (within $\sim$28 kJ mol^{-1}) compared to the very different dissociation energies for 5i and 5ii. This makes it difficult to observe the different populations in the IRMPD kinetics. Based solely on the theoretical binding energies, the observation of dissociation of 6i to 5i and 6ii to 5ii would require the energy equivalent of at least 14 and 12 photons from the CO_2 laser, respectively, whereas the 5i to 4i and 5ii to 4ii dissociation would require 21 and 6 photons, respectively. In contrast, one would expect to observe some indication of multiple populations in BIRD kinetics based on the difference in dissociation energies even for the n = 6 to n = 5 dissociations. However, the two different structures, 6i and 6ii, would also have quite different transition states for the loss of uracil. For 6ii, where two uracil molecules are in the second solvation shell, the transition state for loss of uracil will be early (tight) with a smaller entropy difference between the reactant and the transition state. However, for 6i where all uracil molecules are bound directly to Ca^{2+}, one would expect a much later (loose) transition state since uracil

needs to get farther away from the ion to reach the transition state. As such, the 6i to 5i dissociation with a larger transition state enthalpy requirement will also have a larger pre-exponential factor. In fact, a roughly 10^4 difference in pre-exponential factor would not be uncommon between a tight and loose transition state. As a result, the opposing effects of the activation energy and entropy of activation could leave the BIRD rate constants similar for loss of uracil from 6i and 6ii. At best, this would make it difficult to differentiate between the two structures in the kinetics plots. The opposing effects on the rate constants of the entropy and enthalpy differences between reactants and the transition state could even explain the slight curvature for the n = 6 Arrhenius plot at higher energies.

3.4.3.2 U_nCa^{2+} (n = 7-14).

HMBI MP2/6-311++G(2df,2p)/Amoeba binding energies were calculated for calcium bound clusters with 5-14 uracil molecules and are listed in table 3.2 and plotted in figure 3.7 as a function of cluster size.

Representative structures for larger cluster (n = 7-14) are shown in figures 3.6, 3.9, S2, S3. For clusters larger than n = 6, calculated structures indicate that additional uracil molecules bind to an inner octahedral core containing calcium coordinated to six uracil molecules to form the second solvent shell. This observation agrees with both the experiments performed here and with earlier work,[69] and is supported by simulated annealing calculations, which did not identify any structures with more than six uracil molecules coordinated to the calcium ion. Both octahedral and prism like hexamers are obvious candidates for the inner core of the larger clusters. Representative binding schemes for the octahedral core larger clusters starting with U_7Ca^{2+} (structure 7i-a)

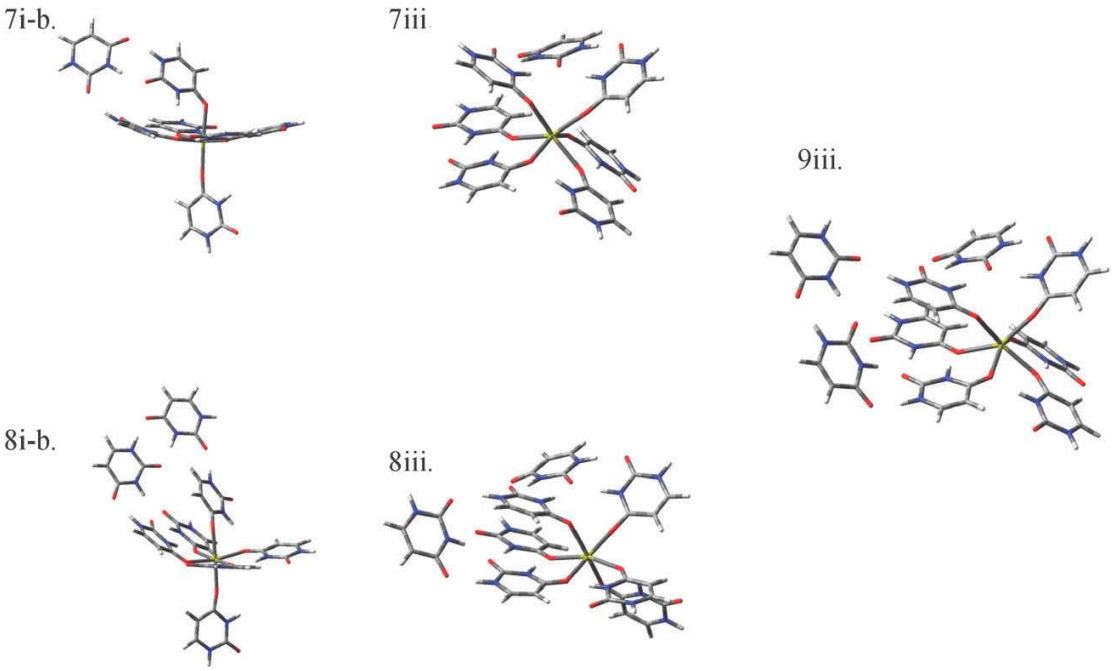


Figure 3.9: Proposed structures based on a simulated annealing study for U_7Ca^{2+}, U_8Ca^{2+}, and U_9Ca^{2+}.

and leading to $U_{14}Ca^{2+}$ (structure 14i-a) are presented in figure 3.6, with structures of $U_{ni-a}Ca^{2+}$ (n = 8-13) presented as supplementary data S2. Structures for the 7 and 8-mer clusters formed by the addition of uracil to the $U_{4ii}Ca^{2+}$ tetrahedral core are presented in supplementary information S3.

The most stable U_7Ca^{2+} structure (7iii) found is similar to the prism-like 6iii except with the top two uracil molecules rotated to orient their aromatic rings facing toward the lower uracil quartet, and with the seventh uracil molecule hydrogenbonded between them (figure 3.9). Presumably, the rotation of the top two uracil molecules occurs to allow this seventh to approach the calcium as closely as possible. The result of this is that the oxygen of the seventh uracil resides only 4.1 Å away from the calcium even though it is in the second solvent shell, (n.b. OCa distance is ∼2.4 Å for the inner uracils). One can also form U_7Ca^{2+} structures based on addition to the octahedral

core. The most stable such structure (7i-b) was found to be 16 kJ mol^{-1} higher in energy than structure 7iii. Structure 7ii, based on the tetrahedral core and presented in supplementary data S3, is 184 kJ mol^{-1} less stable than 7iii.

Similar results are found for U_8Ca^{2+}. Structure 8iii, which looks like 7iii with an additional uracil in the second solvent shell, is the most stable while the octahedral core structure 8i-b, is even less favorable for this cluster size, lying 50 kJ mol^{-1} above 8iii. By the time one reaches U_9Ca^{2+}, the large number of degrees of freedom makes the calculations expensive and clear structural identification difficult. For this U_9Ca^{2+} cluster, a total of 40 simulated annealing runs were performed with Amoeba and optimized with DFT. These resulted in a mixture of structures involving either the pi-stacking[86] or the octahedral core motifs. After running the HMBI MP2/Amoeba single point energies on some of the most stable of these structures, we found that structure 9iii (pi-stacking core) lies roughly 30 kJ mol^{-1} below the lowest energy octahedral structures. For the larger U_nCa^{2+} with n = 10-14 clusters, only the initial octahedral core structures were examined.

The theoretical calculations qualitatively reproduce the overall experimental binding energy trends (i.e. odd-even alternation) as shown in figure 3.7. The decrease in the binding energy for the seventh and higher uracil molecules indicated by experiments is consistent with the prediction of a six-uracil core structure. Once six uracil molecules are present, the additional molecules bind in a second solvent shell and cannot participate in the strong ion-dipole interactions that characterize the inner core. Calculated binding energies (figure 3.7 and table 3.2) indicate that the seventh and eighth uracil molecule bind somewhat stronger than additional uracil molecules, which is understandable since upon each addition of uracil the charge becomes more dispersed. In structure 7iii (figure 3.9), the seventh uracil molecule binds to the inner shell uracil

molecules and can therefore more directly interact with the calcium ion, leading to a relatively high binding energy. The eighth uracil (structure 8iii in figure 3.9) is further away and does not interact as strongly with the calcium, but it forms multiple hydrogen bonds with other uracil molecules and reduces overall strain in the hydrogen bonding network. In particular, one of the hydrogen bonds of the seventh uracil molecule in structure 7iii decreases from 2.02 Å to 1.91 Å in structure 8iii, while its angle increases from 153 to 163 degrees. This leads to a greater stability for U_8Ca^{2+} (8iii) resulting in a higher binding energy (121 vs. 113 kJ mol^{-1}) for U_8Ca^{2+} (8iii) compared with U_7Ca^{2+} (7iii). The ninth uracil (structure 9iii) attaches rather far from the calcium, so its binding energy is somewhat weaker (85 kJ mol^{-1}).

The octahedral core structures for n = 6-14 also exhibit an oscillatory binding trend in the theoretical HMBI binding energies between odd and even numbered clusters for n > 6, as shown in figure 3.7. In this case, the clusters containing an odd number of uracil molecules all have one uracil molecule which forms two hydrogen bonds with one uracil molecule in the core. Adding one more uracil to produce an even-numbered cluster saturates the hydrogen bonding sites of this "dangling" uracil by making two hydrogen bonds to it and one hydrogen bond to a core uracil, effectively "closing the loop." In other words, all uracil molecules in the even-numbered clusters form three hydrogen bonds to two other uracil molecules, making it more strongly bound than the n-1 uracil which only makes two hydrogen bonds. It is interesting to note that the differences in activation energies between the odd and even clusters, for n = 9-14, are between 10 and 15 kJ mol^{-1} (figure 3.4), roughly the strength of a hydrogen bond.

Finally, it is important to recognize that the global minimum energy structures for each of the complexes have likely not been found, especially for the larger clusters. There are simply too many degrees of freedom, and not nearly enough cycles of simulated

annealing were performed to fully explore the potential energy landscapes. Moreover, optimizing the geometries with DFT and calculating HMBI single-point energies for the larger systems was simply too computationally expensive to perform on the large number of possible structures. One might hope to use less-computationally expensive levels of theory to pre-screen for the lowest-energy structures in the larger clusters. Unfortunately, the Amoeba and B3LYP energies were too unreliable to predict the minimum energy structures found with HMBI, making them only moderately useful for screening. For these reasons, the simulated annealing studies were performed only for the smaller structures, i.e. U_nCa^{2+} where n $\leq$ 9. It is also important to note that even if the global minimum structures were identified for each cluster size, the real-world relevance of these structures is debatable. For a given cluster size, many structures typically lie within a few kJ mol^{-1} of the lowest-energy structure. These different structures often correspond to simple rotations of individual uracil molecules relative to the others and/or slightly less favorable hydrogen-bonding patterns. However, sometimes the differences are more dramatic. For example, for U_6Ca^{2+} clusters, two stable but very different structures (prism vs. octahedral) are within only 0.7 kJ mol^{-1} of each other. The lack of a clear thermodynamic minimum also implies that there may be ensembles of structures in the gas-phase and that the dynamics of the cluster dissociation may also be quite important. In this context, the calculations reported here aim to identify common binding motifs and trends and to explain the experimental binding energy trends qualitatively.

3.5 Conclusion

Uracil clusters, U_nCa^{2+} with n = 14-4, were investigated with both experimental (BIRD, IRMPD) and theoretical methods. BIRD studies and master equation modeling of the BIRD data were performed on clusters with 5-14 uracil molecules to determine Arrhenius parameters and obtain threshold dissociation energies. These results showed that the threshold dissociation energy for U_6Ca^{2+} is significantly higher than that for larger sized clusters (n = 7-14). This suggests that the first solvent shell is composed of six uracil molecules about the calcium ion which is consistent with findings from previous CID studies[69]. The threshold dissociation energies obtained from master equation modeling of the Arrhenius data for clusters with n > 6 showed an odd-even alternation consistent with the measured activation energies and theoretically calculated binding energies. This trend suggests that larger uracil clusters beyond the first solvation shell may be composed of dimeric units of uracil. Additionally, theoretical calculations indicate that cluster geometries characterized by the formation of dimers bound to the U_6Ca^{2+} center are favored due to increased hydrogen bonding interactions between the uracil molecules forming the dimer.

IRMPD kinetics of U_5Ca^{2+} showed that at least two distinct populations of the cluster are present in the gas phase. It is likely that both a structure wherein all five uracil molecules are bound to Ca^{2+} and a structure where the fifth uracil is bound to a four-membered inner shell are present in the experiments, with the latter mainly contributing to the observed BIRD kinetics.

3.6 Supplementary Information

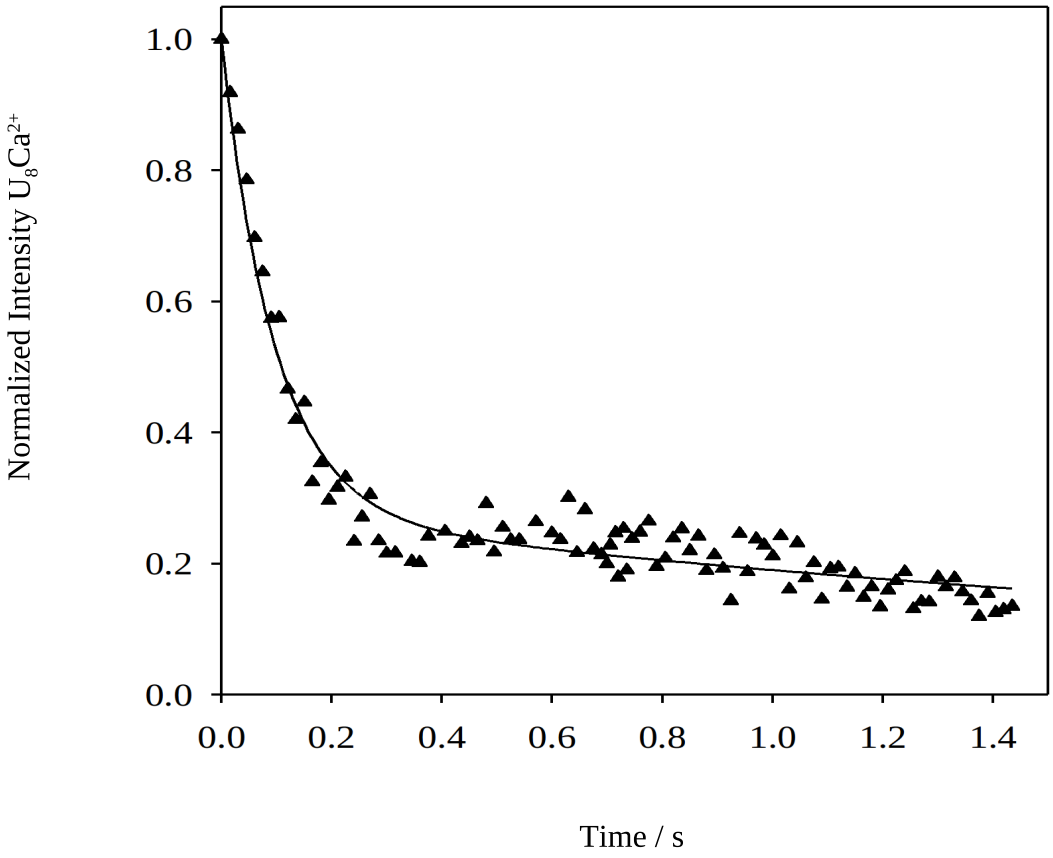


Figure S1: IRMPD kinetics of U_8Ca^{2+}.

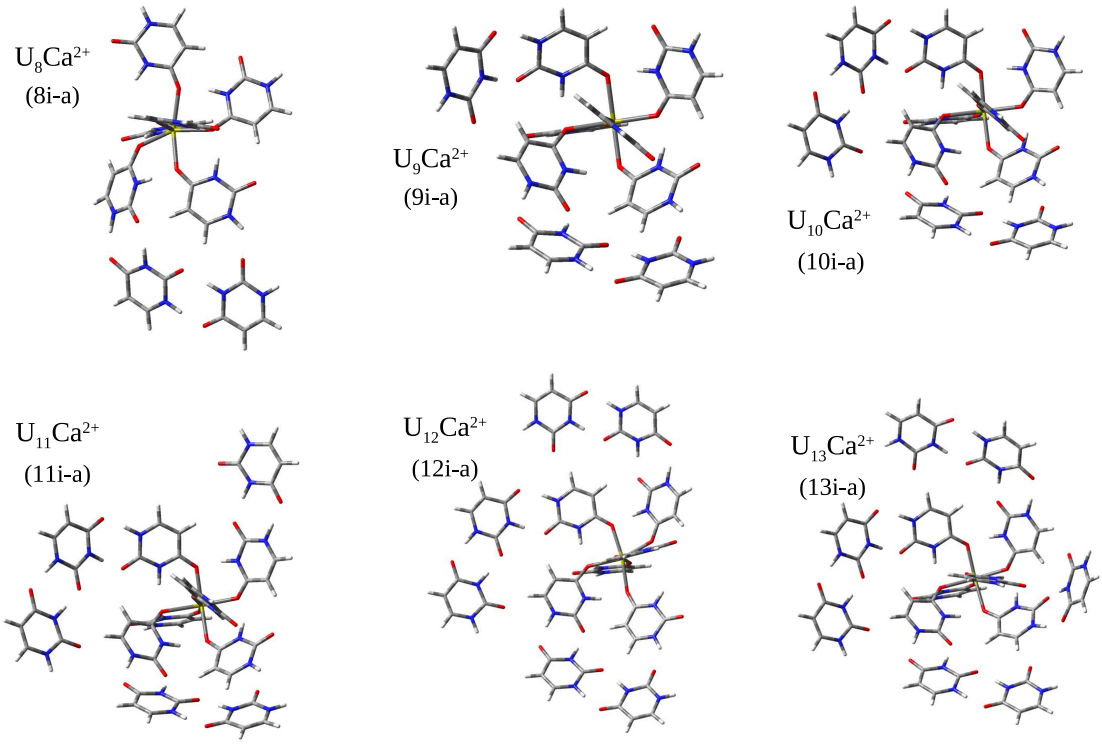
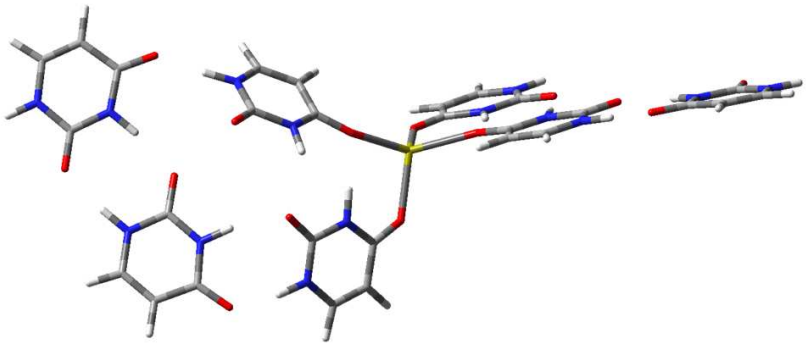


Figure S2: Proposed structures for U_nCa^{2+} with n = 8-13.

7ii.



8ii.

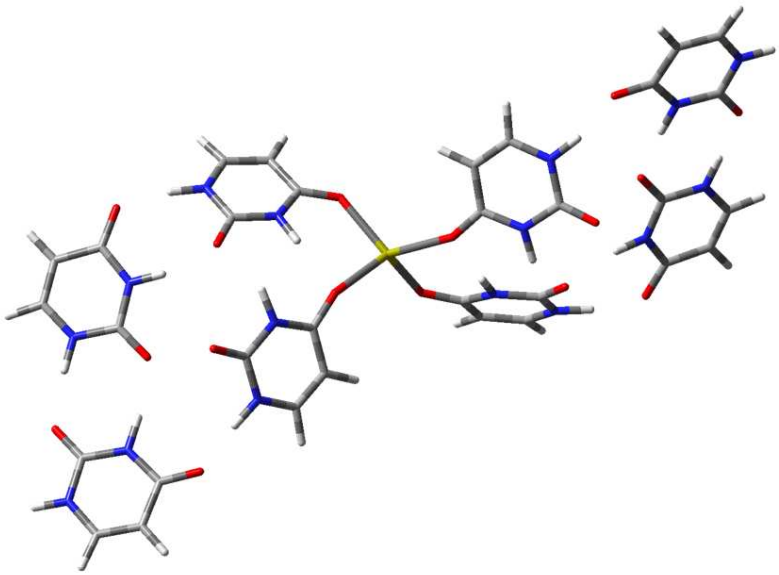


Figure S3: Proposed structures for U_nCa^{2+} with n = 7-8 obtained by adding uracil molecules to the tetrahedral based inner core structure, 4ii.

Chapter 4

Prediction of Organic Molecular Crystal Geometries From MP2-Level Fragment Quantum Mechanical/Molecular Mechanical Calculations

4.1 Outline

The fragment-based hybrid many-body interaction (HMBI) model provides a computationally affordable means of applying electronic structure wavefunction methods to molecular crystals. It combines a quantum mechanical treatment of individual molecules in the unit cell and their short-range pairwise interactions with a polarizable molecular mechanics force-field treatment of long-range and many-body interactions.

Here, we report the implementation of analytic nuclear gradients for the periodic model to enable full relaxation of both the atomic positions and crystal lattice parameters. Using a set of five, chemically diverse molecular crystals, we compare the quality of the HMBI MP2/aug-cc-pVDZ-level structures with those obtained from dispersion-corrected periodic density functional theory, B3LYP-D*, and from the Amoeba polarizable force field. The MP2-level structures largely agree with the experimental lattice parameters to within 2%, and the root-mean-square deviations in the atomic coordinates are less than 0.2 Å. These MP2 structures are almost as good as those predicted from periodic B3LYP-D*/TZP and are significantly better than those obtained with B3LYP-D*/6-31G(d,p) or with the Amoeba force field.

4.2 Introduction

Molecular crystal structure impacts diverse properties ranging from the solubility of a pharmaceutical to the charge-carrier mobility in an organic semiconductor material. Substantial interest lies in understanding, predicting, or engineering different molecular crystal packing motifs, or polymorphs. Theoretical molecular crystal structure prediction requires the ability to compute both the energetics and optimal structure for a given packing motif with high accuracy.

Molecular mechanics (MM) force-field methods have traditionally been used for molecular crystal modeling.[87] However, given the need to discriminate small energy differences between viable polymorphs (1–10 kJ/mol) and the sensitivity of these energetics to the crystal packing (e.g. competition among intramolecular interactions, hydrogen bonding, electrostatics, and van der Waals dispersion), quantum mechanical (QM) methods now play an increasingly important role in such predictions.

Most QM molecular crystal calculations rely on periodic density functional theory (DFT). When van der Waals dispersion is included, density functional theory often works well for molecular crystals.[88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104] However, its predictions can be sensitive to the choice of functional and the nature of the (usually empirical) dispersion correction.[105, 93, 11, 38] For this reason, researchers are actively searching for practical alternatives to periodic DFT for molecular crystals. Recent algorithmic developments have made periodic second-order Møller-Plesset perturbation theory (MP2) much more practical for molecular crystal problems,[106, 33, 107, 108, 109] but it remains computationally expensive for larger unit cells. Also, three-dimensional periodic MP2 structure optimizations are currently rare, and analytical nuclear gradients[110] have not been implemented for the most efficient periodic local MP2 algorithms.

Fragment-based electronic structure methods provide another potentially accurate and computationally practical approach for modeling molecular crystals. As discussed in two recent reviews,[111, 14] such methods typically partition the crystal into individual molecules and include interactions with other molecules using a many-body expansion or a hierarchical scheme. Using one such fragment method, Sode and coworkers demonstrated accurate crystal structure optimizations for the hydrogen bonded formic acid and hydrogen fluoride crystals at the MP2 level.[112, 113] Collins and coworkers have examined the performance of their systematic fragmentation method for reproducing the structures of non-conducting covalent crystals.[114, 115] However, the performance of fragment methods and the benefits of employing methods beyond DFT have not been addressed for a broader range of molecular crystals.

One fragment strategy combines a high-level treatment of the most important intermolecular interactions with a lower-level treatment of the entire periodic system

(e.g. refs [116, 117, 118, 119, 120, 121, 122]. See also the aforementioned reviews for more examples). Our recently developed hybrid quantum/classical fragment model combines a QM treatment of the individual unit cell molecules and their short-range pairwise intermolecular interactions with a classical polarizable force-field approximation for long-range and many-body interactions.[9, 12] We have demonstrated that our hybrid many-body interaction (HMBI) model predicts crystal lattice energies for several small molecules to within 1–2 kJ/mol of the nominal experimental values.[11, 12] Of course, the experimental errors are often larger than that. It also provided new insights into aspirin[15] and oxalyl dihydrazide[123] polymorphism, and it has explained the experimentally-observed oscillatory binding energy trends in gas-phase calcium-uracil clusters.[13]

However, none of our previous studies fully optimized the crystal structures using the fragment method. Here, we present complete nuclear gradients for the HMBI energy with respect to both the atomic positions and the lattice parameters. Then we test the ability of the HMBI model to reproduce experimental crystal structures in a set of five molecular crystals with diverse types of intermolecular interactions, and we compare these results to those obtained with a polarizable force field and with a dispersion-corrected B3LYP density functional.

We demonstrate that MP2-level HMBI calculations in a double-zeta basis set combined with the Amoeba polarizable force field reproduce the experimental lattice parameters to within a few percent, and they also exhibit small root-mean-square deviations in the predicted atomic coordinates. These MP2-level results favorably with those obtained with the B3LYP-D* density functional or the Amoeba force field.

4.3 Theory

4.3.1 The Hybrid Many-Body Interaction Model

The hybrid many-body interaction model for periodic molecular crystals decomposes a crystal into individual molecules ("monomers") and their intermolecular interactions according to a many-body expansion.[12] Monomers in the central unit cell and short-range pairwise interactions are treated quantum mechanically (QM), while longer-range two-body and many-body terms are approximated with a classical molecular mechanics (MM) polarizable force field,

$$E^{HMBI} = E^{QM}_{\text{1-body}} + E^{QM}_{\substack{short\text{-}range \\ \text{2-body}}} + E^{MM}_{\substack{long\text{-}range \\ \text{2-body}}} + E^{MM}_{many\text{-}body} \tag{4.1}$$

For the two-body and higher interactions, at least one monomer resides in the central unit cell.

Typically, the largest MM contribution comes from the many-body terms. Nevertheless, the partitioning of the two-body interactions merits two further comments. First, longer-range interactions can be well-approximated with MM, and substantial computational savings are reaped by treating them classically instead of quantum mechanically. In fact, because each unit-cell molecule interacts quantum mechanically with only a finite number of nearby neighbors, the total number of QM short-range pairwise interactions grows only linearly with the number of molecules in the unit cell. Second, the long-range lattice sums are performed classically and inexpensively. Specialized periodic variants of QM electronic structure methods are unnecessary.

In practice, the HMBI energy expression in Eq (5.1) is rearranged and evaluated

according to:

$$E^{HMBI} = E^{MM}_{PBC} + \sum_i \left(E^{QM}_i - E^{MM}_i\right) + \sum_{ij(0)} d_{ij(0)} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(0)}\right) + \frac{1}{2}\sum_i \sum_{j(\mathbf{n})}^{images} d_{ij(\mathbf{n})} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(\mathbf{n})}\right) \tag{4.2}$$

In this expression, the full crystal energy, E^{MM}_{PBC} is computed using polarizable force field via Ewald summation. The one-body and short-range two-body MM contributions are subtracted from the full MM energy and replaced with their QM counterparts. Specifically, the E_i's are the monomer energies and the $\Delta^2 E_{ij(0)}$'s are the short-range pairwise interaction energies between monomers i and j in the central unit cell. The $\Delta^2 E_{ij(\mathbf{n})}$'s are the short-range pairwise interaction energies between monomer i in the central unit cell and the periodic image monomer j in the $\mathbf{n}^{th}$ image cell. This image cell has been shifted by $n_{v_1}\mathbf{v_1} + n_{v_2}\mathbf{v_2} + n_{v_3}\mathbf{v_3}$ from the central unit cell, where $\mathbf{v_1}$, $\mathbf{v_2}$ and $\mathbf{v_3}$ are the unit cell vectors and n_{v_1}, n_{v_2} and n_{v_3} are the components of $\mathbf{n}$.

The interpolating function d_{ij} ensures a smooth transition between the short-range QM and long-range MM pairwise interactions by smoothly interpolating between the two regimes over the interval defined by c_1 and c_0:

$$d_{ij}(R) = \begin{cases} 1 & \text{if } R \leq r_1 \\ \frac{1}{1+e^{2|c_1-c_0|/(c_1-R)-|c_1-c_0|/(R-c_0)}} & \text{if } c_1 < R < c_0 \\ 0 & \text{if } R \geq c_0 \end{cases} \tag{4.3}$$

where R is defined as the shortest interatomic distance between the two molecules in the dimer. The parameters c_1 and c_0 are user-defined. We typically choose fairly conservative values of 9 and 10 Å, respectively, though more aggressive cutoffs can sometimes be used.[14] At these distances, the QM and MM pairwise interaction energies are nearly identical.

Finally, one must choose appropriate model chemistries for the QM and MM

terms. Any electronic structure method, such as MP2 or coupled cluster theory can be used to evaluate the QM monomer and pairwise interaction terms. The MM force field must be polarizable to capture the many-body induction effects, though many-body dispersion can also be important.[12, 14, 124, 125, 104] Here, we use the Amoeba force field, which in this context includes two-body electrostatics, induction and dispersion, and many-body induction. Even better results may be obtained when the force field is parameterized on-the-fly from quantum mechanical calculations.[10, 12, 119]

4.3.2 Gradient of the HMBI Energy

Fully relaxing a crystal structure requires one to optimize both the positions of each atom in the unit cell and the overall cell lattice parameters. The lattice cell parameters can be represented either in terms of three 3-dimensional vectors $\mathbf{v_1}$, $\mathbf{v_2}$ and $\mathbf{v_3}$, or, to eliminate the arbitrary overall rotation of the cell, as three lattice constants (a, b, c) and three angles (α, β, and γ). We adopt the six-parameter form, with the convention that the a axis is coincident with the x axis in the global coordinate system, the b axis lies in the xy plane of the global coordinate system, and the c axis lies anywhere in xyz. With this convention, the $\mathbf{v_1}$, $\mathbf{v_2}$ and $\mathbf{v_3}$ row vectors are expressed in terms of a, b, c, α, β and γ according to:

$$\begin{pmatrix} \mathbf{v_1} \\ \mathbf{v_2} \\ \mathbf{v_3} \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ b\cos\gamma & b\sin\gamma & 0 \\ c\cos\beta & c\frac{\cos\alpha-\cos\beta\cos\gamma}{\sin\gamma} & c\frac{\sqrt{\sin^2\gamma-\cos^2\beta+2\cos\alpha\cos\beta\cos\gamma}}{\sin\gamma} \end{pmatrix} \tag{4.4}$$

Moreover, we represent the atomic positions in absolute Cartesian, rather than fractional, coordinates.

With these conventions, the overall gradient for a unit cell containing N atoms

has length $3N+6$. Because the HMBI energy expression is linearly additive and does not involve any embedding potentials, the derivatives can be computed easily. We require only the gradients of individual monomers and dimers computed at both the QM and MM levels, along with the MM gradient for the full crystal. We assume that the basic force-field parameters are geometry independent, as is standard in molecular mechanics.

The first $3N$ terms in the gradient correspond to the derivative of the total energy with respect to the q-th atomic Cartesian coordinate. They are given by:

$$\begin{aligned}\frac{\partial E^{HMBI}}{\partial q} &= \frac{\partial E^{MM}_{PBC}}{\partial q} + \sum_i \left(\frac{\partial E^{QM}_i}{\partial q} - \frac{\partial E^{MM}_i}{\partial q} \right) \\ &+ \sum_{ij(0)} \frac{\partial d_{ij(0)}}{\partial q} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(0)} \right) + \sum_{ij(0)} d_{ij(0)} \left(\frac{\partial \Delta^2 E^{QM}_{ij(0)}}{\partial q} - \frac{\partial \Delta^2 E^{MM}_{ij(0)}}{\partial q} \right) \\ &+ \frac{1}{2} \sum_i \sum_{j(\mathbf{n})}^{images} \frac{\partial d_{ij(\mathbf{n})}}{\partial q} \left(\Delta^2 E^{QM}_{ij(\mathbf{n})} - \Delta^2 E^{MM}_{ij(\mathbf{n})} \right) \\ &+ \frac{1}{2} \sum_i \sum_{j(\mathbf{n})}^{images} d_{ij(\mathbf{n})} \left(\frac{\partial \Delta^2 E^{QM}_{ij(\mathbf{n})}}{\partial q} - \frac{\partial \Delta^2 E^{MM}_{ij(\mathbf{n})}}{\partial q} \right)\end{aligned} \tag{4.5}$$

The individual monomer and dimer nuclear derivatives needed to evaluate this overall gradient expression can be computed with standard electronic structure and molecular mechanics packages. The gradient of the QM to MM transition interpolating function is given by:

$$\begin{aligned}\frac{\partial d_{ij}}{\partial q^i_l} &= -\frac{\partial d_{ij}}{\partial q^j_{l'}} \\ &= -\delta_{lm}\delta_{l'm'} \frac{e^{g(c_1,c_0,R_{ij})} \left(q^i_l - q^j_{l'} \right) \left(\frac{1}{(c_0 - R_{ij})^2} + \frac{2}{(c_1 - R_{ij})^2} \right) |c_0 - c_1|}{\left(1 + e^{g(c_1,c_0,R_{ij})} \right)^2 R_{ij}}\end{aligned} \tag{4.6}$$

where $g\left(c_1, c_0, R_{ij}\right) = 2|c_1 - c_0| / \left(c_1 - R_{ij}\right) - |c_1 - c_0| / \left(R_{ij} - c_0\right)$, q^i_l is the q (= x, y or z) coordinate of the l^{th}-atom in monomer i, R_{ij} is the shortest distance between the monomers i and j, m and m' are the atoms in monomers i and j respectively which define R_{ij}, and δ is the Kronecker delta.

The remaining six elements in the gradient correspond to derivatives with respect to the six lattice parameters (a, b, c, α, β, and γ). When evaluating these derivatives, it proves convenient to break up the total HMBI energy into two pieces: the MM contribution for the full periodic crystal and the remaining terms involving the monomer and dimer energies,

$$E^{HMBI} = E^{MM}_{PBC} + E' \tag{4.7}$$

so that the partial derivative of the total energy with respect to one of the six lattice parameters u is given by,

$$\frac{\partial E^{HMBI}}{\partial u} = \frac{\partial E^{MM}_{PBC}}{\partial u} + \frac{\partial E'}{\partial u} \tag{4.8}$$

Since E' depends on the nuclear coordinates q_k of all atoms in the central unit cell and the unit cell vectors $\mathbf{v_1}$, $\mathbf{v_2}$ and $\mathbf{v_3}$, its gradient with respect to lattice parameter u is given by:

$$\frac{\partial E'}{\partial u} = \sum_{\epsilon=1}^{3} \sum_{q=x,y,z} \frac{\partial E'}{\partial v_{\epsilon q}} \frac{\partial v_{\epsilon q}}{\partial u} + \sum_{k=0}^{N} \frac{\partial E'}{\partial q_k} \frac{\partial q_k}{\partial u} \tag{4.9}$$

According to the rotation convention used to define the unit cell vectors $\mathbf{v_1}$, $\mathbf{v_2}$ and $\mathbf{v_3}$, $\frac{\partial v_{1y}}{\partial a} = \frac{\partial v_{1z}}{\partial a} = \frac{\partial v_{2z}}{\partial b} = 0$. Since we use Cartesian coordinates instead of fractional coordinates, the q_k are independent of the unit cell parameters, $\frac{\partial q_k}{\partial u} = 0$. Therefore, Eq (4.9) reduces to

$$\frac{\partial E'}{\partial u} = \frac{\partial E'}{\partial v_{1x}} \frac{\partial v_{1x}}{\partial u} + \frac{\partial E'}{\partial v_{2x}} \frac{\partial v_{2x}}{\partial u} + \frac{\partial E'}{\partial v_{2y}} \frac{\partial v_{2y}}{\partial u} + \frac{\partial E'}{\partial v_{3x}} \frac{\partial v_{3x}}{\partial u} + \frac{\partial E'}{\partial v_{3y}} \frac{\partial v_{3y}}{\partial u} + \frac{\partial E'}{\partial v_{3z}} \frac{\partial v_{3z}}{\partial u} \tag{4.10}$$

The $\frac{\partial v_{\epsilon q}}{\partial u}$ terms on the right-hand-side of Eq (4.10) are computed by taking partial derivatives of the x-, y- and z-components of the unit cell vectors $\mathbf{v_1}$, $\mathbf{v_2}$ or $\mathbf{v_3}$ with respect to a particular unit cell parameter, u (a, b, c, α, β or γ) using Eq (4.4).

Each $\frac{\partial E'}{\partial v_{\epsilon q}}$ term in Eq (4.10) is given by:

$$\frac{\partial E'}{\partial v_{\epsilon q}} = \frac{1}{2} \sum_{i(0)} \sum_{j(n_v)} n_v \sum_{k} \left\{ \frac{\partial d_{i(0)j(n_v)}}{\partial q_k} \left(\left(E^{QM}_{i(0)j(n_v)} - E^{QM}_{i(0)} - E^{QM}_{j(n_v)} \right) - \left(E^{MM}_{i(0)j(n_v)} - E^{MM}_{i(0)} - E^{MM}_{j(n_v)} \right) \right) + d_{i(0)j(n_v)} \left(\left(\frac{\partial E^{QM}_{i(0)j(n_v)}}{\partial q_k} - \frac{\partial E^{QM}_{j(n_v)}}{\partial q_k} \right) - \left(\frac{\partial E^{MM}_{i(0)j(n_v)}}{\partial q_k} - \frac{\partial E^{MM}_{j(n_v)}}{\partial q_k} \right) \right) \right\} \tag{4.11}$$

Index i sums over monomers in the central unit cell, j sums over periodic image monomers that lie within the short-range QM distance cutoffs for a given monomer i, and k sums over the atoms in monomer j. Also, q_k is the x, y, or z coordinate of the k-th atom in the image monomer j in the n_v-th image cell along the $\tilde{\mathbf{v}}$-vector. The factor n_v arises from the fact that $\frac{\partial E'}{\partial v_\epsilon} = \sum \frac{\partial E'}{\partial q} \frac{\partial q}{\partial v_{\epsilon q}}$ and $\frac{\partial q}{\partial v_{\epsilon q}} = n_v$. Note that the use of Cartesian coordinates instead of fractional coordinates means that monomers and dimers that lie only within the central unit cell contribute nothing to the lattice parameter gradient terms. Only dimer terms involving one periodic image molecule contribute.

The gradient of E^{MM}_{PBC} with respect to the six lattice parameters can be obtained in two ways. If analytic gradients are available, one could use those directly. Alternatively, because there are only six unit-cell parameters and the MM calculations are typically inexpensive compared to the quantum calculations in our model, they can also be obtained easily using finite difference. We use the latter approach here.

Combining all these pieces, the overall derivatives of the HMBI energy with respect to the six unit cell parameters are:

$$\frac{\partial E^{HMBI}}{\partial a} = \frac{\partial E^{MM}_{PBC}}{\partial a} + \frac{\partial E'}{\partial v_{1x}} \tag{4.12}$$

$$\frac{\partial E^{HMBI}}{\partial b} = \frac{\partial E^{MM}_{PBC}}{\partial b} + \frac{\partial E'}{\partial v_{2x}} \cos\gamma + \frac{\partial E'}{\partial v_{2y}} \sin\gamma \tag{4.13}$$

$$\frac{\partial E^{HMBI}}{\partial c} = \frac{\partial E^{MM}_{PBC}}{\partial c} + \cos\beta\left(\frac{\partial E'}{\partial v_{3x}} - \frac{\partial E'}{\partial v_{3y}}\cot\gamma\right) + \frac{\partial E'}{\partial v_{3y}}\cos\alpha\csc\gamma + \frac{\partial E'}{\partial v_{3z}}\sqrt{1 + 2\cos\beta\cot\gamma\csc\gamma - \cos^2\alpha\csc^2\gamma - \cos^2\beta\csc^2\gamma} \tag{4.14}$$

$$\frac{\partial E^{HMBI}}{\partial \alpha} = \frac{\partial E^{MM}_{PBC}}{\partial \alpha} + c\csc\gamma\sin\alpha\left(-\frac{\partial E'}{\partial v_{3y}} + \frac{\partial E'}{\partial v_{3z}}\frac{\cos\alpha - \cos\beta\cos\gamma}{\sqrt{\sin^2\gamma - \cos^2\alpha - \cos^2\beta + 2\cos\alpha\cos\beta\cos\gamma}}\right) \tag{4.15}$$

$$\frac{\partial E^{HMBI}}{\partial \beta} = \frac{\partial E^{MM}_{PBC}}{\partial \beta} + c\sin\beta\left(-\frac{\partial E'}{\partial v_{3x}} + \frac{\partial E'}{\partial v_{3y}}\cot\gamma + \frac{\partial E'}{\partial v_{3z}}\frac{\cos\alpha - \cos\beta\cos\gamma}{\sqrt{\sin^2\gamma - \cos^2\alpha - \cos^2\beta + 2\cos\alpha\cos\beta\cos\gamma}}\right) \tag{4.16}$$

$$\frac{\partial E^{HMBI}}{\partial \gamma} = \frac{\partial E^{MM}_{PBC}}{\partial \gamma} + b\cos\gamma\frac{\partial E'}{\partial v_{2y}} + c\csc\gamma\left(\cos\beta\csc\gamma - \cos\alpha\cot\gamma\right)\frac{\partial E'}{\partial v_{3y}} - b\sin\gamma\frac{\partial E'}{\partial v_{2x}} + c\csc^2\gamma\frac{\partial E'}{\partial v_{3z}}\frac{\cos^2\alpha\cos\gamma + \cos^2\beta\cos\gamma - 1/2\cos\alpha\cos\beta\left(3 + \cos 2\gamma\right)}{\sqrt{\sin^2\gamma - \cos^2\alpha - \cos^2\beta + 2\cos\alpha\cos\beta\cos\gamma}} \tag{4.17}$$

The computationally intensive steps of computing the dimer interaction energies and nuclear gradients can easily be run in parallel across many computer processors, with each individual dimer calculation distributed to a different processor (or processor group).

4.4 Computational Details

To test this model and compare it against force field and DFT predictions, full molecular geometry optimizations were performed for five molecular crystals: ice, formamide, acetamide, imidazole, and benzene. These five crystals include a diverse

range of intermolecular interactions: the first three are held together through hydrogen bonding, while benzene relies on van der Waals dispersion interactions, and imidazole contains both hydrogen-bonding and dispersion interactions. Experimental, low-temperature crystal structures exist for all five crystals.

The HMBI fragment model accurately predicts the lattice energies of these five crystals, at least when high-level QM methods are used. The best lattice energy results are obtained when estimated complete-basis set MP2 or CCSD(T) QM calculations were coupled to an *ab initio* force field (AIFF) that is parameterized on the fly,[12] but the Amoeba polarizable force field[41] has parameters for these five species and also works well in this context.[11] Here, we use the Amoeba force field for the MM contributions in HMBI when optimizing the crystals. As shown below, the optimized structures are in good agreement with experiment. Slightly better results might be expected if the AIFF were used instead, but implementation of its more complicated gradients is on-going.

For HMBI, the QM monomer and dimer calculations were performed at the counterpoise-corrected, resolution-of-the-identity second-order Møller-Plesset (RI-MP2) theory[42, 43, 44] with the Dunning aug-cc-pVDZ basis set[45] and its corresponding auxiliary set[126] using a developmental version of the Q-Chem[47] package. The Amoeba MM calculations were carried out using the Tinker[48] software package. For four of the crystals, we use conservative QM to MM transition distances of $c_1 = 9$ and $c_0 = 10$ Å. For ice, almost identical energies are obtained even when the transition distance is shortened to $c_1 = 6$ and $c_0 = 7$ Å, so we use the shorter cutoff. See Refs. [11] and [14] for further discussion of these cutoffs. The HMBI energy and gradient calculations were performed using a parallel implementation of our code that distributes the individual monomer and dimer calculations over a set of processors. Geometry optimizations were performed using L-BFGS[127] as implemented in the open-source optimizer DL-FIND

[49].

For each crystal, geometry optimization with HMBI is done using two different approaches. First, unless otherwise specified, we partially enforce space group symmetry by freezing all symmetry-determined (90 and 120 degree) angular lattice parameters during the optimization (to ensure the correct Bravais lattice type). In the future, it would be desirable to implement a proper treatment of space group symmetry to ensure strict adherence to the overall space group and to reap the associated computational savings. Second, we also perform fully relaxed optimizations, with no constraints. This case mimics the true crystal structure prediction problem, where the space group symmetry is generally unknown in advance. The ability to obtain the correct symmetry is an important criterion for crystal structure prediction.

The HMBI optimized lattice parameters and atomic coordinates are compared with those from the purely classical Amoeba force field and from dispersion-corrected periodic density functional theory. The DFT crystal structure calculations were performed with CRYSTAL09[128, 129] using the B3LYP-D* functional,[92] which includes a Grimme-type empirical dispersion correction.[130] Two different basis sets, 6-31G(d,p)[131] and the triple zeta quality Ahlrichs' TZP[132] were used for each crystal. These basis sets were chosen to avoid the numerical troubles that often occur with diffuse Gaussian-type basis sets in periodic crystal calculations. Full space-group symmetry was employed in these DFT calculations. A pruned grid containing 75 angular and 974 Lebedev angular points was employed to ensure accurate gradients. Due to a software bug in how CRYSTAL09 v1.0 handles the dispersion correction gradient in certain space groups, the acetamide DFT optimizations used finite difference to compute the lattice parameter gradients. The other optimizations used fully analytic gradients. All optimized crystal structures are provided as Supplementary information.[133]

Finally we calculated lattice energies for each of the optimized structures. For HMBI, we combine our recent high-level benchmark energies[12] on partially optimized structures with new corrections due to the full structure relaxations performed here (see Section 4.5.4). For DFT, counterpoise-corrected single-point B3LYP-D* lattice energies were calculated on the DFT-optimized structures using over 200 ghost atoms.

4.5 Results and Discussion

4.5.1 Experimental Crystal Structures

For all crystals except ice, the experimental crystal structures were taken from the Cambridge Structural Database (CSD). To minimize the potential discrepancies between predicted and experimental structures due to thermal expansion, we selected the lowest-temperature structure available. The CSD reference codes for the selected crystals are ACEMID05 (acetamide),[134] BENZEN14 (benzene),[135] FORMAM02 (formamide),[136] and IMAZOL06 (imidazole).[137] For acetamide, the ACEMI- -D05 refinement based on combined x-ray and neutron scattering data was used over the earlier structure[138] at the same temperature.

For icosahedral ice, which is a disordered crystal, we adjusted the sixteen-molecule non-polar supercell of Morrison et al[139] to the experimental lattice parameters and atomic positions to their values at 15 K.[140, 141] Note that the ice a parameter reported here corresponds to this supercell, and it is twice the crystallographic unit cell parameter. The disordered protons reduce the overall symmetry of the ice unit cell, so the DFT optimizations were performed in the P_1 space group with the constraint that the cell remains orthorhombic. Other implications of the disorder are discussed in Section 4.5.3.

Table 4.1: Optimized lattice parameters and unit-cell volumes.

	Amoeba	B3LYP-D* 6-31G(d,p)	B3LYP-D* TZP	HMBI MP2 aug-cc-pVDZ	Expt.
Ice (Expt @ T=15 K)					
a (Å)	9.075 (0.9%)	8.717 (-3.1%)	8.950 (-0.5%)	9.099 (1.2%)	8.994
b (Å)	7.740 (-0.6%)	7.546 (-3.1%)	7.709 (-1.0%)	7.816 (0.3%)	7.789
c (Å)	7.344 (0.3%)	7.040 (-3.8%)	7.284 (-0.5%)	7.389 (0.9%)	7.321
V (Å^3)	515.8 (0.6%)	463.1 (-9.7%)	502.6 (-2.0%)	525.5 (2.5%)	512.9
Formamide (Expt @ T=90 K)					
a (Å)	3.549 (-1.5%)	3.533 (-2.0%)	3.621 (0.5%)	3.580 (-0.7%)	3.604
b (Å)	9.341 (3.3%)	8.908 (-1.5%)	9.029 (-0.1%)	9.459 (4.6%)	9.041
c (Å)	6.745 (-3.5%)	6.923 (-1.0%)	6.933 (-0.9%)	6.975 (-0.3%)	6.994
β (°)	104.5 (3.9%)	109.5 (8.9%)	107.2 (6.6%)	99.6 (-0.9%)	100.5
V (Å^3)	216.6 (-3.4%)	205.4 (-8.3%)	216.6 (-3.3%)	232.9 (3.9%)	224.1
Acetamide (Expt @ T=23 K)[a]					
a (Å)	11.556 (0.6%)	11.431 (-0.5%)	11.422 (-0.6%)	11.646 (1.3%)	11.492
b (Å)	11.556 (0.6%)	11.431 (-0.5%)	11.422 (-0.6%)	11.648 (1.4%)	11.492
c (Å)	12.317 (-4.5%)	12.305 (-4.6%)	12.725 (-1.3%)	12.971 (0.6%)	12.892
V (Å^3)	1424.3 (-3.4%)	1392.5 (-5.6%)	1437.6 (-2.5%)	1525.8 (3.5%)	1474.5
Benzene (Expt @ T=4 K)					
a (Å)	6.973 (-5.2%)	7.350 (-0.1%)	7.368 (0.2%)	7.263 (-1.3%)	7.355
b (Å)	9.591 (2.3%)	9.195 (-1.9%)	9.359 (-0.1%)	9.362 (-0.1%)	9.371
c (Å)	6.866 (2.5%)	6.582 (-1.7%)	6.681 (-0.3%)	6.598 (-1.5%)	6.699
V (Å^3)	459.2 (-0.6%)	444.9 (-3.7%)	460.7 (-0.2%)	448.6 (-2.8%)	461.8
Imidazole (Expt @ T=103 K)					
a (Å)	7.358 (-2.8%)	7.714 (1.9%)	7.777 (2.7%)	7.370 (-2.6%)	7.569
b (Å)	5.253 (-2.1%)	5.067 (-5.6%)	5.224 (-2.6%)	5.282 (-1.6%)	5.366
c (Å)	10.03 (2.5%)	9.763 (-0.2%)	9.759 (-0.3%)	9.801 (0.2%)	9.785
β (°)	116.9 (-1.9%)	120.5 (1.2%)	120.5 (1.2%)	120.0 (0.8%)	119.08
V (Å^3)	352.2 (1.4%)	328.7 (-5.4%)	341.6 (-1.6%)	330.5 (-4.9%)	347.3
RMS[b]	2.8%	3.4%	2.0%	1.6%	

[a] Experimentally, $a = b$ in acetamide ($R3c$ space group), but our HMBI software currently does not enforce this symmetry constraint.

[b] Root-mean-square error of lattice parameters (excludes cell volumes)

4.5.2 Structure Optimization

We begin by examining the optimized lattice parameters calculated with DFT, Amoeba, and HMBI (MP2/aug-cc-pVDZ). Comparing against experimental structures can be difficult, since thermal expansion can change crystal structure parameters by up to a few percent between 0 K (predictions) and the finite experimental temperature at which the structure was measured.[142] Fortunately, the comparisons here are facilitated by the fact that all five experimental crystal structures used here were determined at low temperatures ($\sim$100 K or less).

As shown in Table 4.1, all methods perform reasonably well for many of the lattice parameters. The Amoeba force field performs particularly well for ice, where all three lattice parameters are predicted to within less than 1%. Its performance for the other crystals is also quite good, though the errors in different crystals vary rather broadly. Still, as will be seen below, it performs about as well as B3LYP-D*/6-31G(d,p). Given that the Amoeba force field parameters were not optimized for the solid state, its overall root-mean-square (RMS) error of 2.8% in the lattice parameters that are not fixed by symmetry is impressive.

Dispersion-corrected B3LYP-D* optimizations were performed using two different basis sets, and we observe that dispersion-corrected DFT performs quite well for structures. As expected, the larger TZP basis performs significantly better than 6-31G(d,p). For instance, the reduction in basis-set superposition error (BSSE) in the more complete TZP basis leads to larger unit cells (Counterpoise corrections were not employed during the DFT optimization), and the corresponding errors in the cell volumes decrease by a factor of two or more. Typical DFT-D lattice parameter errors are on the order a few percent or less, which is consistent with previous benchmark

studies.[92, 95] The β angle in formamide is the most notable exception, with errors of 8.9% and 6.6% in the 6-31G(d,p) and TZP basis sets, respectively. In the TZP basis, the imidazole a and b parameters show errors of 2.6-2.7%, while the rest of the lattice parameters are predicted to within ∼1% or less. For comparison, the Amoeba-optimized structures are roughly on par with those calculated at the B3LYP-D*/6-31G(d,p) level.

Next, consider the HMBI MP2/aug-cc-pVDZ level structures. Taken as a whole, the lattice parameters predicted at the HMBI MP2 level are significantly more accurate than the B3LYP-D*/6-31G(d,p) predictions and almost as good as the B3LYP-D*/TZP results, as can be seen in the box plots in Figure 4.1. For all but two lattice parameters, the HMBI MP2 errors are less than 2%. The HMBI MP2-predicted a lattice parameter for ice is very close to that determined by Hermann and Schwerdtfeger when they combined periodic HF with an MP2 treatment of short-range dimer interactions.[117]

The largest HMBI MP2 errors occur for formamide (4.6% in b) and imidazole (2.6% in a and 1.6% in b) and are smaller than the largest 6.6% B3LYP-D*/TZP error (for the β angle in formamide). The artifactual expansion along the b crystallographic axis in formamide with HMBI MP2 may reflect the Amoeba force field's systematic underestimation of many-body polarization effects[10] along the crystal's hydrogen-bonded chains. Like for the B3LYP-D*/TZP results, the other lattice parameters are predicted to within ∼1% or less.

Of course, MP2 exhibits well-known sensitivities to basis set, and the aug-cc-pVDZ basis set used here is fairly small. The inclusion of Counterpoise corrections during the HMBI geometry optimizations mitigates this sensitivity somewhat. The ease of including Counterpoise corrections[143] is one of the strengths of fragment methods. In benzene, the MP2-level structures are in good agreement with experiment despite

the known problems MP2 has in treating dispersion. While MP2 significantly overestimates the binding energy of the benzene dimer[144] and crystal[145, 11] Counterpoise-corrected MP2/aug-cc-pVDZ fortuitously predicts the optimal gas-phase benzene dimer intermolecular separation to within 0.1–0.2 Å of the estimated CCSD(T)/aug-cc-pVQZ result.[144] Similar factors presumably account for the good HMBI MP2 performance in imidazole as well.

Interestingly, the errors in specific lattice parameters between B3LYP-D* and HMBI MP2 appear uncorrelated. For example, whereas B3LYP-D* overestimates the β angle in formamide by 7-9°, HMBI MP2 predicts that bond angle to within less than 1°. On the other hand, HMBI MP2 significantly overestimates the b lattice parameter in the same crystal, while B3LYP-D* has no difficulty with that particular parameter. Similar variability occurs for the other crystals as well.

Furthermore, whereas the predicted B3LYP-D* volumes underestimate the experimental volumes in all cases tested here, the HMBI MP2 volumes sometimes overestimate the experimental volume. On the one hand, crystals typically expand upon heating, so one expects volumes predicted at 0 K to be smaller than those at finite temperatures, though the low temperatures at which these experimental structures were obtained means that thermal expansion effects will be small. On the other hand, the DFT optimizations will spuriously over-bind the crystals due to BSSE, which might shrink the lattices. In contrast, Counterpoise correction used in the HMBI MP2 optimizations reduces this effect.

Comparison of the atomic coordinates and structure overlays provide further perspective on these optimized structures. Figure 4.2 plots the root-mean-square deviations of the non-hydrogen atoms between the predicted and experimental structures.

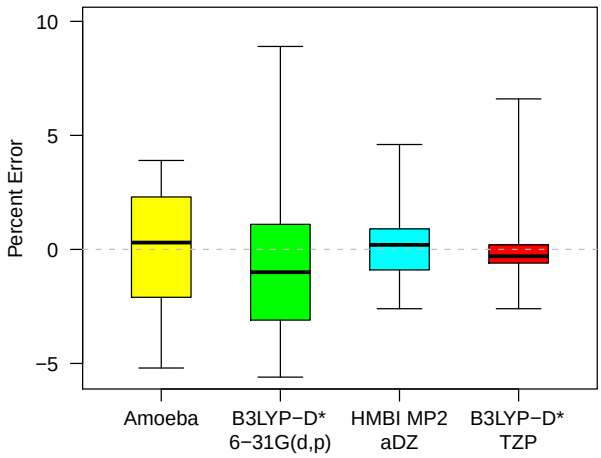


Figure 4.1: Box plots showing the median (dark line), middle 50% (colored boxes), and full range of the errors (whiskers) in the optimized lattice parameters of the five crystals relative to experiment.

These $rmsd_{15}$ values[146] were computed in Mercury 3.0[147] by overlaying a central molecule and its fifteen nearest neighbors. With the exception of benzene, the quality of the atomic coordinates from Amoeba and B3LYP-D*/6-31G(d,p) are comparable. Increasing the DFT basis to TZP improves the results significantly. In terms of overall accuracy, once again the HMBI MP2/aug-cc-pvdz $rmsd_{15}$ values are better than those from B3LYP-D*/6-31G(d,p) and almost as good as those from B3LYP-D*/TZP. However, the HMBI MP2 values exhibit more uniform errors than those from B3LYP-D*/TZP. Error consistency is particularly valuable when assessing predictions for unknown structures. The structure overlays in Figure 4.3 tell a similar story to the lattice parameter and $rmsd_{15}$ errors: the B3LYP-D*/TZP and HMBI MP2 structures are difficult to distinguish from the experimental ones visually. The B3LYP-D*/6-31g(d,p) structures, which are omitted for the sake of clarity, overlap slightly less well with experiment.

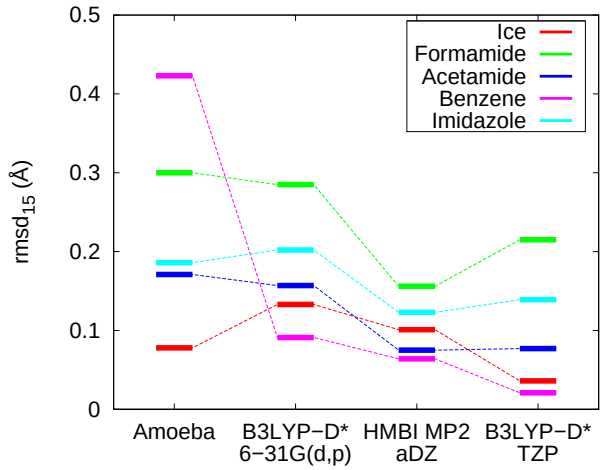


Figure 4.2: Root-mean-square deviations (rmsd_{15}) in the optimized non-hydrogen atom coordinates relative to experiment.

4.5.3 Crystal Symmetry

In all of these HMBI MP2 calculations described above, symmetry was partially enforced. Specifically, all 90° unit cell angles were held fixed. In *ab initio* crystal structure prediction, however, one typically does not know the space-group symmetry in advance. For this reason, we also optimized all the crystals with HMBI without any constraints, to determine how well they reproduce the proper space-group symmetry.

In most cases, these optimizations were performed from the experimental lattice parameters. However, for acetamide, where the large unit cell made optimization from the experimental structure slow and computationally expensive, we simply distorted the lattice angles slightly from the symmetry-constrained structure. In that case, it quickly re-converged to the symmetric structure to within numerical accuracy.

The fully relaxed HMBI MP2 structures very nearly reproduce the correct Bravais lattice, as shown in Table 4.2. Most of the optimized angles deviate from the ideal cell angles by no more than a couple tenths of a degree. This additional relaxation

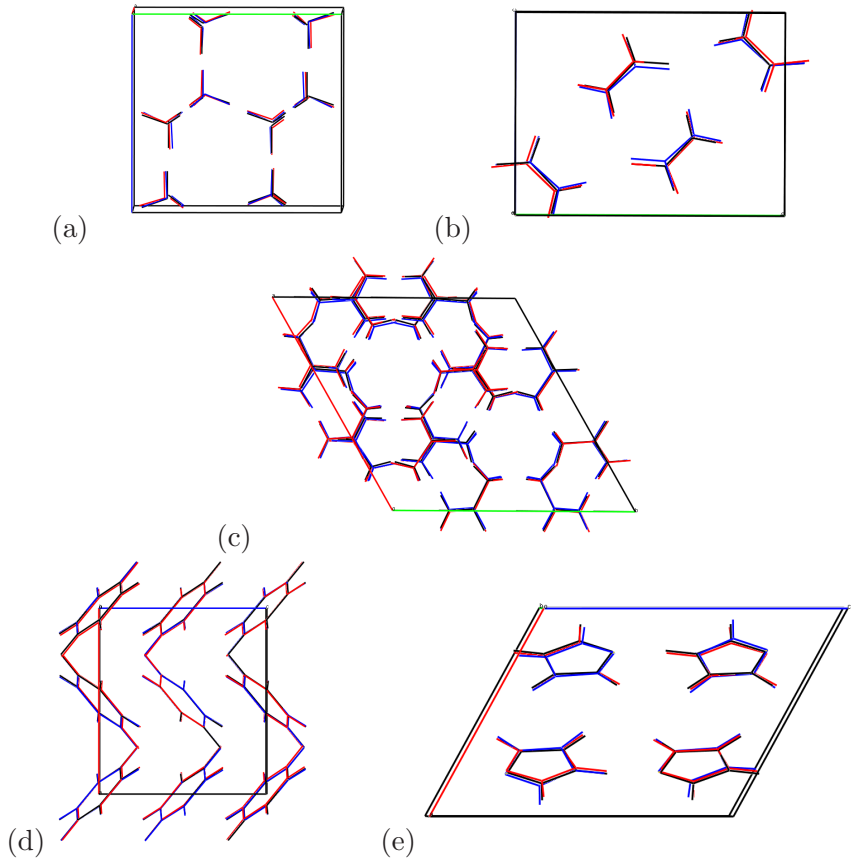


Figure 4.3: Overlays of the experimental (black), B3LYP-D*/TZP (blue) and HMBI MP2 (red) structures for (a) ice, (b) formamide, (c) acetamide, (d) benzene, and (e) imidazole. For clarity, only the experimental cell boundaries are drawn.

also typically introduces small changes to the other structure parameters. The cell lengths change by only a few hundredths of an Angstrom, the angles not determined by symmetry usually change by a few tenths of a degree, and the predicted volumes change by 0.2% or less.

Similarly, for the most part, the atomic positions maintain the overall space group symmetry, regardless of whether or not the cell angles were constrained during optimizations. The tolerances (defined as the maximum atomic deviation, in Å) with which they reproduce the appropriate space group is shown in the last column of Ta-

ble 4.2. For all crystals except ice, these tolerances lie within a reasonable ∼0.1 Å.

Ice provides an interesting case. The idealized experimental structure represents an averaged structure, but the disorder means that individual water molecules adjust to their local environments. This has two implications here. First, regardless of whether or not the cell is held orthorhombic, individual water molecules rotate and shift off their idealized lattice parameters, leading to a reduction in symmetry. The experimental $P6_3/mmc$ space-group symmetry is only maintained for the average oxygen positions. This explains the large 0.3 Å symmetry tolerances in Table 4.2. Nevertheless, Figure 4.3(a) shows the overall good agreement between the experimental, HMBI MP2, and B3LYP-D* structures.

Second, the lattice distorts once the orthorhombic Bravais lattice constraint is lifted, and we observe a 1.6° deviation in the α angle. Performing the same symmetry-free P1 space-group optimization with B3LYP-D*/TZP produces similar distortions ($\alpha = 89.1°$, $\beta = 89.9°$, and $\gamma = 90.0°$). This cell distortion reflects the use of a periodic supercell to approximate the disordered crystal rather than an inherent flaw in the electronic structure methods.

4.5.4 Lattice Energies

In our previous benchmark lattice energy calculations,[11, 12] we optimized the atomic positions within the unit cell while holding the lattice parameters fixed. We then used complete-basis set (CBS) MP2, smaller-basis CCSD(T) calculations, and the *ab initio force-field model* to estimate the complete-basis-set CCSD(T) lattice energies. Finally, to compensate for artificially constraining the lattice parameters at their experimental values, we estimated a relaxation energy correction $\Delta_{lattice}^{relax}$ based on a simple

Table 4.2: Effect of relaxing the symmetry constraints on the HMBI MP2/aug-cc-pVDZ optimized lattice parameters. (constr. = constrained, unconstr. = unconstrained, Sym. tol. = Symmetry tolerance)

	a(Å)	b(Å)	c(Å)	α(°)	β(°)	γ(°)	V(Å^3)	Sym. tol[a](Å)
Ice ($P6_3/mmc$)								
constr.	9.099	7.816	7.389	90	90	90	525.5	0.31[b]
unconstr.	9.124	7.788	7.394	88.63	89.89	90.14	525.2	0.34[b]
% change	-0.27%	0.36%	-0.07%	1.55%	0.12%	-0.16%	0.06%	
Formamide ($P2_1/n$)								
constr.	3.580	9.459	6.975	90	99.64	90	232.9	0.03
unconstr.	3.597	9.416	6.982	90.616	99.77	90.33	233.0	0.13
% change	-0.47%	0.46%	-0.10%	-0.68%	-0.13%	-0.37%	-0.07%	
Acetamide (R3c)								
constr.	11.646	11.648	12.971	90	90	120	1525.8	0.06
unconstr.	11.646	11.648	12.971	89.99	90.00	119.87	1525.8	0.06
% change	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%	
Benzene (Pbca)								
constr.	7.263	9.362	6.598	90	90	90	448.6	0.03
unconstr.	7.261	9.352	6.591	90.01	90.00	89.98	447.6	0.04
% change	0.03%	0.11%	0.11%	-0.01%	0.00%	0.02%	0.24%	
Imidazole ($P2_1/c$)								
constr.	7.370	5.282	9.801	90	119.99	90	330.5	0.07
unconstr.	7.377	5.291	9.832	89.8	120.37	89.97	331.1	0.04
% change	-0.09%	-0.17%	-0.32%	0.22%	-0.32%	0.03%	-0.19%	

[a] Tolerance with which the optimized structure reproduces the experimental space-group symmetry.

[b] Excludes hydrogen atoms.

model for isotropically expanding or contracting the lattice. Combining the single-point energies with this relaxation energy correction gave a best estimate for the lattice energy. In the end, the model predicted lattice energies for four of the five crystals here to within 1–2 kJ/mol of experiment. Only acetamide had a larger (6 kJ/mol) error, as summarized in Table 4.3.[12] The experimental errors in the crystals are probably as high as several kJ/mol in each case.[11]

With the fully optimized unit cells computed here, we can reassess those benchmark lattice energies. Because the estimated complete-basis-set CCSD(T) benchmarks required large amounts of computer time, we do not repeat the full calculations. Rather, we simply revise our estimates of the lattice parameter relaxation energies $\Delta_{lattice}^{relax}$ using the difference between the HMBI MP2/aug-cc-pVDZ energies computed on the old (unrelaxed lattice parameters) and new fully-relaxed geometries obtained here.

These new $\Delta_{lattice}^{relax}$ values are listed in Table 4.3, and they differ from the previous ones by 1 kJ/mol or less. Similarly, the revised best estimates for the lattice energies are very close to the previous values and in excellent agreement with experiment. Four of the five lattice energies lie within 1 kJ/mol of experiment, which is easily within the experimental error bars of a few kJ/mol. The fact that the crystal structures and lattice energies do not change significantly from the earlier results also supports the decision not to recompute the lattice energies directly on the new, fully optimized structures.

Once again, only acetamide exhibits a large error with HMBI. Based on the new results here, it appears that this unusually large error is not a lattice relaxation effect, as we previously speculated. Therefore, this error presumably stems from the relatively small 6-31+G* basis used to compute the small-basis CCSD(T) correlation correction in our previous work, the temperature extrapolation and zero-point energy estimates used to convert the experimental heat of sublimation to lattice energy, or the

Table 4.3: Predicted and experimental lattice energies, in kJ/mol.

	Ice	Form--amide	Acet--amide	Benzene	Imid--azole
HMBI, optimized with fixed lattice parameters (ref. [12])					
Est. CCSD(T)/CBS with AIFF	60.2	80.4	79.7	51.2	88.6
Old $\Delta_{lattice}^{relax}$	0.2	0.0	n/a	2.8	2.2
Previous HMBI best estimate	60.4	80.4	79.7	54.0	90.8
HMBI, fully optimized (this work)					
New $\Delta_{lattice}^{relax}$	0.2	1.1	0.6	1.8	1.3
Revised HMBI best estimate	60.4	81.5	80.3	53.0	89.9
B3LYP-D*/6-31G(d,p)	73.3	75.3	86.4	46.4	86.3
B3LYP-D*/TZP	64.7	77.0	88.2	45.8	86.9
Experiment (see ref. [11])	59	82 ± 0.3	86 ± 2	52 ± 3	91 ± 4

underlying experimental data.[11, 12] Table 4.3 also gives the counterpoise-corrected B3LYP-D* lattice energies. B3LYP-D*/TZP also performs fairly well, with errors 6 kJ/mol or less. With the exception of acetamide, those errors are several times larger than the best near-CCSD(T)-quality HMBI results in these crystals.

4.6 Conclusions

We presented an implementation of analytic nuclear gradients for the fragment-based HMBI model chemistry in terms of both the atomic positions and the lattice parameters. Using these gradients, we optimized five different molecular crystal structures at the HMBI MP2/aug-cc-pVDZ level. These MP2-quality structures improve significantly upon dispersion-corrected periodic B3LYP-D*/6-31G(d,p) structures and are almost as good as the B3LYP-D*/TZP ones. HMBI MP2 also predicts the correct space group symmetry for the crystals considered here. Overall, these results suggest that optimizing the structures at the HMBI MP2/aug-cc-pVDZ level offers modest bene-

fits compared to using a periodic dispersion-corrected density functional treatment. For instance, the errors in the MP2-level predicted atomic coordinates were more uniform across the different crystals than those from B3LYP-D*, and the fragment approach allows facile inclusion of Counterpoint corrections for BSSE during the optimization.

Moreover, fragment methods like HMBI offer the ability systematically converge predictions by improving the quality of the fragment wavefunctions. As noted above, predicted crystal lattice energies can be systematically improved to within experimental accuracy. In principle, even better optimized structures ought to be obtained using larger basis sets and/or higher-level correlation methods than those used here, though such geometry optimizations are very computationally expensive at this point. Further algorithmic improvements to exploit space-group symmetry will help, both by lowering the computational cost and by reducing the number of degrees of freedom that need to be optimized. Furthermore, we are currently implementing gradients of the more elaborate *ab initio force field* which will eliminate the need for pre-determined force field parameters and should provide additional accuracy. For instance, three-body dispersion contributions can be large in even some polar crystals,[12, 14, 125, 104] so it will be interesting to see what effects such terms have on the predicted structures.

In the gas phase, one turns to wavefunction-based methods when one needs accuracy beyond what DFT can provide. By making the full spectrum of wavefunction methods available, fragment methods will fill a similar role in the condensed phase. The ability to converge the results with respect to the model chemistry is crucial to making robust predictions in molecular crystals, especially those involving flexible molecules, where crystal polymorphs are often separated by small energy gaps that are highly dependent on the calculated structures.[15, 93, 94]

Chapter 5

Predicting Thermal and Vibrational Properties of Molecular Crystals With a Fragment-Based QM/MM Method

5.1 Outline

Studying the crystal properties of organic crystals experimentally is not always straightforward as these chemical systems suffer from polymorphism. Accurate theoretical predictions of these crystal properties have an important role supplementing experiments in screening these different polymorphs. Here, we present the formalism and benchmark results for the lattice dynamics implementation with the fragment-based

QM/MM method called Hybrid Many-Body Interaction (HMBI) method for predicting crystal properties like sublimation heats, phonon dispersion curves, phonon density of states, specific heats, etc. on test crystals like I_h ice and ammonia. We also emphasize the role played by such fragment-based methods for getting electronic structure quality predictions efficiently for these large chemical systems.

5.2 Introduction

Applications of organic crystals are especially affected by polymorphism[18, 19, 148] because of the weak intermolecular forces which make various packing motifs thermodynamically competing. A priori identification and screening of polymorphs for applications of organic crystals as pharmaceutical drugs, organic semi-conductors and energetic materials could potentially be achieved using tried and tested accurate theoretical tools. Accurate predictions of crystal properties of the various polymorphs is *prima facie* to this end. For example, it has been known that the 0-6 Terahertz region of the vibrational spectra, which in organic crystals is dominated by weak intermolecular interactions, is different for different polymorphic packing[149, 150, 151, 152, 153]. Potentially, with accurate predictions of these fingerprint region, theoretical methods can help distinguish between polymorphs. Furthermore, since there are no associated functional group frequencies in this frequency range, theoretical methods can assist experiments by assigning the vibrational modes. In some cases like crystalline glycine, the relative vibrational enthalpies could become important in establishing the correct stability order of the polymorphs at finite temperatures[154, 155].

However, efficient treatment of these large systems with gold-standard electronic structure methods like coupled cluster methods is unaffordable. Methods based on

density functional theory (DFT) have been the method-of-choice lately[156, 38, 39], but these are not systematically improvable[105]. Alternatively, fragment-based methods[3, 4, 5, 6, 7, 8, 9] have an advantage in studying these systems as these methods are systematically improvable.

Recently, we have shown that the hybrid many-body interaction (HMBI) method[9, 10, 11, 12, 13, 14, 15, 16] which is a fragment-based QM/MM method can predict crystal structures[16] and 0 K lattice energies of a wide variety of organic crystals accurately. However, in order to distinguish between competing polymorphs and assign correct thermodynamic polymorphic stability order at finite temperatures, one needs to account for the thermal effects like thermal expansion and the entropic contribution in the energetics. Moreover, reliable theoretical methods should be able to correctly predict measurable crystal properties like sublimation heats, specific heat capacities, vibrational spectra, etc. To this effect, we have implemented lattice dynamics[157] with the HMBI method.

Lattice dynamics implementations are well-known for predicting the vibrational spectra of the chemical systems, phase stability analysis of the optimized structures and phase change reactions. Using lattice dynamics, Strassle et al[158] have successfully simulated the softening of the transverse acoustic mode in I_h ice which helps explain the pressure-induced amorphization of this crystal under moderate pressures. Recently, lattice dynamics implementations similar to the one discussed here have been implemented in other fragment-based methods. One of these implementations has been used to study the effects of external pressure on the Phonon Density of States (PDOS) in hydrogen flouride[159] and to successfully reproduce the vibrational spectrum of the proton-disordered I_h ice phase consistent with the experimental inelastic neutron scattering spectrum[160]. Reilly and Tkatchenko[161] have done similar lattice dynamics calcula-

tions for predicting sublimation heats accurately using a DFT approach which also included a many-body dispersion effects. Meanwhile, Otero-de-la-Roza and Johnson[162] have performed accurate sublimation heats calculations on a number of crystals with various dispersion-corrected non-local density functionals, where the dispersion correction is treated using the electrostatic interaction of the electron-exchange hole dipoles in the system.

On similar lines, in this study we illustrate that accurate prediction of the vibrational spectrum and other crystal properties can be performed using our QM/MM type fragment-based HMBI approach. We show here lattice dynamics calculations for the following organic crystals: I_h ice, ammonia and imidazole. Ice and ammonia crystals are mostly hydrogen bonded crystals while imidazole shows both hydrogen bonds and significant dispersion effects. We also emphasize the computational cost advantage of using the HMBI method for lattice dynamics.

5.3 Theory

The HMBI method decomposes the system into fragments and treats the fragment and short-wise pairwise inter-fragment interactions with high level electronic structure methods while the long-range pairwise and many-body inter-fragment interactions are treated using classical polarizable force fields.

$$E^{HMBI} = E^{QM}_{1\text{-}body} + E^{QM}_{\substack{short\text{-}range \\ 2\text{-}body}} + E^{MM}_{\substack{long\text{-}range \\ 2\text{-}body}} + E^{MM}_{many\text{-}body} \tag{5.1}$$

The use of the force fields in dealing with expensive many-body interaction energies makes this method affordable for large chemical systems like organic crystals and molecular clusters. Moreover, polarizable force-fields can accurately capture the long-range electrostatic and induction effects. More importantly, the use of highly accurate elec-

tronic structure methods for describing the most important interactions allows predictions of lattice energies within the chemical accuracy of a kcal/mol. The HMBI energy formalism for a periodic system is given by

$$\begin{aligned} E^{HMBI} = E^{MM}_{PBC} + \sum_{i} \left(E^{QM}_{i} - E^{MM}_{i} \right) + \sum_{ij(0)} d_{ij(0)} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(0)} \right) \\ + \frac{1}{2} \sum_{i} \sum_{j(n_v)}^{images} d_{ij(n_v)} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(n_v)} \right) \end{aligned} \tag{5.2}$$

where one can consider the quantum mechanical treatment of the 1-body and short-range 2-body interactions as a correction to the classical treatment of the full crystal with force fields using Ewald sum. The factor of 1/2 arises in order to avoid double counting of the pairwise interactions between fragments within the central unitcell and image fragments. The damping functions d_{ij} for each pairwise interactions ensure a smooth spatial truncation from the quantum mechanical treatment to the classical treatment giving a continuously differentiable potential energy surface. Such a spatial truncation scheme usually introduces less than 0.1 kJ/mol errors in the lattice energies, mainly because the classical and quantum mechanical potential energy surfaces for a pairwise interaction converge in the long-range limit.

The first derivative of the HMBI energy with respect to any nuclear coordinate q_l within the central unit cell is given by

$$\begin{aligned} \frac{\partial E^{HMBI}}{\partial q_l} = \frac{\partial E^{MM}_{PBC}}{\partial q} + \sum_{i} \left(\frac{\partial E^{QM}_{i}}{\partial q_l} - \frac{\partial E^{MM}_{i}}{\partial q_l} \right) \\ + \sum_{ij} \frac{\partial d_{ij}}{\partial q_l} \left(\Delta^2 E^{QM}_{ij} - \Delta^2 E^{MM}_{ij} \right) + \sum_{ij} d_{ij} \left(\frac{\partial \Delta^2 E^{QM}_{ij}}{\partial q_l} - \frac{\partial \Delta^2 E^{MM}_{ij}}{\partial q_l} \right) \\ + \frac{1}{2} \sum_{i} \sum_{k}^{images} \frac{\partial d_{ik}}{\partial q_l} \left(\Delta^2 E^{QM}_{ik} - \Delta^2 E^{MM}_{ik} \right) \\ + \frac{1}{2} \sum_{i} \sum_{k}^{images} d_{ik} \left(\frac{\partial \Delta^2 E^{QM}_{ik}}{\partial q_l} - \frac{\partial \Delta^2 E^{MM}_{ik}}{\partial q_l} \right) \end{aligned} \tag{5.3}$$

Here, q is any x, y or z coordinate of the l-th atom in the central unit cell.

Similarly, the second derivative of the HMBI energy with respect to any two nuclear coordinates within the central unit cell is given by

$$\begin{aligned}\frac{\partial^2 E^{HMBI}}{\partial q_l \partial q_{l'}} = & \frac{\partial E^{MM}_{PBC}}{\partial q_l \partial q_{l'}} \\ & + \sum_{i(0)}^{Monomers} \left(\frac{\partial E^{QM}_{i(0)}}{\partial q_l \partial q_{l'}} - \frac{\partial E^{MM}_{i(0)}}{\partial q_l \partial q_{l'}} \right) \\ & + \sum_{i(0)j(n)}^{Dimers} \zeta d_{i(0)j(n)} \left(\frac{\partial^2 \left(\Delta^2 E^{QM}_{i(0)j(n)} \right)}{\partial q_l \partial q_{l'}} - \frac{\partial^2 \left(\Delta^2 E^{MM}_{i(0)j(n)} \right)}{\partial q_l \partial q_{l'}} \right) \\ & + \sum_{i(0)j(n)}^{Dimers} \zeta \frac{\partial d_{i(0)j(n)}}{\partial q_l} \frac{\partial \left(\Delta^2 E^{QM} - \Delta^2 E^{MM} \right)}{\partial q_{l'}} \\ & + \sum_{i(0)j(n)}^{Dimers} \zeta \frac{\partial d_{i(0)j(n)}}{\partial q_{l'}} \frac{\partial \left(\Delta^2 E^{QM} - \Delta^2 E^{MM} \right)}{\partial q_l} \\ & + \sum_{i(0)j(n)}^{Dimers} \zeta \frac{\partial^2 d_{i(0)j(n)}}{\partial q_l \partial q_{l'}} \left(\Delta^2 E^{QM} - \Delta^2 E^{MM} \right)\end{aligned} \tag{5.4}$$

Here, $\zeta = 1$ for pairwise interactions of two fragments both within the central unit cell, while $\zeta = 0.5$ for pairwise interactions of a fragment within the central unit cell with an image fragment.

The Γ-point harmonic vibrational frequencies can be recovered by mass-weighting and then diagonalizing this HMBI hessian matrix. However, recovering the harmonic frequencies at other reciprocal space points of the first Brillouin zone requires decoupling of the motions of the fragment atoms within the central unit cell from the motions of the image atoms. A well-known procedure to achieve this is by calculating the supercell hessian and then building the dynamical matrix or the effective mass-weighted unit cell hessian matrix at each reciprocal space point. If $\Phi_{\alpha,\beta}\left(l, 0; l', \kappa\right)$ represents a supercell hessian term corresponding to the double derivative of supercell HMBI energy with respect to the α (x, y or z) coordinate on atom l within the central unit cell (unit cell index = 0) and the β (x, y or z) coordinate on atom l' in the unit cell with index κ,

then the dynamical matrix is given by

$$D_{\alpha,\beta}\left(l,l';\mathbf{k}\right)=\frac{1}{\sqrt{M_l M_{l'}}}\sum_{\kappa'}\Phi_{\alpha,\beta}\left(l,0;l',\kappa'\right)exp\left(-2\pi i\mathbf{k}\cdot\delta\mathbf{x}\left(l,l',\kappa'\right)\right) \tag{5.5}$$

Here, M is the atomic mass, $\mathbf{k}$ is the reciprocal space vector and $\delta x\left(l,l',\kappa'\right)$ is the distance between the atom l in the central unit cell and the atom l' in the unit cell with index κ. In the above equation, the exponential introduces the phase shift in the harmonic motions of the image atoms relative to the atoms in the central unit cell.

Lattice dynamics calculations with HMBI are relatively inexpensive in that the full supercell hessian need not be evaluated. A closer look at the right hand side of the dynamical matrix equation suggests that instead of evaluating the full supercell hessian, the task is reduced to evaluating the force constants involving the nuclear coordinates within the central unit cell. However, because the expensive 1-body and 2-body quantum mechanical hessians involving the central unit cell are already in the fragmented form, once the unit cell HMBI hessian has been evaluated, one can effectively find the HMBI supercell hessian terms involving the central unit cell's nuclear coordinates by performing just one additional supercell hessian calculation with the force field and rearranging the 1-body and 2-body hessians, which is quite inexpensive compared to the quantum mechanical calculations for the pairwise interaction hessians. In short, the unit cell hessian calculation with HMBI plus a trivial force field supercell hessian calculation allows calculation of the harmonic frequencies at all the reciprocal space gridpoints. Moreover, the accuracy of the thermal-vibrational properties evaluated using the harmonic frequencies often depend on the density of the reciprocal space grid which in turn is controlled by the supercell size. Since the only calculation that depends on the supercell size is the supercell hessian calculation with the force field which is relatively inexpensive, the density of the reciprocal space grid can be trivially controlled.

Once the harmonic frequencies in the Brillouin zone have been computed, one can then form the vibrational partition function, Z and compute the thermal properties like the vibrational enthalpy "E_{vib}", Helmholtz free energy of vibration "F", specific heat at constant volume "C_v" and vibrational entropy "S" of the crystal.

$$Z = \prod_{\mathbf{k},j} \frac{e^{-\frac{1}{2}\beta\hbar\omega_j(\mathbf{k})}}{1 - e^{-\beta\hbar\omega_j(\mathbf{k})}} \tag{5.6}$$

$$F = -kTlnZ \tag{5.7}$$

$$E_{vib} = F - T\left(\frac{\delta F}{\delta T}\right)_v \tag{5.8}$$

$$C_v = \left(\frac{\delta E}{\delta T}\right)_v \tag{5.9}$$

$$S = -\left(\frac{\delta F}{\delta T}\right)_v \tag{5.10}$$

The sublimation heat for the crystal can be computed using the formula[161]

$$\Delta H_{sub}\left(T\right) = \Delta E_{lattice} + E_{vib}\left(gas\right) - E_{vib}\left(crystal\right) + 4RT \tag{5.11}$$

5.4 Computational Details

We have performed geometry optimizations and lattice dynamics calculations on the I_h ice and ammonia crystals. For ice, these calculations were performed using the resolution-of-the-identity second order Moller-Plesset perturbation theory (RI-MP2)[42, 43, 44] with the Dunning aug-cc-pVDZ[45, 46] basis set for the quantum mechanical part and polarizable AMOEBA[41] force field for the classical part of the

HMBI calculations. For ammonia crystal, we used the Dunning aug-cc-pVTZ basis set. The electronic structure calculations were performed using a developmental version of Q-Chem software[47] while the classical calculations were done using the Tinker package[48]. The geometry optimizations were done using an open-source optimizer called DL-FIND[49] with the L-BFGS[127] approach. For ice, the spatial truncation cutoffs were set as 6.0 $\mathring{A}$ and 7.0 $\mathring{A}$ as these have been found to work well previously. For the ammonia crystal, these cutoffs for the spatial damping functions were set to 9.0 $\mathring{A}$ and 10.0 $\mathring{A}$. No space group symmetry constraints were imposed during the geometry optimizations or lattice dynamics. However, the lattice system was ensured by constraining the unit cell angles to 90° during the geometry optimizations of both the crystals.

For the lattice dynamics, $3 \times 3 \times 3$ supercells of the optimized unit cell geometries were used. Monkhorst-Pack scheme[163] for reciprocal space gridpoints were used consistent with the supercell size for predicting the thermal properties. However, for phonon dispersion curves, simple interpolating scheme along the various reciprocal space directions was used.

While the experimental sublimation heats for ice were reproduced from Ref. [164, 165], for the ammonia crystal, the experimental sublimation heats at different temperatures were computed using the experimental values of the specific heats at constant pressure (C_p) at different temperatures for the solid, liquid and the gas phases of the crystal, along with the heats of vaporization and fusion.[166, 167]

$$\begin{aligned}\Delta H_{sub}^{expt}(T) = &\int_{T}^{M.P.} C_p\,(crystal)\,\mathrm{d}T + \Delta H_{Fusion} \\ &+ \int_{M.P.}^{B.P.} C_p\,(liquid)\,\mathrm{d}T + \Delta H_{Vaporization} + \int_{B.P.}^{T} C_p\,(gas)\,\mathrm{d}T\end{aligned} \tag{5.12}$$

In this study, the integrals in the above equation were computed as sum of discrete

products of the temperature step and the average value of the specific heat in that temperature range.

5.5 Results and Discussion

5.5.1 Ice I_h

The most common phase of ice found on earth, ice I_h is proton disordered hexagonal phase. Numerous studies[168, 169, 170, 171] have been performed to understand and estimate the crystal properties of this organic crystal. Successful modelling of this system requires a method which can balance the treatment of strong intermolecular interactions and the intramolecular interactions. It is well known that ice displays strong many-body induction effects due to the polar nature of the water molecule and the hydrogen bonding cooperativity. Previously, we have demonstrated that the HMBI scheme with RI-MP2/aug-cc-pVDZ and AMOEBA force field predicts the geometry for this crystal accurately. HMBI predictions reported in our previous study were within a couple kJ/mol in the lattice energy and 1.5% in the lattice parameters. Here, we test this scheme again for predicting the thermal and vibrational properties of I_h ice.

Previously, the inelastic neutron scattering experiments revealed two signature peaks at 229 cm^{-1} and 306 cm^{-1}. At that time, it was hypothesized that there exist two different types of hydrogen bonds differing in the bond strengths which were the source of these 2 peaks. Subsequently however, this model has been discarded due to the successful reproduction of these peaks without the use of this assumption using force field[172], density functional theory[173] and recently with another embedded fragment-based approach[160]. With the HMBI method, we are successfully able to reproduce these two peaks as can be seen in the phonon density of state curve (figure 5.1), although

the 229 cm^{-1} peak is slightly shifted towards the right at 245 cm^{-1} while the 306 cm^{-1} peak is slightly shifted towards the left at 300 cm^{-1}. These small errors arise to a number of reasons suggested in reference [160] such as lack of anharmonic effects, inherent errors with RI-MP2 level of theory and AMOEBA force field, basis set convergence of aug-cc-pVDZ basis with RI-MP2, etc. Furthermore, the average O-H bond length in the optimized structure is 0.9797 Å with a very small standard deviation of 0.0002 Å. Similarly, the average O$\cdots$O distance is 2.7780 Å with a standard deviation of just 0.0111 Å. This clearly conflicts the existence of two different hydrogen bonds of differing strengths.

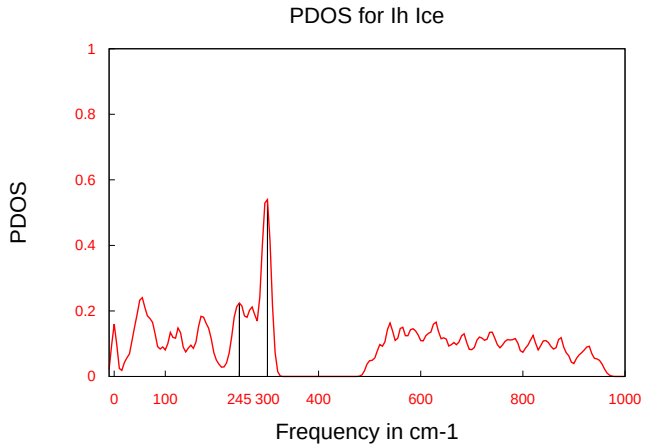


Figure 5.1: Phonon Density of States predicted with HMBI for I_h ice

HMBI predicted sublimation heats of I_h at different temperatures have been tabulated in table 5.1 using the best estimated lattice energy of 60.2 kJ/mol with HMBI and equation 5.11. HMBI estimates reproduce the experimental sublimation heats obtained from reference [164, 165] quite accurately within a kJ/mol, unlike purely classical AMOEBA calculation which shows larger errors of about 5-8 kJ/mol.

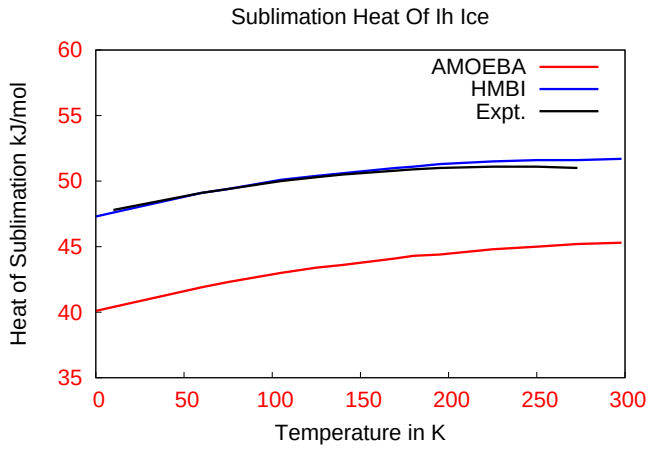


Figure 5.2: Comparison of HMBI and AMOEBA predicted sublimation heats of I_h ice with those predicted with experiments

5.5.2 Phase-1 Ammonia

Previous studies[174, 175, 176, 177, 1], both experimental and computational, on crystalline ammonia indicate the discrepancies in the assignment of the various peaks in the lattice mode region of the vibrational spectrum. Significantly, the phonon dispersion curves and the density of states of this crystal indicate strong LO-TO splitting of some optical modes around the Γ-point, thanks to long-range nature of the electrostatics. Hence, correct comparisons between experiments and theory requires the computational lattice dynamics method to account for this LO-TO splitting. Recently, Reilly et al[1] employed lattice dynamics with molecular dynamics to obtain the full phonon spectrum for this crystal and reassigned the symmetry for each peaks of ammonia. They accounted for the LO-TO splitting by adding a correction term that depends on the Born effective charges and the dielectric tensor. In our method, similar implementation can be achieved by empirically adding this term during the calculation of the dynamical matrix at the Γ-point. However, in this study we do not use such a correction. Never-

Table 5.1: Comparison of HMBI estimated sublimation heats (in kJ/mol) for I_h ice with AMEOBA and experiments at different temperatures. (Errors with respect to experiments shown in parenthesis)

T(K)	ΔH_{sub}^{expt}	ΔH_{sub}^{HMBI}	ΔH_{sub}^{AMOEBA}
0	-	47.3	40.1
10	47.8	47.6 (-0.2)	40.4 (-7.4)
60	49.1	49.1 (0.0)	41.9 (-7.2)
75	49.4	49.4 (0.0)	42.3 (-7.1)
105	50.0	50.1 (0.1)	43.0 (-7.0)
125	50.3	50.4 (0.1)	43.4 (-6.9)
140	50.5	50.6 (0.1)	43.6 (-6.9)
170	50.8	51.0 (0.2)	44.1 (-6.7)
180	50.9	51.1 (0.2)	44.3 (-6.6)
195	51.0	51.3 (0.3)	44.4 (-6.6)
225	51.1	51.5 (0.4)	44.8 (-6.3)
250	51.1	51.6 (0.5)	45.0 (-6.1)
273	51.0	51.6 (0.6)	45.2 (-5.9)

theless, the errors introduced in the sublimation heats and specific heat capacities by not properly treating the LO-TO splitting around the Γ-point are tolerable if a dense reciprocal space grid is used.

The predicted phonon dispersion curves for the lattice mode region are shown in figure 5.4. The phonon frequencies are all real (aside from a few small imaginary frequencies in the acoustic branches near the Γ-point due to numerical noise) suggesting no lattice instabilities. This spectrum more or less reproduces similar trends seen in the dispersion curves by Della Valle et al[174] except for the fact that some degeneracies are lost due to the lack of any symmetry constraints during the geometry optimizations. Hence, symmetry assignments for these peaks becomes difficult. Nevertheless, in table 5.2, we present the Γ-point frequencies in the lattice mode region for comparison against previous computational and experimental studies. Significantly, like the previous study by A. Reilly, we did not encounter any peaks around 260 cm^{-1} in the phonon density

of states (figure 5.3) unlike some experimental studies. This study backs hypothesis of Reilly and co-workers that the weak peaks observed around 260 cm^{-1} might result from lattice defects or to regions of metastable forms.

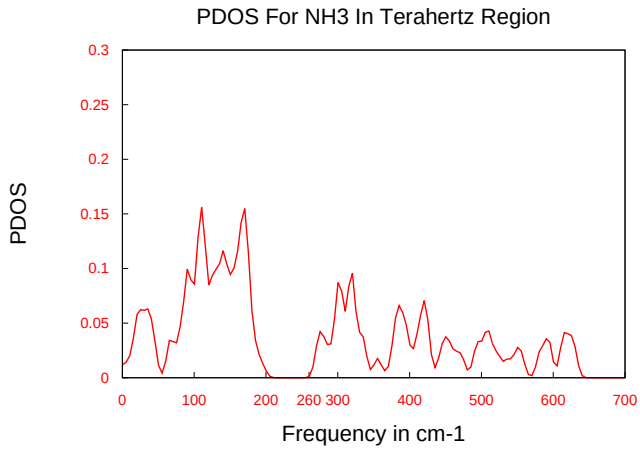


Figure 5.3: Phonon Density of States predicted with HMBI for Phase-1 ammonia

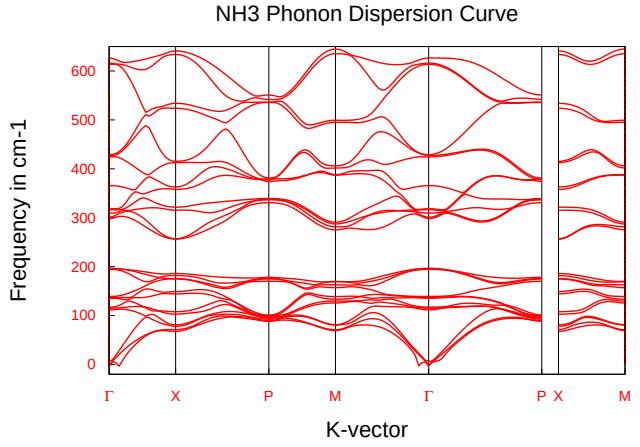


Figure 5.4: Phonon dispersion curves predicted with HMBI for Phase-1 ammonia

In a previous work with HMBI, the best estimated lattice energy for this ammonia crystal was computed to be 40.2 kJ/mol. Using this, the sublimation heats

Table 5.2: Comparison of the predicted Γ-point frequencies (in cm^{-1}) for Phase-1 ammonia lattice modes in this study with the ones predicted in Ref. [1] and experiments.

[a]Binbrek et al., *Chem. Phys. Lett.*, 15:421, 1972

[b]Powell et al., *Can. J. Phys.*, 58:1703, 1980

Phonon branch	This study	Ref. [1]	Experiments
	Freq. (Irred. Rep.)	Freq. (Irred. Rep.)	
1	0	0	0
2	0	0	0
3	0	0	0
4	112	98 (E)	107[a] (A)
5	115	98 (E)	138[a] (E)
6	117	125 (A)	138[a] (E)
7	136	136 (F)	138[a] (F)
8	137	136 (F)	138[a] (F)
9	139	136 (F)	138[a] (F)
10	194	175 (F)	181[b] (F)
11	196	175 (F)	181[b] (F)
12	197	175 (F)	181[b] (F)
13	298	320 (F)	298[a] (A)
-	-	-	260[a]/258[b] (F)
-	-	-	260[a]/258[b] (F)
-	-	-	260[a]/258[b] (F)
14	301	320 (F)	310[a] (E)
15	309	320 (F)	310[a] (E)
16	316	356 (A)	358[a]/361[b] (F)
17	318	367 (E)	358[a]/361[b] (F)
18	366	367 (E)	358[a]/361[b] (F)
19	424	457 (F)	426[a] (F)
20	427	457 (F)	426[a] (F)
21	429	457 (F)	426[a] (F)
22	614	605 (F)	532[b] (F)
23	616	605 (F)	532[b] (F)
24	627	605 (F)	532[b] (F)

Table 5.3: Comparison of HMBI predicted sublimation heats (in kJ/mol) of solid Phase-1 ammonia with estimated experimental data.

T(K)	ΔH_{sub}^{HMBI}	ΔH_{sub}^{expt}	Error in HMBI
0	31.7	29.0	2.7
60	33.4	30.6	2.8
75	33.7	30.9	2.8
105	34.1	31.2	3.0
125	34.3	31.1	3.2
140	34.4	31.1	3.3
170	34.5	31.0	3.5
180	34.5	30.8	3.6
195	34.5	30.7	3.7

were then computed using equation 5.11 and compared against experimental sublimation heats[178, 167, 166] estimated using the procedure mentioned in the computational details in the 0 K to the melting point (195 K) range. 5.3 suggests that the HMBI predicted sublimation heats are within a kcal/mol of the experimental values.

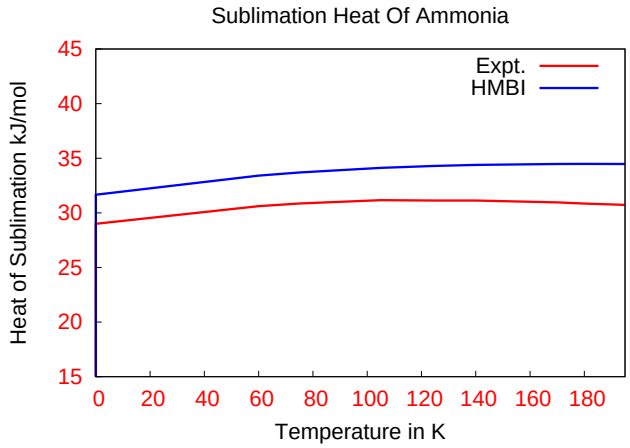


Figure 5.5: Comparison of HMBI predicted sublimation heats of Phase-1 ammonia with experiments

Predicted specific heats at constant volumes with HMBI also compare well

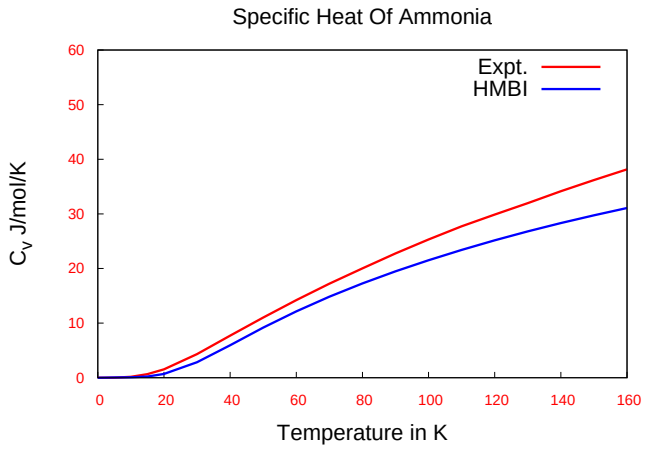


Figure 5.6: Comparison of HMBI predicted specific heats at constant volumes of Phase-1 ammonia with experiments.

with the experimental values[166] as can be seen in 5.6. However, errors in both the sublimation heats and the specific heats at constant volume increase as we near the melting point, which suggests that our approximation to use the geometry optimized at 0 K breaks down at higher temperature due to thermal expansion of the crystal.

5.6 Conclusions

This study demonstrated that accurate crystal properties like vibrational spectra, phonon dispersion curves, sublimation heats and specific heats can be accurately predicted with the fragment-based QM/MM HMBI method. This study also emphasizes the cost-effective advantage in using the lattice dynamics implementation of HMBI owing to its fragment-based formalism. Predictions in the measurable sublimation heats of ice and ammonia crystals indicate that associated errors are within a kcal/mol for this method. Predicted vibrational spectrum for ice reproduces the two signature peaks at 229 cm^{-1} and 306 cm^{-1} quite accurately, thereby consolidating evidence against the

hypothesis supporting the existence of two different types of hydrogen bonds. Predicted vibrational spectrum and phonon dispersion curves for ammonia crystals, albeit without accounting for the LO-TO splitting in optical lattice modes, suggests the non-existence of the contentious peak around 260 cm^{-1} and backs the hypothesis that these observed peaks might be associated with lattice defects.

Chapter 6

Resolving the Discrepancy on the Crystal Structure of Ice XV Between Experiment and Theory: A Fragment-Based QM/MM Study

6.1 Outline

Experimentally the structure of recently identified ordered phase of ice, ice XV has been shown to be antiferroelectric. In clear disagreement with experiments, it was earlier reported with density functional theory calculations that this phase should be ferroelectric. Here, we employ a linear-scaling fragment-based QM/MM method to optimize the geometries of the eighteen possible configurations for ice XV. The energetics

are systematically improved using this method. We report that the antiferroelectric structure is slightly favored by 0.3 kJ/mol in the sublimation enthalpies, in agreement with experiments.

6.2 Introduction

Recently, Salzmann et al[179] have reported the first experimental evidence of the ordered phase of ice VI, named ice XV stable in the 0-130 K range and 0.8-1.5 GPa pressure prepared by cooling ice VI doped with hydrochloric acid. The structure of this phase was shown to be antiferroelectric using powder neutron diffraction with the two interpenetrating C1 and C2 hydrogen networks in the unitcell and the overall space group of P-1. It has also been suggested that the preference for the highly polar C1/C2 networks allows maximum directional cooperativity along the individual hydrogen bonded networks, while the P-1 space group allows for the most effective cancellation of these polarities along the individual networks. However, previous calculations with DFT had predicted a ferroelectric structure with A1/A2 networks and space group Cc as the most stable[180, 181]. While the reasons for this apparent disagreement between DFT results and experiments are still unclear, it has been reported that energy difference between the lowest energy antiferroelectric and lowest energy ferroelectric structures is very small. Elsewhere[182], it has been suggested that ice XV requires highly accurate treatment with expensive methods such as coupled cluster CCSD(T) method or Quantum Monte Carlo simulations. Alternatively, hybrid DFT approach with dispersion correction tailored for this system has been suggested.

In this study, we present another alternative to these highly expensive computational approaches. We employ a fragment-based QM/MM method called Hybrid

Many-Body Interaction (HMBI) method[9, 10, 11, 12, 13, 14, 15, 16]. Fragment-based methods[3, 4, 5, 6, 7, 8, 9] like HMBI target electronic structure quality predictions while remaining computationally affordable for such periodic organic crystals. This method has previously been employed to predict accurate lattice energies and crystal structures[16] of a variety of organic crystals like I_h ice, benzene, formamide, etc. and also to predict and explain the energetics of gas phase electrosprayed clusters of uracils and calcium cation[13].

6.3 HMBI Background

The HMBI method decomposes the system into molecular fragments and builds the electronic energy by appropiate treatment of the various inter-fragment interactions using highly accurate electronic structure methods and polarizable force fields. More specifically, the fragment and short-range pairwise interactions are computed using electronic structure methods while the many-body interactions are computed classically, allowing for the affordable treatment of large periodic systems like organic crystals.

$$E^{HMBI} = E^{QM}_{1-Body} + E^{QM}_{short-range\ 2-Body} + E^{MM}_{long-range\ 2-Body} + E^{MM}_{Many-Body} \quad (6.1)$$

The use of polarizable force fields also allows us to capture the long-range electrostatics accurately. One advantage of using this fragment-based approach over DFT is that we can systematically improve the HMBI predictions by use of highly accurate electronic structure methods and better force fields.

First, we outline the periodic implementation of the HMBI method. The HMBI

energy is given by

$$E^{HMBI} = E^{MM}_{PBC} + \sum_{i} \left(E^{QM}_{i} - E^{MM}_{i} \right) + \sum_{ij(0)} d_{ij(0)} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(0)} \right) + \frac{1}{2} \sum_{i} \sum_{j(n_v)}^{images} d_{ij(n_v)} \left(\Delta^2 E^{QM}_{ij(0)} - \Delta^2 E^{MM}_{ij(n_v)} \right) \tag{6.2}$$

where the energy of the full crystal is approximated by the Ewald sum computed classically with polarizable force field plus the quantum mechanical corrections for the fragment energies for all the central unitcell's fragments and their short-range pairwise interaction energies with other fragments within or outside the central unitcell. The damping function, d_{ij} ensures that the potential energy surface are continuous and differentiable and allow for gradual shift from the short-range quantum mechanical treatment to long-range classical mechanical treatment for inter-fragments interactions.

Previously, we have reported the implementation of the nuclear and lattice gradients of the HMBI energy. In this study, we optimize the crystal structures under external isotropic pressure. As a result, our lattice gradients include an extra term to accomodate this external pressure. Accordingly, the lattice gradient terms $\frac{\delta E'}{\delta v_{1x}}$, $\frac{\delta E'}{\delta v_{2y}}$ and $\frac{\delta E'}{\delta v_{3z}}$ in Ref. [16] are modified to $\frac{\delta E'}{\delta v_{1x}} + p$, $\frac{\delta E'}{\delta v_{2y}} + p$ and $\frac{\delta E'}{\delta v_{3z}} + p$ respectively, where p is the external pressure.

6.4 Geometry Optimization With HMBI

In this study, HMBI was used to first optimize the geometries of the eighteen possible non-P1 structures of ice XV and then to systematically improve the energetics. Experimental evidence suggests that ice XV has a psuedo-orthorhombic unitcell with the lattice angles very close to 90°. We thank Professor Christoph Salzmann for providing us with the initial guess structures. The 18 structures were then optimized with the parallel HMBI scheme employing counterpoise-corrected resolution-of-the-identity

second-order Moller-Plesset level of theory (RI-MP2)[42, 43, 44] and Dunning aug-cc-pVDZ basis set[45] and its auxiliary basis set[46] for the quantum mechanical treatment and the AMOEBA[41] polarizable force field for the classical treatment. The quantum calculations were done using a developmental version of Q-Chem package[47] while the classical force field calculations were done using TINKER package[48]. The DL-FIND optimizer[49] was used to perform these geometry optimizations using the L-BFGS approach[127]. The spatial truncation cutoffs for the damping function were set as 6.0 Å and 7.0 Å as previously[11, 16], we had determined that this rather shorter cutoffs work quite well for ice. Finally, we do not impose any symmetry constraint during the geometry optimization. This is justified since previous performance of this method on I_h ice suggested that the optimized structures with and without these symmetry constraints show very small differences.

Although direct comparisons of the predicted lattice parameters with experiments cannot be made since the experimental data is gathered at finite temperatures (80 K here), fortunately the relatively small changes in the lattice parameters for other ice phases like I_h in the 0-80 K range allow us to approximate that the temperature dependence of the lattice parameters of ice XV could be small. Nevertheless, the thermal effects have been somewhat accounted for in the thermal properties of these ice XV structures using the lattice dynamics implementation, which we discuss later. The basis set convergence error with MP2 with relatively small aug-cc-pVDZ basis is also quite significant but the use of counterpoise correction[183] in our calculations somewhat reduces this error. Gradient implementation with explicit correlation methods[23, 24, 25, 26] or spin-component scaled MP2[27, 28, 29, 30, 31] could be beneficial in the future to address this issue. In any case, experiments at lower temperatures than 80 K would help for a better comparison with our computational results.

Our results for the optimized set of lattice parameters for the 18 possible structures of ice XV are tabulated in Table 1. Clearly, most of the lattice parameters have optimized to similar values with standard deviation of less than 0.100 Å in the unitcell lengths and 1.000° in the unitcell angles. The percent errors in the mean values of all the lattice parameters are less than 1% from the experimental values measured at 80 K. The percent errors for individual optimzied lattice parameters for all the structures are within 3% of both the experimental as well as the mean values. Compared to other lattice parameters, the errors in the unitcell length "c" are larger against both the experimental values and the mean values. As expected, all the optimized volumes have been underestimated compared to the volume at 80 K with experiments. The standard deviation in the unitcell volumes is less than 1.000 $Å^3$. Interestingly, the "2C" structure made up of C1/C2 networks with P-1 space group, which corresponds to the antiferroelectric structure that was found to be the most stable with experiments has the highest optimized volume compared to the other 17 structures. Meanwhile, the ferroelectric structure "9A1" made up of one A1 and one A2 network with Cc space group which was found to be the most stable with DFT calculations had the third smallest optimized volume.

6.5 Systematic Improvements With HMBI

The optimized lattice energies with RI-MP2/aug-cc-pVDZ:AMOEBA HMBI scheme of these 18 structures have been tabulated in Table 2. It has been suggested that the higher directional-polarity along the C1/C2 hydrogen-bonded networks in ice XV makes the anti-ferroelectric structure the most stable. Clearly, accurate determination

Table 6.1: Optimized lattice parameters of all 18 possible structures of ice XV

	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	V (Å^3)
2a	6.260	6.266	5.687	91.67	89.64	90.58	223.9
2b	6.206	6.227	5.769	91.38	90.52	90.80	222.8
2c	6.199	6.186	5.861	89.50	89.24	90.58	224.7
4a1	6.245	6.305	5.648	89.85	90.10	90.26	222.4
4a2	6.282	6.278	5.639	89.97	87.71	89.50	222.2
4b1	6.167	6.302	5.743	89.90	90.84	90.49	223.1
4b2	6.202	6.238	5.771	89.94	90.92	89.80	223.2
4c1	6.168	6.223	5.841	90.06	90.28	89.94	224.2
4c2	6.191	6.203	5.828	89.97	89.52	89.61	223.8
7a	6.261	6.274	5.687	89.90	89.95	89.94	223.4
7b	6.283	6.184	5.750	90.11	90.68	89.96	223.4
7c	6.230	6.171	5.834	90.32	90.22	90.06	224.3
9a1	6.276	6.289	5.635	90.87	90.81	90.62	222.3
9a2	6.268	6.260	5.657	89.57	89.12	89.95	221.9
9b1	6.216	6.223	5.761	90.44	90.70	89.83	222.8
9b2	6.235	6.237	5.753	90.15	90.55	88.56	223.6
9c1	6.218	6.216	5.786	89.82	90.30	89.51	223.6
9c2	6.217	6.219	5.785	89.60	89.90	89.57	223.7
Mean	6.229	6.245	5.744	90.17	90.06	89.98	223.3
Std. Dev.	0.037	0.036	0.075	0.59	0.80	0.54	0.8
Expt.	6.232	6.244	5.790	90.06	89.99	89.92	225.3
Error in Mean	-0.052	0.016	-0.797	0.12	0.07	0.06	-0.9

Table 6.2: Systematic Improvement in the Lattice Energies (in kJ/mol) of All 18 Possible Ice XV Structures

	RIMP2/ ADZ[a]: AMOEBA	RIMP2/ ADZ: AIFF/ Sadlej	MP2/ ATZ[a]: AIFF/ Sadlej	MP2/ AQZ[a]: AIFF/ Sadlej	MP2/ CBS[a]: AIFF/ Sadlej	MP2/ CBS: AIFF/ ATZ	$\Delta^{CCSD(T)}$	Best Esti- -mate
2a	27.8	27.0	38.8	40.7	42.2	42.2	-1.4	40.8
2b	28.1	26.9	38.8	40.7	42.2	42.1	-1.4	40.7
2c	28.2	27.8	39.6	41.4	42.9	42.9	-1.4	41.5
4a1	28.4	24.2	36.1	37.9	39.4			
4a2	28.5	23.8	35.7	37.5	39.0			
4b1	28.1	24.7	36.6	38.4	39.9			
4b2	28.0	23.5	35.5	37.3	38.8			
4c1	28.4	25.3	37.1	38.9	40.4			
4c2	28.4	24.8	36.6	38.4	39.9			
7a	28.0	20.1	32.1	33.8	35.3			
7b	28.3	21.4	33.3	35.0	36.5			
7c	28.4	21.9	33.7	35.5	37.0			
9a1	28.3	27.5	39.4	41.2	42.8	42.8	-1.5	41.3
9a2	28.6	20.3	32.3	34.1	35.6			
9b1	28.2	20.8	32.8	34.6	36.1			
9b2	28.0	27.1	39.0	40.8	42.3	42.2	-1.3	40.9
9c1	27.9	16.9	28.9	30.6	32.1			
9c2	28.2	10.9	22.9	24.6	26.0			

[a] ADZ = aug-cc-pVDZ, ATZ = aug-cc-pVTZ, AQZ = aug-cc-pVQZ, CBS = complete basis set

of the many-body induction is required. The off-the-shelf AMOEBA polarizable force field in general does a good job but is not accurate enough, mainly because AMOEBA totally misses the many-body dispersion interactions and only includes the many-body induction interactions of the induced dipoles. Clearly, with this scheme, neither the "2c" structure or the "9a1" structure shows the highest lattice energy. In fact, the "9a2" and "4a2" structures are the most bound structures with this scheme. Meanwhile, the use of our ab-initio force field (AIFF) which has been shown to give more accurate many-body effects correctly predicts "2c" as the most stable structure in terms of lattice energy. AIFF features force field parameters in terms of atom-centered distributed multipoles and polarizabilities which are calculated on-the-fly separately for each fragment. AIFF includes the 3-Body dispersion terms and polarization effects beyond the induced dipole interactions. With this force field, the "9a1" structure is less stable by only 0.3 kJ/mol compared to "2c", which prompted us to systematically refine our lattice energy predictions by going to the CBS limit with MP2 for all the structures and also performing more accurate but expensive CCSD(T) single point calculations for five highest bound structures. As can be seen from Table 2, when we extrapolate the MP2 results with triple zeta and quadruple zeta basis to the complete basis set limit, "2c" is still the most stable structure by about 0.2 kJ/mol followed by the "9a1". The many-body effects with aug-cc-pVTZ basis were done on the five highest bound structures, "2a", "2b", "2c", "9a1" and "9b2", to further refine their lattice energies. The change in many-body effects with the bigger basis however were not appreciable, meaning that these were sufficiently converged with the Sadlej basis[184, 185]. CCSD(T) corrections at aug-cc-pVTZ basis were then applied to these five highest bound structures. The final lattice energies suggest that the antiferroelectric "2c" structure is the most stable structure in agreement with experiments, while the the ferroelectric "9a1" structure is

only 0.2 kJ/mol less bound.

Table 3 shows the breakdown of the best estimated lattice energies for these five highest bound crystals into various contributions coming from the quantum part, many-body electrostatics, many-body induction and many-body dispersion. With just the 1-body and short-range 2-body quantum interactions, the antiferroelectric "2c" structure would be the least favored of these five structures while the ferroelectric "9a1" structure would be most bound. We further decomposed these quantum interactions into 1-body and short-range 2-body interactions. This decomposition suggests that even though the monomers in the "2c" structure are the most relaxed, as indicated by the least negative lattice energy, the short-range dimer interactions in the "9a1" structure are the strongest, which offset the stability order considering only the 1-body interactions. The many-body electrostatics and many-body dispersion contributions are similar for all of these five crystals. Clearly, the many-body effects are dominated by the induction effects in these ice structures. However, the many-body induction effects are significantly more stabilizing in the "2c" structure compared to other structures, which more than compensate the difference of 1.7 kJ/mol coming from the 1-body and short-range 2-body interactions between "2c" and "9a1". This confirms that the hydrogen bond cooperativity along the individual hydrogen bonded network is greatest in the C1/C2 networks of the "2c" structure. In short, if the long-range interactions are neglected, the local 2-body interactions would make the "9a1" structure most bound, but the accurate consideration of the long-range interactions which are the greatest in the "2c" structure, ensures that "2c" is the most bound consistent with experiments. This decomposition thereby points towards the fact that accurate treatment of long-range interactions in this system is vital for correct structure prediction.

Table 6.3: Lattice Energy Contributions From Different Types of Interactions for the 5 Highest Bound Ice XV Structures

	1-Body QM	2-Body QM	(1+2)-Body QM	MB-ES[a]	MB-Disp.[a]	MB-Ind.[a]	Total
2a	-1.9	48.1	46.2	0.3	0.3	-8.5	40.8
2b	-2.0	48.0	46.0	0.3	0.3	-8.7	40.7
2c	-1.7	47.1	45.4	0.3	0.3	-10.2	41.5
9a1	-1.9	49.1	47.1	0.8	0.3	-8.6	41.3
9b2	-1.8	47.8	46.0	0.1	0.2	-8.7	40.9

[a] MB = Many-Body, ES = Electrostatic, Disp. = Dispersion, Ind. = Induction

6.6 Lattice Dynamics With HMBI

Next, for comparisons against future experiments, we predicted the sublimation heats for the "2c" and "9a1" structures in the 0-130 K range using the lattice dynamics implementation of HMBI. Here, we emphasize the advantages of performing the lattice dynamics[157] calculations with HMBI. Firstly, the nuclear gradients and hessians of the HMBI electronic energy are straighforward to find. Secondly, since all the quantum 1-Body and 2-Body hessian calculations are already in the fragmented form, an inexpensive supercell hessian calculation with the force field gives us all the force constants needed to build the dynamical matrices at different reciprocal space points. Finally, often the accuracy of the thermal-vibrational properties depend on the size of the reciprocal space grid which in turn depends on the size of the supercell. Since the supercell hessian calculation in HMBI is done with the force field, the size of the supercells can be controlled trivially to improve the predictions.

For our calculations, supercells of size $3 \times 3 \times 3$ were used. The Monkhorst-Pack scheme[163] for the reciprocal space gridpoints was used. The dynamical matrix element corresponding to the α-component of the displacement of atom κ in the central unitcell (index = 0) and the β-component of the displacement of atom κ' in a unitcell

Table 6.4: Predicted Sublimation Enthalpies in kJ/mol for "2c" and "9a1" structures

T (in K)	2c	9a1
0	29.2	28.8
5	29.3	29.0
10	29.5	29.1
25	30.0	29.6
40	30.4	30.1
60	31.0	30.7
75	31.3	31.0
105	32.0	31.7
125	32.3	32.0

with index l' at a reciprocal space wavevector, $\mathbf{k}$, is given by

$$D_{\alpha,\beta}\left(\kappa,\kappa';\mathbf{k}\right)=\frac{1}{\sqrt{M_{\kappa}M_{\kappa'}}}\sum_{l'}\Phi_{\alpha,\beta}\left(\kappa,0;\kappa',l'\right)exp\left(-2\pi i\mathbf{k}\cdot\left(\mathbf{R}\left(\kappa,\mathbf{0}\right)-\mathbf{R}\left(\kappa',\mathbf{l}'\right)\right)\right) \tag{6.3}$$

where Φ is the supercell hessian matrix, M and R are the atomic mass and position vector respectively. The harmonic vibrational frequencies are then calculated at each point in the reciprocal space grid by diagonalizing these dynamical matrices. Except for some small imaginary frequencies around the Γ-point, our calculations did not predict any imaginary frequencies suggesting instability of the lattice at any other reciprocal space point. Table 4 lists the sublimation enthalpies, ΔH_{sub} predicted using the formula[161],

$$\Delta H_{sub}=\Delta E_{lattice}+E_{vib}\left(gas\right)-E_{vib}\left(crystal\right)+4RT \tag{6.4}$$

where $\Delta E_{lattice}$ are the lattice energies, $E_{vib}\left(gas\right)$ and $E_{vib}\left(crystal\right)$ are the vibrational enthalpies of the monomer in the gas phase and in the crystal respectively. Clearly, the sublimation heats in the $0-130$ K range for the antiferroelectric structure, "2c" are about $0.2-0.3$ kJ/mol higher than the ferroelectric structure, "9a1".

6.7 Conclusions

In conclusion, this is the first computational study which correctly predicts that ice XV has an antiferroelectric P-1 structure in agreement with experiments, which the previous DFT studies failed to do. Here, we also emphasize the advantage of using fragment-based methods which allow systematic refinements to get highly accurate electronic structure quality predictions while studying large chemical systems like organic crystals. The role played by the stronger hydrogen-bond cooperativity along the individual hydrogen-bonded networks in making the antiferroelectric structure most bound was confirmed. Predicted sublimation enthalpies in the $0 - 130$ K range suggest that the antiferroelectric P-1 structure is more stable by just $0.2 - 0.3$ kJ/mol than the ferroelectric Cc structure.

Chapter 7

Conclusions

Predicting structure and properties of molecular crystals and clusters is important in many applications of these materials like pharmaceutical drugs, organic semiconductors, explosives, etc. Polymorphism affects these materials due to inherent weak intermolecular interactions allowing multiple near-isoenergetic packing motifs. A priori detection and screening of the unwanted but competing polymorphs requires reliable theoretical methods that can accurately distinguish between polymorphs even when the energy difference is within a kcal/mol. Conventional methods like classical force field methods are not quite accurate while methods depending on Density Functional Theory, though quite accurate after empirical or semi-empirical dispersion corrections, are not systematically reliable. Alternatively, fragment-based methods which decompose the system into more important short-range interactions and less important long-range interactions are well-suited to get highly accurate results for these molecular aggregates. Moreover, these methods are trivially parallelizable which makes them highly efficient for dealing with large systems.

The Hybrid Many-Body Interaction (HMBI) method is a fragment-based QM/MM

method which has been shown here to predict lattice energies within a kcal/mol accuracy after employing systematic improvements of the predicted results by use of better levels of theory and basis sets. This method employs the electronic structure methods for short-range interactions of the fragments and classical polarizable force fields help capture the long-range many-body effects. This method has been shown to accurately balance the treatment of QM and MM for a wide variety of benchmark moelcular crystals. It has also been shown to correctly predict the energetics of gas phase charged clusters of uracils and calcium cation. Crystal structure predictions with this method yielded unitcells which were within a couple percent of the experimental lattice parameters' values with the correct space group symmetry, thereby suggesting high reliability.

Accurate prediction of crystal properties like vibrational spectra, sublimation heats, etc. can help distinguish polymorphs and understand the polymorphic stability order. The HMBI method with the help of lattice dynamics has been shown here to predict crystal properties very accurately. Crystal properties like sublimation heats at finite temperatures were found to be within a kcal/mol of the experimental values. Moreover, the vibrational spectra of benchmark crytals were shown to be reproduced quite accurately under harmonic approximation with HMBI. Moreover, crystal structure and property predictions for materials under external stress can be reliably made with HMBI as was demonstrated by its successful treatment in distinguishing between competing polymorphs of ordered ice XV phase.

Here, we also emphasize that for the forseeable future, efficient fragment-based methods like HMBI would play a role for molecular crystals similar to what highly accurate electronic structure methods like MP2, CI and coupled cluster play for treatment of gas phase systems, while being computationally affordable.

Appendix A

Nuclear Gradient and Hessian of the Spatial Damping Function

The interpolating function d_{ij} ensures a smooth, differentiable transition between the short-range QM and long-range MM pairwise interactions by smoothly transitioning between the two regimes over the interval defined by the cutoffs c_1 and c_0:

$$d_{ij}(R) = \begin{cases} 1 & \text{if } R \leq r_1 \\ \frac{1}{1+e^{2|c_1-c_0|/(c_1-R)-|c_1-c_0|/(R-c_0)}} & \text{if } c_1 < R < c_0 \\ 0 & \text{if } R \geq c_0 \end{cases} \tag{A.1}$$

where R is defined as the shortest interatomic distance between the two molecules in the dimer. Since R depends on the nuclear coordinates of the two atoms defining the shortest distance between the two interacting monomers, the spatial damping function, d_{ij} also depends on these nuclear coordinates. Hence, we also need to include the nuclear derivatives of these damping functions in the HMBI nuclear gradient and hessian formalisms. If q_l and $q_{l'}$ are any two nuclear coordinates within the central unitcell, the

gradient of the damping function is given by

$$\begin{aligned}\frac{\partial d_{i(0)j(n)}}{\partial q_l} &= \frac{\partial d_{i(0)j(n)}}{\partial q_{l(0)}} = \frac{\partial d_{i(0)j(n)}}{\partial q_{l(n)}} \\ &= -\left(\delta_{lm} - \delta_{lm'}\right) \frac{e^{g(c_1,c_0,R_{ij})} \left(q_m^i - q_{m'}^j\right) \left(\frac{1}{(c_0-R_{ij})^2} + \frac{2}{(c_1-R_{ij})^2}\right) |c_0 - c_1|}{\left(1 + e^{g(c_1,c_0,R_{ij})}\right)^2 R_{ij}} \\ &= -\Omega_{ij}^a \Omega_{ij}^b \Omega_{ij}^c \Omega_{ij}^d\end{aligned} \tag{A.2}$$

where $g\left(c_1, c_0, R_{ij}\right) = 2|c_1 - c_0|/\left(c_1 - R_{ij}\right) - |c_1 - c_0|/\left(R_{ij} - c_0\right)$, q_l^i is the q ($= x$, y or z) coordinate of the l^{th}-atom in the i^{th} or the j^{th} monomer, R_{ij} is the shortest distance between the monomers i and j, m and m' are the atoms in monomers i and j respectively which define R_{ij}, and δ is the Kronecker delta.

The hessian of the damping function is then given by the chain rule

$$\begin{aligned}\frac{\partial d_{ij}^2}{\partial q_l \partial q_{l'}} =& - \\ & \left(\frac{\partial \Omega_{ij}^a}{\partial q_{l'}^{i'}} \Omega_{ij}^b \Omega_{ij}^c \Omega_{ij}^d + \Omega_{ij}^a \frac{\partial \Omega_{ij}^b}{\partial q_{l'}^{i'}} \Omega_{ij}^c \Omega_{ij}^d + \Omega_{ij}^a \Omega_{ij}^b \frac{\partial \Omega_{ij}^c}{\partial q_{l'}^{i'}} \Omega_{ij}^d + \Omega_{ij}^a \Omega_{ij}^b \Omega_{ij}^c \frac{\partial \Omega_{ij}^d}{\partial q_{l'}^{i'}}\right)\end{aligned} \tag{A.3}$$

where the functions Ω's are given by

$$\Omega_{ij}^a = e^{g(c_1,c_0,R_{ij})} \tag{A.4}$$

$$\Omega_{ij}^b = \left(\delta_{lm} - \delta_{lm'}\right)\left(q_m - q_{m'}\right) \tag{A.5}$$

$$\Omega_{ij}^c = \left(\frac{1}{\left(c_0 - R_{ij}\right)^2} + \frac{2}{\left(c_1 - R_{ij}\right)^2}\right) |c_0 - c_1| \tag{A.6}$$

$$\Omega_{ij}^d = \frac{1}{\left(1 + e^{g(c_1,c_0,R_{ij})}\right)^2 R_{ij}} \tag{A.7}$$

Note that Ω^b is a function of the type of component (x, y or z) of the nuclear coordinate, q_l. The derivatives of these Ω functions are given by

$$\frac{\partial \Omega_{ij}^{a}}{\partial q_{l'}} = -\frac{\Omega_{ij}^{a}\left(q_m - q_{m'}\right)\Omega_{ij}^{c}}{R_{ij}} \tag{A.8}$$

$$\frac{\partial \Omega_{ij}^{b}}{\partial q_{l'}} = \left(\delta_{lm} - \delta_{lm'}\right)\left(\delta_{l'm} - \delta_{l'm'}\right) \tag{A.9}$$

$$\frac{\partial \Omega_{ij}^{c}}{\partial q_{l'}} = -\frac{2\left(q_m - q_{m'}\right)\left(\frac{1}{\left(c_0 - R_{ij}\right)^3} + \frac{2}{\left(c_1 - R_{ij}\right)^3}\right)\left|c_0 - c_1\right|}{R_{ij}} \tag{A.10}$$

$$\frac{\partial \Omega_{ij}^{d}}{\partial q_{l'}} = \frac{\left(q_m - q_{m'}\right)\left(1 + \Omega_{ij}^{a} - 2\Omega_{ij}^{a}\left(-\Omega_{ij}^{c} R_{ij}\right)\right)}{\left(\left(1 + \Omega_{ij}^{a}\right) R_{ij}\right)^3} \tag{A.11}$$

Appendix B

HMBI Lattice Hessian

Lattice hessians could prove to be useful for stability analysis of the crystal structure as well for evaluating the elastic constants as will be shown in Appendix 3.

We start with the lattice gradient of the HMBI energy formalism:

$$E^{HMBI} = E^{MM}_{PBC} + E' \tag{B.1}$$

$$\begin{aligned}\frac{\partial E'}{\partial u} &= \frac{\partial E'}{\partial a_x}\frac{\partial a_x}{\partial u} + \frac{\partial E'}{\partial b_x}\frac{\partial b_x}{\partial u} + \frac{\partial E'}{\partial b_y}\frac{\partial b_y}{\partial u} + \frac{\partial E'}{\partial c_x}\frac{\partial c_x}{\partial u} + \frac{\partial E'}{\partial c_y}\frac{\partial c_y}{\partial u} + \frac{\partial E'}{\partial c_z}\frac{\partial c_z}{\partial u} \\ &= \sum_{v_q} \frac{\partial E'}{\partial v_q}\frac{\partial v_q}{\partial u}\end{aligned} \tag{B.2}$$

Note that the lattice gradient of the E^{MM}_{PBC} term is found using finite difference in our HMBI code.

For the HMBI hessian, we need terms like $\frac{\partial^2 E'}{\partial u \partial u'}$ and $\frac{\partial^2 E'}{\partial u \partial q_k}$ where u' is a lattice parameter and q_k is a nuclear coordinate. We differentiate equation B.2 throughout with respect to another lattice parameter/nuclear coordinate to get,

$$\frac{\partial^2 E'}{\partial u \partial u'} = \sum_{v_q} \left(\frac{\partial}{\partial v_q}\left(\frac{\partial E'}{\partial u'}\right)\frac{\partial v_q}{\partial u} + \frac{\partial E'}{\partial v_q}\frac{\partial^2 v_q}{\partial u \partial u'} \right) \tag{B.3}$$

The second term in the above equation can be found trivially since the lattice gradient code has already been implemented. Expanding the first term using equation B.2, we get

$$\begin{aligned}\sum_{v_q}\frac{\partial}{\partial v_q}\left(\frac{\partial E'}{\partial u'}\right)\frac{\partial v_q}{\partial u} &= \sum_{v_q}\frac{\partial}{\partial v_q}\left(\sum_{w_r}\frac{\partial E'}{\partial w_r}\frac{\partial w_r}{\partial u'}\right)\frac{\partial v_q}{\partial u} \\ &= \sum_{v_q}\frac{\partial^2 E'}{\partial v_q^2}\left(\frac{\partial v_q}{\partial u}\right)^2 + \sum_{v_q\neq w_r}\frac{\partial^2 E'}{\partial v_q \partial w_r}\frac{\partial v_q}{\partial u}\frac{\partial w_r}{\partial u'}\end{aligned} \tag{B.4}$$

The geometric terms $\frac{\partial v_q}{\partial u}$ are based purely on lattice parameters, so are constant for a set of lattice parameters. On the other hand, the terms $\frac{\partial^2 E'}{\partial w_r \partial v_q}$ need careful evaluation. For formulating these terms, we will first develop the the double differential operators that will operate on the term E' in the HMBI energy formalism.

$$\begin{aligned}\hat{P_{v_q}} &= \frac{\partial}{\partial v_q} \\ &= \frac{1}{2}\sum_{i(0)}\sum_{j(n_v)} n_v \sum_k \frac{\partial}{\partial q_k}\end{aligned} \tag{B.5}$$

$$\begin{aligned}\hat{P_{w_r}}\hat{P_{v_q}}E' &= \frac{1}{2}\sum_{i(0)}\sum_{j(n_v)} n_v \sum_k \frac{\partial^2\left(f_{i(0)j(n_v)}\Delta^2 E_{i(0)j(n_v)}\right)}{\partial q_k \partial w_r} \\ &= \frac{1}{2}\sum_{i(0)}\sum_{j(n_v)} n_v \sum_k \frac{1}{2}\sum_{a(0)}\sum_{b(n_w)} n_w \sum_l \frac{\partial^2\left(f_{i(0)j(n_v)}\Delta^2 E_{i(0)j(n_v)}\right)}{\partial q_k \partial r_l}\end{aligned} \tag{B.6}$$

Now, $\Delta^2 E_{i(0)j(n_v)} = E_{i(0)j(n_v)} - E_{i(0)} - E_{j(n_v)}$ and

$$\frac{\partial^2\left(E_{i(0)j(n_v)}\right)}{\partial q_k \partial r_l} = 0 \quad if \quad b \neq \quad j \quad and \quad n_v = n_w \tag{B.7}$$

$$\frac{\partial^2 E_{j(n_v)}}{\partial q_k \partial r_l} = 0 \quad if \quad b \neq j \quad and \quad n_v = n_w \tag{B.8}$$

Then we have,

$$\hat{P_{w_r}}\hat{P_{v_q}}E' = \frac{1}{2}\sum_{i(0)}\sum_{j(n_v)} n_v \sum_k \frac{1}{2}\sum_{i(0)}\sum_{j(n_v)} n_v \sum_l \frac{\partial^2\left(f_{i(0)j(n_v)}\Delta^2 E_{i(0)j(n_v)}\right)}{\partial q_k \partial r_l} \tag{B.9}$$

Note that $b = j$ (or $b = i$) means that the monomer b in the n_w^{th} cell along the unit cell axis $\tilde{\mathbf{w}}$ has the same reference monomer as that of the monomer j in the n_v^{th} cell along the unit cell axis $\tilde{\mathbf{v}}$. Secondly, $l = k$ means that we are talking about the same atom (the atom which is important as far as spatial damping function goes) in these image monomers.

Terms based on the spatial damping function can be computed using its analytical expression while the other terms can be extracted from the QM or MM output files of the HMBI job.

Appendix C

Formalisms of HMBI Stress Tensor, Stiffness Tensor and Elastic Constants

C.1 Few Related Comments

C.1.1 Tensor Equation

Tensor equations of the form $\mathbf{B} = \mathbf{CA}$ relate the tensors $\mathbf{B}$ and $\mathbf{A}$ which are called field tensors as they represent an action on the system and a reaction of the system to the action (Mathematically, one cannot say which is an action and which is a reaction since one can equivalently write the above equation as $\mathbf{A} = \mathbf{C^{-1}B}$). The tensor (of appropriate rank) is a property of the system and hence is called Matter tensor. As an example, if $\mathbf{B}$ is the strain tensor and $\mathbf{A}$ is the stress tensor, then $\mathbf{C}$ is the tensor of elastic constants.

C.1.2 How do We Start the Stress Tensor Derivation?

Each point in the system can be assumed to be an infinitesimal cube upon which the stress acts. This cube is called the volume element.

C.1.3 Homogeneous Stress

Homogeneous stress means that the forces are independent of the location of the volume element in the system.

C.1.4 Features of Stress Tensor

Nye[186] proves that the stress tensor at an equilibrium or non-equilibrium point is a symmetric tensor for both the homogeneous and inhomogeneous cases provided that there are no body torques. Hydrostatic stress means that the system is in a liquid. A liquid cannot apply shear stress and hence the hydrostatic stress tensor is diagonal.

C.1.5 Strain Tensor and Deformation

Nye distinguishes the deformation tensor e_{ij} and the strain tensor ϵ_{ij}. The deformation tensor is the sum of a symmetric strain tensor and an asymmetric rotation tensor ω_{ij} i.e.

$$\begin{aligned} e_{ij} &= \frac{(e_{ij} + e_{ji})}{2} + \frac{(e_{ij} - e_{ji})}{2} \\ &= \epsilon_{ij} + \omega_{ij} \end{aligned} \tag{C.1}$$

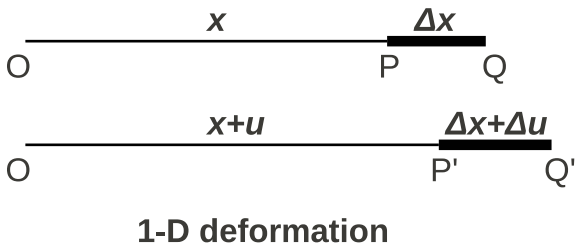


Figure C.1: 1-D Stretch

C.2 Strain Tensor: Why is It Symmetric?

C.2.1 1-D Strain

First we need to understand what each term in the matrix e_{ij} means. Let us start with strain in 1-D Consider two points P and Q originally separated by a distance Δx. When the strain is applied, let the relative displacement in 1-D be Δu. So, by definition,

$$\begin{aligned} e &= strain \\ &= \frac{relative \quad displacement}{original \quad length} \\ &= \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \\ &= \frac{\mathrm{d}u}{\mathrm{d}x} \end{aligned} \tag{C.2}$$

The homogenous strain is a constant and on integration, we end up with

$$u = u_0 + ex \tag{C.3}$$

C.2.2 2-D Strain

In 2-D, strain causes a displacement of two proximate points P and Q joined by the vector $\vec{\Delta x}\left(x_1, x_2\right)$ to $\vec{\Delta x}\left(x_1, x_2\right) + \vec{\Delta u}\left(u_1, u_2\right)$ Now, since u_1 and u_2 are functions

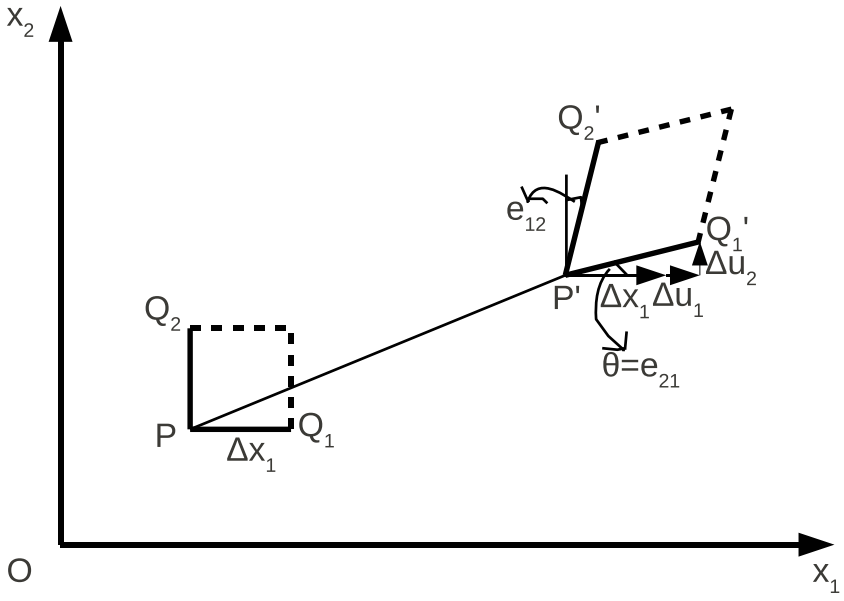


Figure C.2: 2-D Deformation

of the position, i.e. x_1 and x_2, we have

$$\begin{aligned} \Delta u_1 &= \frac{\partial u_1}{\partial x_1}\Delta x_1 + \frac{\partial u_1}{\partial x_2}\Delta x_2 \\ \Delta u_2 &= \frac{\partial u_2}{\partial x_1}\Delta x_1 + \frac{\partial u_2}{\partial x_2}\Delta x_2 \end{aligned} \tag{C.4}$$

$$\begin{aligned} \Delta u_1 &= e_{11}\Delta x_1 + e_{12}\Delta x_2 \\ \Delta u_2 &= e_{21}\Delta x_1 + e_{22}\Delta x_2 \end{aligned} \tag{C.5}$$

Now consider a rectangular area element at P.

$$\begin{aligned} \tan\theta &\approx \theta \\ &= \frac{\Delta u_2}{\Delta x_1 + \Delta u_1} \end{aligned} \tag{C.6}$$

For pure rotations (which do not have any significance as our crystal systems are rotationally invariant), the rotation operator equals the strain.

$$\begin{pmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} \tag{C.7}$$

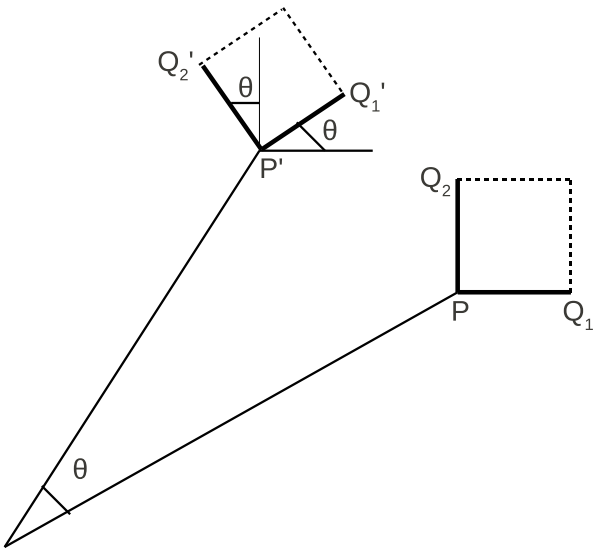


Figure C.3: Pure Rotation

So, for pure rotations, the e_{ij} do not vanish, which means that we need to project them out to obtain the effect of shear and tensile strains.

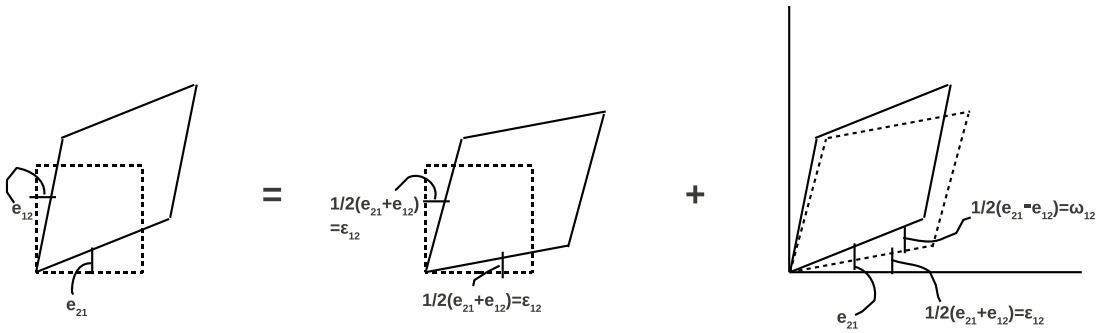


Figure C.4: Components Of Deformation

Any matrix can be written as the sum of symmetric and anti-symmetric matrices in the following manner:

$$\begin{aligned} e_{ij} &= \frac{(e_{ij} + e_{ji})}{2} + \frac{(e_{ij} - e_{ji})}{2} \\ &= \epsilon_{ij} + \omega_{ij} \end{aligned} \tag{C.8}$$

Here ϵ_{ij} is the symmetric strain tensor and ω_{ij} is the anti-symmetric rotation tensor.

C.2.3 3-D Strain

Analogous to the above proof, we can prove that a deformation that causes pure rotation is anti-symmetric in 3-D case as well. u_i are the relative displacements.

$$u_i = \sum_j e_{ij} x_j \tag{C.9}$$

For a small rotation, these u_i's will be orthogonal to the corresponging x_i's. So the dot-product will be zero.

$$u_i x_i = \sum_j e_{ij} x_i x_j = 0 \tag{C.10}$$

Since this is true for all values of x_i, we add all the above equations and get

$$\begin{aligned} e_{ij} &= 0 \quad if \quad i = j \\ &= -e_{ji} \quad if \quad i \neq j \end{aligned} \tag{C.11}$$

which means that e_{ij} is anti-symmetric for pure rotation. So, the 3-D strain tensor ϵ_{ij} is a symmetric tensor with $\epsilon_{ij} = \frac{1}{2}(e_{ij} + e_{ji})$.

C.2.4 Strain and Crystal Symmetry

Strain is not a crystal property but a response to an influence on the crystal system. So, it need not conform to the symmetry of the crystal unless the influence itself conforms to the crystal symmetry. An example of a case where strain conforms to the symmetry is the strain caused by thermal expansion.

C.3 How Does That Affect HMBI?

A symmetric strain tensor means that we cannot use $\epsilon_{12} = 0.001$ in the strain tensor's numerical calculation while keeping $\epsilon_{21} = 0.000$ without avoiding the rotational

component which would eventually lead to asymmetric stress tensor. In the convention used, $\frac{\epsilon_k}{2} = \epsilon_{ij} = \epsilon_{ji}$.

$$strain \quad tensor = \begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{pmatrix} = \begin{pmatrix} \epsilon_1 & \frac{1}{2}\epsilon_6 & \frac{1}{2}\epsilon_5 \\ \frac{1}{2}\epsilon_6 & \epsilon_2 & \frac{1}{2}\epsilon_4 \\ \frac{1}{2}\epsilon_5 & \frac{1}{2}\epsilon_4 & \epsilon_3 \end{pmatrix} \tag{C.12}$$

$$E^{HMBI} = E\left(nuc.coords, latticeparameters, strain\right) \tag{C.13}$$

$$\mathrm{d}E = \sum_{i \neq j} \frac{\partial E}{\partial \epsilon_{ij}} \mathrm{d}\epsilon_{ij} \tag{C.14}$$

Stress tensor terms are given by[187]

$$stress \quad tensor = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} = \begin{pmatrix} \sigma_1 & \frac{1}{2}\sigma_6 & \frac{1}{2}\sigma_5 \\ \frac{1}{2}\sigma_6 & \sigma_2 & \frac{1}{2}\sigma_4 \\ \frac{1}{2}\sigma_5 & \frac{1}{2}\sigma_4 & \sigma_3 \end{pmatrix} \tag{C.15}$$

and

$$\sigma_i = \frac{1}{V} \frac{\partial E}{\partial \epsilon_i} \tag{C.16}$$

For $i \neq j$, that is for off-diagonal terms of the stress tensor,

$$\begin{aligned} \frac{\partial E}{\partial \left(\epsilon_k\right)} &= \frac{\partial E}{\partial \epsilon_{ij}} \frac{\partial \epsilon_{ij}}{\partial \left(\epsilon_k\right)} + \frac{\partial E}{\partial \epsilon_{ji}} \frac{\partial \epsilon_{ji}}{\partial \left(\epsilon_k\right)} \\ &= \frac{\partial \mathbf{E}}{\partial \boldsymbol{\epsilon}_{\mathbf{ij}}} \left(1\right) + \frac{\partial \mathbf{E}}{\partial \boldsymbol{\epsilon}_{\mathbf{ji}}} \left(1\right) \end{aligned} \tag{C.17}$$

For $i = j$, that is for diagonal terms of the stress tensor,

$$\frac{\partial E}{\partial \epsilon_i} = \frac{\partial \mathbf{E}}{\partial \boldsymbol{\epsilon}_{\mathbf{ii}}} \tag{C.18}$$

The terms on the right hand side of the last two equations are already known with HMBI.

C.4 Stiffness Tensor

Stiffness tensor "**C**" relates the stress and strain tensors by the relationship, $\sigma = \mathbf{C}\epsilon$

Each term of the Stiffness tensor is given by $\mathbf{C}_{lml'm'} = \frac{\partial^2 E}{\partial \epsilon_{lm} \partial \epsilon_{l'm'}}$

We know that,

$$\frac{\partial E}{\partial \epsilon_{lm}} = \sum_{i,j=1}^{3} \frac{\partial E}{\partial a'_{ij}} \frac{\partial a'_{ij}}{\partial \epsilon_{lm}} |_{\epsilon_{lm} \to 0} \tag{C.19}$$

$$\frac{\partial^2 E}{\partial \epsilon_{lm} \partial \epsilon_{l'm'}} = \sum_{i,j=1}^{3} \left(\frac{\partial}{\partial \epsilon_{l'm'}} \left(\frac{\partial E}{\partial a'_{ij}} \right) \frac{\partial a'_{ij}}{\partial \epsilon_{lm}} + \frac{\partial E}{\partial a'_{ij}} \frac{\partial}{\partial \epsilon_{l'm'}} \left(\frac{\partial a'_{ij}}{\partial \epsilon_{lm}} \right) \right) \tag{C.20}$$

$$\frac{\partial^2 E}{\partial \epsilon_{lm} \partial \epsilon_{l'm'}} = \sum_{i,j,i',j'=1}^{3} \frac{\partial^2 E}{\partial a'_{ij} \partial a_{i'j'}} \frac{\partial a'_{i'j'}}{\partial \epsilon_{l'm'}} \frac{\partial a'_{ij}}{\partial \epsilon_{lm}} + \sum_{i,j=1}^{3} \frac{\partial E}{\partial a'_{ij}} \frac{\partial^2 a'_{ij}}{\partial \epsilon_{lm} \epsilon_{l'm'}} \tag{C.21}$$

$\frac{\partial^2 a'_{ij}}{\partial \epsilon_{lm} \epsilon_{l'm'}} = 0$ as a'_{ij} are all linear and not quadratic in ϵ. So the above equation reduces to

$$\frac{\partial^2 E}{\partial \epsilon_{lm} \partial \epsilon_{l'm'}} = \sum_{i,j,i',j'=1}^{3} \frac{\partial^2 E}{\partial a'_{ij} \partial a_{i'j'}} \frac{\partial a'_{i'j'}}{\partial \epsilon_{l'm'}} \frac{\partial a'_{ij}}{\partial \epsilon_{lm}} \tag{C.22}$$

Since $\frac{\partial a'_{ij}}{\partial \epsilon_{lm}} = \delta_{jl} a_{im}$, the above equation reduces to

$$\begin{aligned} \frac{\partial^2 E}{\partial \epsilon_{lm} \partial \epsilon_{l'm'}} &= \sum_{i,j,i',j'=1}^{3} \frac{\partial^2 E}{\partial a'_{ij} \partial a'_{i'j'}} \delta_{j'l'} a_{i'm'} \delta_{jl} a_{im} \\ &= \sum_{i,i'=1}^{3} \frac{\partial^2 E}{\partial a'_{il} \partial a'_{i'l'}} a_{i'm'} a_{im} \end{aligned} \tag{C.23}$$

where the term in red is the 9×9 lattice hessian term.

C.5 Stiffness Tensor in Voigt Notation

Remember that in Voigt's notation, $11 = 1$, $22 = 2$, $33 = 3$, $23 = 32 = 4$, $13 = 31 = 5$ and $12 = 21 = 6$.

$$\begin{aligned}\frac{\partial^2 E}{\partial\epsilon_k \partial\epsilon_{k'}} &= \frac{\partial}{\partial\epsilon_k}\left(\frac{\partial E}{\partial\epsilon_{k'}}\right) \\ &= \frac{\partial}{\partial\epsilon_k}\left(\frac{\partial E}{\partial\epsilon_{i'j'}} + \frac{\partial E}{\partial\epsilon_{j'i'}}\right) \\ &= \left(\frac{\partial}{\partial\epsilon_{ij}} + \frac{\partial}{\partial\epsilon_{ji}}\right)\left(\frac{\partial E}{\partial\epsilon_{i'j'}} + \frac{\partial E}{\partial\epsilon_{j'i'}}\right) \\ &= \frac{\partial^2 E}{\partial\epsilon_{ij}\partial\epsilon_{i'j'}} + \frac{\partial^2 E}{\partial\epsilon_{ij}\partial\epsilon_{j'i'}} + \frac{\partial^2 E}{\partial\epsilon_{ji}\partial\epsilon_{i'j'}} + \frac{\partial^2 E}{\partial\epsilon_{ji}\partial\epsilon_{j'i'}}\end{aligned} \tag{C.24}$$

Now the stiffness tensor is a 6×6 matrix.

C.6 Elastic Constants

Elastic constants are the second derivatives of the HMBI energy w.r.t. the six strain components.

$$\begin{aligned}C_{pq} &= \frac{1}{V}\frac{\partial^2 E}{\partial\epsilon_p \partial\epsilon_q} \\ &= \frac{\partial^2 E}{\partial\epsilon_{lm}\partial\epsilon_{l'm'}}\end{aligned} \tag{C.25}$$

Again using the operator math, we get

$$\frac{\partial}{\partial\epsilon_{lm}} = \sum_{i=1}^{3} u_{im}\frac{\partial}{\partial u_{il}} \tag{C.26}$$

and similarly,

$$\frac{\partial}{\partial\epsilon_{l'm'}} = \sum_{j=1}^{3} u_{jm'}\frac{\partial}{\partial u_{jl'}} \tag{C.27}$$

$$\frac{\partial^2 E}{\partial\epsilon_{lm}\partial\epsilon_{l'm'}} = \sum_{i,j=1}^{3} u_{im}\frac{\partial u_{jm'}}{\partial u_{il}}\frac{\partial E}{\partial u_{jl'}} + \sum_{i,j=1}^{3} u_{im}u_{jm'}\frac{\partial^2 E}{\partial u_{il}\partial u_{jl'}} \tag{C.28}$$

As far as our HMBI code goes, the calculations for $\frac{\partial^2 E'}{\partial u_{il} \partial u_{jl'}}$ are already done when we calculate the lattice hessian. So, the only additional piece is to convert the derivatives of E_{PBC}^{MM} terms w.r.t. the six lattice parameters to nine parameters, which is trivial but symmetry dependent.

Bibliography

[1] A. M. Reilly, D. S. Middlemiss, M. M. Siddick, D. A. Wann, G. J. Ackland, C. C. Wilson, D. W. H. Rankin, , and C. A. Morrison. The Phonon Spectrum of Phase-I Ammonia: Reassignment of Lattice Mode Symmetries from Combined Molecular and Lattice Dynamics Calculations. *J. Phys. Chem. A*, 112:1322–1329, 2008.

[2] M. Svensson, S. Humbel, R. D. Froese, T. Matasubara, S. Sieber, , and K. Morokuma. ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels -Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.*, 100:19357, 1996.

[3] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayasi. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* , 313:701706, 1999.

[4] D. G. Fedorov and K. Kitaura. Extending the Power of Quantum Chemistry to Large Systems with the Fragment Molecular Orbital Method. *J. Phys. Chem. A* , 111:69046914, 2007.

[5] S. Hirata, M. Valiev, M. Dupuis, S. S. Xantheas, S. Sugiki, and H. Sekino. Fast electron correlation methods for molecular clusters in the ground and excited states. *Mol. Phys.* , 103:22552265, 2005.

[6] E. E. Dahlke and D. G. Truhlar. Assessment of the Pairwise Additive Approximation and Evaluation of Many-Body Terms for Water Clusters. *J. Phys. Chem. B* , 3:1059510601, 2006.

[7] B. W. Hopkins and G. S. Tschumper. Multicentered QM/QM methods for Overlapping Model Systems. *Mol. Phys.* , 103:309315, 2005.

[8] A. Hermann and P. Schwerdtfeger. Complete basis set limit second-order Møller-Plesset calculations for the fcc lattices of neon, argon, krypton and xenon. *J. Chem. Phys.* , 131:244508, 2009.

[9] G. J. O. Beran. Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields. *J. Chem. Phys.*, 130:164115, 2009.

[10] A. Sebetci and G. J. O. Beran. Spatially homogeneous QM/MM for systems of interacting molecules with on-the-fly *ab initio* force-field parameterization. *J. Chem. Theory Comput.*, 6:155–167, 2010.

[11] G. J. O. Beran and K. Nanda. Predicting organic crystal lattice energies with chemical accuracy. *J. Phys. Chem. Lett.*, 1:3480–3487, 2010.

[12] S. Wen and G. J. O. Beran. Accurate molecular crystal lattice energies from a fragment QM/MM approach with on-the-fly *ab initio* force-field parameterization. *J. Chem. Theory Comput.*, 7:3733–3742, 2011.

[13] E. A. L. Gillis, M. Demireva, K. Nanda, G. J. O. Beran, E. R. Williams, and T. D. Fridgen. Structures and energetics of electrosprayed uracil$_n$Ca^{2+} clusters (n=14-4) in the gas phase. *Phys. Chem. Chem. Phys.*, 14:3304–3315, 2012.

[14] S. Wen, K. Nanda, Y. Huang, and G. J. O. Beran. Practical quantum mechanics-based fragment methods for predicting molecular crystal properties. *Phys. Chem. Chem. Phys.*, 14:7578–7590, 2012.

[15] S. Wen and G. J. O. Beran. Accidental degeneracy in crystalline aspirin: New insights from high-level ab initio calculations. *Cryst. Growth Des.*, 12:2169–2172, 2012.

[16] K. D. Nanda and G. J. O. Beran. Prediction of organic molecular crystal geometries from MP2-level fragment quantum mechanical/molecular mechanical calculations. *J. Chem. Phys.* , 137:174106, 2012.

[17] S. Haas, A. F. Stassen, G. Schuck, K. P. Pernstich, D. J. Gundlach, B. Batlogg, U. Berens, and H. J. Kirner. High charge-carrier mobility and low trap density in a rubrene derivative. *Phys. Rev. B: Condens. Matter*, 76:115203, 2007.

[18] S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley, and K. Mcfarland. Dealing with the impact of ritnavir polymorphs on the late stages of bulk. *Org. Process Res. Dev.*, 4:413–417, 2000.

[19] J. Bauer, S. Spanton, R. Quick, J. Quick, W. Dziki, W. Porter, and J. Morris. Ritonavir: an extraordinary example of conformational polymorphism. *Pharm. Res.*, 18:859–866, 2001.

[20] M. Born and R. Oppenherimer. Zur Quantentheorie der Molekeln. *Annalen der Physik*, 389(20):457–484, 1927.

[21] F. Jensen. Introduction to Computational Chemistry and 2nd ed., 2007.

[22] Chr. Møller and M. S. Plesset. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.*, 46:618–622, 1934.

[23] L. Kong, F. A. Bischoff, and E. F. Valeev. Explicitly Correlated R12/F12 Methods for Electronic Structure . *Chem. Rev.*, 112:75–107, 2012.

[24] C. Hattig, W. Klopper, A. Kohn, and D. P. Tew. Explicitly Correlated Electrons in Molecules. *Chem. Rev.*, 112:4–74, 2011.

[25] O. Marchetti and H.-J. Werner. Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions. *Phys. Chem. Chem. Phys.*, 10:3400–3409, 2008.

[26] K. M. de Lange and J. R. Lane. Explicit correlation and intermolecular interactions: Investigating carbon dioxide complexes with the CCSD(T)-F12 method. *J. Chem. Phys.*, 134:034301, 2011.

[27] Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon. Scaled opposite-spin second order MøllerPlesset correlation energy: An economical electronic structure method . *J. Chem. Phys.*, 121:97939802, 2004.

[28] T. Takatani, E. G. Hohenstein, and C. D. Sherrill. Improvement of the Coupled-Cluster Singles and Doubles Method Via Scaling Same- and Opposite-Spin Components of the Double Excitation Correlation Energy. *J. Chem. Phys.*, 128:124111, 2010.

[29] M. Pitonak, J. Rezac, and P. Hobza. Spin-component scaled coupled-clusters singles and doubles optimized towards calculation of noncovalent interactions. *Phys. Chem. Chem. Phys.*, 12:96119614, 2010.

[30] M. Gerenkamp and S. Grimme. Spin-component scaled second-order MøllerPlesset perturbation theory for the calculation of molecular geometries and harmonic vibrational frequencies. *Chem. Phys. Lett.*, 392:229–235, 2004.

[31] R. A. Distasio and M. Head-Gordon. Optimized spin-component scaled second-order Møller-Plesset perturbation theory for intermolecular interaction energies. *Mol. Phys.*, 105:1073–1083, 2007.

[32] D. Usvyat, L. Maschio, F. R. Manby, S. Casassa, C. Pisani, , and M. Schtz. Fast local-MP2 method with density-fitting for crystals. II. Test calculations and applications to the carbon dioxide crystal. *Phys. Rev. B*, 76:075102, 2007.

[33] C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schtz, , and D. Usvyat. Periodic local MP2 method for the study of electronic correlation in crystals: Theory and preliminary applications. *J. Comput. Chem.*, 29:2113, 2008.

[34] T. D. Crawford and H. F. Schaefer III. An introduction to coupled cluster theory for computational chemists. *Rev. Comp. Chem.*, 14:33–136, 2000.

[35] K. Raghavachari, G. Trucks, J. A. Pople, , and M. Head-Gordon. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.*, 157:479–483, 1989.

[36] J. D. Watts, Jürgen Gauss, , and R. J. Bartlett. Coupled-cluster methods with non-iterative triple excitations for restricted open-shell Hartree-Fock and other general single determinant reference functions. Energies and analytical gradients. *J. Chem. Phys.*, 98(11):8718–8733, 1993.

[37] R. G. Parr and W. Yang. Density-Functional Theory of Atoms and Molecules, 1990.

[38] B. Civalleri, C. M. Zicovich-Wilson, L. Valenzano, and P. Ugliengo. B3LYP augmented with an empirical dispersion term (B3LYP-D*) as applied to molecular crystals. *CrystEngComm* , 10:405–410, 2008.

[39] S. Grimme. Density functional theory with London dispersion corrections. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* , 1:211–228, 2011.

[40] J. E. Subotnik, A. Sodt, and M. Head-Gordon. The limits of local correlation theory: Electronic delocalization and chemically smooth potential energy surfaces. *J. Chem. Phys.*, 128:034103, 2005.

[41] P. Ren and J. W. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B*, 107:5933–5947, 2003.

[42] B. I. Dunlap. Fitting the Coulomb potential variationally in $X\alpha$ molecular calculations. *J. Chem. Phys.*, 78:3140, 1983.

[43] M. W. Feyereisen, G. Fitzgerald, , and A. Komornicki. Use of approximate integrals in *ab initio* theory. An application in MP2 energy calculations. *Chem. Phys. Lett.*, 208:359, 1993.

[44] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 240:283, 1995.

[45] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007, 1989.

[46] F. Weigend, A. Köhn, and C. Hättig. Efficient use of correlation consistent basis sets in the resolution of the identity MP2 calculations. *J. Chem. Phys.*, 116:3175, 2002.

[47] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. Distasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A .D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Z. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.*, 8:3172–3191, 2006.

[48] J. W. Ponder and TINKER v4.2, 2004 and `http://dasher.wustl.edu/tinker/`. Accessed January 23 and 2008.

[49] J. Kästner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander, and P. Sherwood. DL-FIND: An open-source geometry optimizer for atomistic simulations. *J. Phys. Chem. A*, 113:11856–11865, 2009.

[50] R. P. Steele, R. A. Distasio Jr., Y. Shao, J. Kong, and M. Head-Gordon. Dual-basis second-order Møller-Plesset perturbation theory: A reduced cost reference for correlation calculations. *J. Chem. Phys.*, 125:074108, 2006.

[51] T. D. Crawford, C. D. Sherill, E. F. Valeev, J. T. Fermann, R. A. King, M. L. Leininger, S. T. Brown, C. L. Janssen, E. T. Seidl, and J. P. Kenny et al. PSI3: An open-source Ab Initio electronic structure package. *J. Comput. Chem.*, 28:1610–1616, 2007.

[52] Y. Eichen, E. Braun, U. Sivan, and G. Ben-Yoseph. Self-assembly of nanoelectronic components and circuits using biological templates. *Acta Polym.*, 49:663, 1998.

[53] K. Ariga and T. Kunitake. Supramolecular Chemistry Fundamentals and Applications, Springer and New York and 2006.

[54] J. T. Davis. G-Quartets 40 Years Later: From 5'-GMP to Molecular Biology and Supramolecular Chemistry . *Angew. Chem. and Int. Ed*, 43:668, 2004.

[55] T. Aggerholm, S. C. Nanita, K. J. Koch, and R. G. Cooks. Clustering of nucleosides in the presence of alkali metals: Biologically relevant quartets of guanosine, deoxyguanosine and uridine observed by ESIMS/MS. *J. Mass Spectrom.*, 38:87, 2003.

[56] M. Cai, X. Shi, V. Sidorov, D. Fabris, Y.-F. Lam, and J. T. Davis. Cation-directed self-assembly of lipophilic nucleosides: the cation's central role in the structure and dynamics of a hydrogen-bonded assembly . *Tetrahedron*, 58:661, 2002.

[57] K. Fukushima and H. Iwahashi. 1:1 Complex of guanine quartet with alkali metal cations detected by electrospray ionization mass spectrometry. *Chem. Commun.*, page 895, 2000.

[58] J. Gu and J. Leszczynski. A Remarkable Alteration in the Bonding Pattern: An HF and DFT Study of the Interactions between the Metal Cations and the Hoogsteen Hydrogen-Bonded G-Tetrad. *J. Phys. Chem. A*, 104:6308, 2000.

[59] I. Manet, L. Francini, S. Masiero, S. Pieraccini, Spada, P. Gian, and G. Gottarelli. An ESI-MS and NMR Study of the Self-Assembly of Guanosine Derivatives. *Helv. Chim. Acta*, 84:2096, 2001.

[60] R. Otero, M. Schock, L. M. Molina, E. Lgsgaard, I. Stensgaard, B. Hammer, and F. Besenbacher. Guanine Quartet Networks Stabilized by Cooperative Hydrogen Bonds. *Angew. Chem. and Int. Ed.*, 44:2270, 2005.

[61] W. I. Sundquist and A. Klug. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, 342:825, 1989.

[62] C. Cheong and P. B. Moore. Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry*, 31:8406, 1992.

[63] E. Freisinger, A. Schimanski, and B. Lippert. Thymine-metal ion interactions: Relevance for thymine quartet structures. *J. Biol. Inorg. Chem.*, 6:378, 2001.

[64] K. J. Koch, T. Aggerholm, S. C. Nanita, and R. G. Cooks. Clustering of nucleobases with alkali metals studied by electrospray ionization tandem mass spectrometry: implications for mechanisms of multistrand DNA stabilization. *J. Mass Spectrom.*, 37:676, 2002.

[65] M. Meyer, T. Steinke, M. Brandl, and J. Suhnel. Density functional study of guanine and uracil quartets and of guanine/metal ion complexes. *J. Comput. Chem.*, 22:109, 2001.

[66] H. Witkowski, E. Freisinger, and B. Lippert. Uracil quartet formation through non-covalent interaction with a neutral metal ammine complex. *Chem. Commun.*, page 1315, 1997.

[67] B. Fischer, H. Preut, B. Lippert, H. Schollhorn, and U. Thewalt. Alkali-Nucleobase Interactions: Involvement of Exocyclic Oxygens of 1-Methyluracil and 1-Methylthymine in Na^+ Binding . *Polyhedron*, 9:2199, 1990.

[68] B. Qiu, J. Liu, Z. Qin, G. Wang, and H. Luo. Quintets of uracil and thymine: a novel structure of nucleobase self-assembly studied by electrospray ionization mass spectrometry. *Chem. Commun.*, page 2863, 2009.

[69] E.-L. Zins, S. Rochut, and C. Pepe. Formation of complexes between uracil and calcium ions: an ESI/MS/MS study in combination with theoretical calculations. *J. Mass Spectrom.*, 44:813, 2009.

[70] E.-L. Zins, S. Rochut, and C. Pepe. Theoretical and experimental studies of cationized uracil complexes in the gas phase. *J. Mass Spectrom.*, 44:40, 2009.

[71] R. C. Dunbar. BIRD (Blackbody Infrared Radiative Dissociation). *Mass Spectrom. Rev.*, 23:127, 2004.

[72] W. D. Price and E. R. Williams. Activation of peptide ions by blackbody radiation: Factors that lead to dissociation kinetics in the rapid energy exchange limit. *J. Phys. Chem. A*, 101:8844, 1997.

[73] J. S. Prell, T. M. Chang, J. A. Biles, G. Berden, J. Oomens, and E. R. Williams. Isomer Population Analysis of Gaseous Ions from Infrared Multiple Photon Dissociation Kinetics. *J. Phys. Chem. A*, 115:2745, 2011.

[74] R. C. Dunbar. Kinetics of Thermal Unimolecular Dissociation by Ambient Infrared Radiation. *J. Phys. Chem.*, 98:8705, 1994.

[75] R. C. Dunbar, T. B. McMahon, D. Tholmann, D. S. Tonner, D. W. Salahub, and D. Wei. Zero-Pressure Thermal Radiation Induced Dissociation (ZTRID) of Gas-Phase Cluster Ions: Comparison of Theory and Experiment for $(H_2O)_2Cl^-$ and $(H_2O)_3Cl^-$. *J. Am. Chem. Soc.*, 117:12819, 1995.

[76] W. D. Price, P. D. Schnier, and E. R. Williams. Binding Energies of the Proton-Bound Amino Acid Dimers Gly - Gly, Ala - Ala, Gly - Ala, and Lys - Lys Measured by Blackbody Infrared Radiative Dissociation. *J. Phys. Chem. B*, 101:664, 1997.

[77] D. Zhang and R. G. Cooks. Doubly charged cluster ions (NaCl)(m)(Na)(2) (2+): magic numbers, dissociation, and structure . *Int. J. Mass Spectrom.*, 195-196:667, 2000.

[78] N. J. Kim, H. Kang, G. Jeong, Y. S. Kim, K. T. Lee, and S. K. Kim. Even-odd alternation in mass spectrum of thymine and uracil clusters: evidence of intracluster photodimerization. *Proc. Natl. Acad. Sci. U. S. A.*, 98:4841, 2001.

[79] S. E. Rodriguez-Cruz, R. A. Jockusch, and E. R. Williams. Hydration Energies and Structures of Alkaline Earth Metal Ions, M^{2+} (H2O) n, n= 5-7, M= Mg, Ca, Sr, and Ba. *J. Am. Chem. Soc.*, 121:8898, 1999.

[80] Considering that our laser probably does not completely overlap with the orbiting ion packet, the estimation of half in this case is greater than the true ratio. From the baseline of Fig. 8iii, and the ratio is likely 60/40 at best.

[81] E. A. L. Gillis, K. Rajabi, and T. D. Fridgen. Structures of Hydrated Li+-Thymine and Li+-Uracil Complexes by IRMPD Spectroscopy in the N-H/O-H Stretching Region. *J. Phys. Chem. A*, 113:824, 2009.

[82] M. T. Rodgers and P. B. Armentrout. Noncovalent Interactions of the Nucleic Acid Bases (Uracil, Thymine, and Adenine) with Alkali Metal Ions. Threshold Collision-Induced Dissociation and Theoretical Studies. *J. Am. Chem. Soc.*, 122:8548, 2000.

[83] N. Russo, M. Toscano, and A. Grand. Lithium affinity for DNA and RNA nucleobases. The role of theoretical information in the elucidation of the mass spectrometry data. *J. Phys. Chem. B*, 105:4735, 2001.

[84] N. Russo, M. Toscano, and A. Grand. Bond energies and attachments sites of sodium and potassium cations to DNA and RNA nucleic acid bases in the gas phase. *J. Am. Chem. Soc.*, 123:10272, 2001.

[85] C. Trujillo, A. M. Lamsabhi, O. Mo, M. Yanez, and J.-Y. Salpin. Interaction of Ca^{2+} with uracil and its thio derivatives in the gas phase. *Org. Biomol. Chem.*, 6:3695, 2008.

[86] for example An attractive interaction between aromatic rings. For more information, see, C. A. Hunter, and J. K. M. Sanders. The nature of pi-pi interactions . *J. Am. Chem. Soc.*, 112:5525–5534, 1990.

[87] S L Price. Computational prediction of organic crystal structures and polymorphism. *Int. Rev. Phys. Chem.*, 27(3):541–568, July 2008.

[88] T. Li and S. Feng. Empirically augmented density functional theory for predicting lattice energies of aspirin, acetaminophen polymorphs, and ibuprofen homochiral and racemic crystals. *Pharm. Res.*, 23:2326–2332, 2006.

[89] J. Kleis, B. I. Lundqvist, D. C. Langreth, and E. Schröder. Towards a working density-functional theory for polymers: First-principles determination of the polyethylene crystal structure. *Phys. Rev. B*, 76:100201, 2007.

[90] M. A. Neumann and M. A. Perrin. Energy ranking of molecular crystals using density functional theory calculations and an empirical van der Waals correction. *J. Phys. Chem. B*, 109:15531–15541, 2005.

[91] M. A. Neumann, F. J. J. Leusen, and J. Kendrick. A major advance in crystal structure prediction. *Angew. Chem. Int. Ed.*, 47:2427–2430, 2008.

[92] Bartolomeo Civalleri, Claudio M. Zicovich-Wilson, Loredana Valenzano, and Piero Ugliengo. B3LYP augmented with an empirical dispersion term (B3LYP-D*) as applied to molecular crystals. *CrystEngComm*, 10:405–410, 2008.

[93] Panagiotis G Karamertzanis, Graeme M Day, Gareth W A Welch, John Kendrick, Frank J J Leusen, Marcus A Neumann, and Sarah L Price. Modeling the interplay of inter- and intramolecular hydrogen bonding in conformational polymorphs. *J. Chem. Phys.*, 128(24):244708, June 2008.

[94] Matthew Habgood, S L Price, Gustavo Portalone, and Simona Irrera. Testing a Variety of Electronic-Structure-Based Methods for the Relative Energies of 5-Formyluracil Crystals. *J. Chem. Theory Comput.*, 7:2685–2688, 2011.

[95] D. C. Sorescu and B. M. Rice. Theoretical predictions of energetic molecular crystals at ambient and hydrostatic compression conditions using dispersion corrections to conventional density functionals (DFT-D). *J. Phys. Chem. C*, 114:6734–6748, 2010.

[96] Radhakrishnan Balu, Edward F C Byrd, and Betsy M Rice. Assessment of dispersion corrected atom centered pseudopotentials: application to energetic molecular crystals. *J. Phys. Chem. B*, 115(5):803–10, February 2011.

[97] Fuyuki Shimojo, Zhongqing Wu, Aiichiro Nakano, Rajiv K Kalia, and Priya Vashishta. Density functional study of 1,3,5-trinitro-1,3,5-triazine molecular crystal with van der Waals interactions. *J. Chem. Phys.*, 132(9):094106, March 2010.

[98] D. Lu, Y. Li, D. Rocca, and G. Galli. *Ab initio* calculation of van der Waals bonded molecular crystals. *Phys. Rev. Lett.*, 102:206411, 2009.

[99] Y. Li, D. Lu, H.-V. Nguyen, and G. Galli. van der Waals interactions in molecular assemblies from first-principles calculations. *J. Phys. Chem. A*, 114:1944–1952, 2010.

[100] Lorenzo Maschio, Bartolomeo Civalleri, Piero Ugliengo, and Angelo Gavezzotti. Intermolecular interaction energies in molecular crystals: comparison and agreement of localized Møller-Plesset 2, dispersion-corrected density functional, and classical empirical two-body calculations. *J. Phys. Chem. A*, 115(41):11179–86, October 2011.

[101] John Kendrick, Frank J. J. Leusen, Marcus A. Neumann, and Jacco van de Streek. Progress in Crystal Structure Prediction. *Chem. Eur. J.*, 17(38):10736–10744, September 2011.

[102] N Marom, A Tkatchenko, S Kapishnikov, L Kronik, and L Leiserowitz. Structure and Formation of Synthetic Hemozoin: Insights from First-Principles Calculations. *Cryst. Growth Des.*, 11(8):3332–3341, August 2011.

[103] B Schatschneider, J-J Liang, S Jezowski, and A Tkatchenko. Phase transition between cubic and monoclinic polymorphs of the tetracyanoethylene crystal: the role of temperature and kinetics. *CrystEngComm*, 14:4656–4663, 2012.

[104] R. A. DiStasio, O. A. von Lilienfeld, and A. Tkatchenko. Collective many-body van der Waals interactions in molecular systems. *Proc. Nat. Acad. Sci.*, 109:14791–14795, August 2012.

[105] K. Hongo, M. A. Watson, R. S. Sanchez-Carrera, T. Iitaka, and A. Aspuru-Guzik. Failure of conventional density functionals for the prediction of molecular crystal polymorphism: A quantum Monte Carlo study. *J. Phys. Chem. Lett.*, 1:17891794, 2010.

[106] D. Usvyat, L. Maschio, F. R. Manby, S. Casassa, C. Pisani, and M. Schütz. Fast local-MP2 method with density-fitting for crystals. II. Test calculations and applications to the carbon dioxide crystal. *Phys. Rev. B*, 76:075102, 2007.

[107] A. Erba, C. Pisani, S. Casassa, L. Maschio, M. Schütz, and D. Usvyat. MP2 versus density-functional theory study of the Compton profiles of crystalline urea. *Phys. Rev. B*, 81:165108, 2010.

[108] L. Maschio, D. Usvyat, M. Schütz, and B. Civalleri. Periodic local Møller-Plesset second order perturbation theory method applied to molecular crystals: Study of solid NH_3 and CO_2 using extended basis sets. *J. Chem. Phys.*, 132:134706, 2010.

[109] M. Marsman, A. Grueneis, J. Paier, and G. Kresse. Second-order Møller-Plesset perturbation theory applied to extended systems. I. Within the projector-augmented-wave formalism using a plane wave basis set. *J. Chem. Phys*, 130:184103, 2009.

[110] So Hirata and Suehiro Iwata. Analytical energy gradients in second-order Moller-Plesset perturbation theory for extended systems. *J. Chem. Phys.*, 109(11):4147–4155, 1998.

[111] Mark S Gordon, Dmitri G Fedorov, Spencer R Pruitt, and L Slipchenko. Fragmentation Methods: A Route to Accurate Calculations on Large Systems. *Chem. Rev.*, 112:632–672, August 2012.

[112] So Hirata. Fast electron-correlation methods for molecular crystals: an application to the alpha, beta(1), and beta(2) modifications of solid formic acid. *J. Chem. Phys.*, 129(20):204104, November 2008.

[113] Olaseni Sode, Murat Keceli, So Hirata, and Kiyoshi Yagi. Coupled-Cluster and Many-Body Perturbation Study of Energies, Structures, and Phonon Dispersions of Solid Hydrogen Fluoride Phonon Dispersions. *Int. J. Quant. Chem.*, 109:1928–1939, 2009.

[114] Heather M Netzloff and Michael A Collins. Ab initio energies of nonconducting crystals by systematic fragmentation. *J. Chem. Phys.*, 127(13):134113, October 2007.

[115] Michael A Collins. Ab initio lattice dynamics of nonconducting crystals by systematic fragmentation. *J. Chem. Phys.*, 134(16):164110, April 2011.

[116] E. E. Dahlke and D. G. Truhlar. Electrostatically embedded many-body correlation energy, with applications to the calculation of accurate second-order Møller-Plesset perturbation theory energies for large water clusters. *J. Chem. Theory Comput.*, 3:1342–1348, 2007.

[117] Andreas Hermann and Peter Schwerdtfeger. Ground-State Properties of Crystalline Ice from Periodic Hartree-Fock Calculations and a Coupled-Cluster-Based Many-Body Decomposition of the Correlation Energy. *Phys. Rev. Lett.*, 101(18):183005, October 2008.

[118] O. Bludsky, M. Rubes, and P. Soldan. *Ab initio* investigation of intermolecular interactions in solid benzene. *Phys. Rev. B*, 77:092103, 2008.

[119] Darragh P O Neill, Neil L Allan, and F R Manby. Ab initio Monte Carlo simulations of liquid water. In F. Manby, editor, *Accurate Quantum Chemistry in the Condensed Phase*, pages 163–193. CRC Press, Boca Raton, FL, 2010.

[120] S. Tsuzuki, H. Orita, K. Honda, and M. Mikami. First-principles lattice energy calculation of urea and hexamine crystals by a combination of periodic DFT and MP2 two-body interaction energy calculations. *J. Phys. Chem. B*, 114:6799–6805, 2010.

[121] Christopher R Taylor, Peter J Bygrave, Judy N Hart, Neil L Allan, and Frederick R Manby. Improving density functional theory for crystal polymorph energetics. *Phys. Chem. Chem. Phys.*, 14:7739–7743, March 2012.

[122] Carsten Müller and Beate Paulus. Wavefunction-based electron correlation methods for solids. *Phys. Chem. Chem. Phys.*, 14:7605–7614, February 2012.

[123] S. Wen and G. J. O. Beran. . *J. Chem. Theory Comput.*, 8:2698–2705, 2012.

[124] R. Podeszwa, B. M. Rice, and K. Szalewicz. Predicting structures of molecular crystals from first-principles. *Phys. Rev. Lett.*, 101:115503, 2008.

[125] O Anatole von Lilienfeld and Alexandre Tkatchenko. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. *J. Chem. Phys.*, 132(23):234109, June 2010.

[126] F. Weigend, A. Köhn, and C. Hättig. Efficient use of correlation consistent basis sets in the resolution of the identity MP2 calculations. *J. Chem. Phys.*, 116:3175–3183, 2002.

[127] Liu, D C, Nocedal, and J. On the limited memory BFGS method for large scale optimization. *Math. Program.*, 45:503–528, 1989.

[128] R. Dovesi, R. Orlando, B. Civalleri, C. Roetti, V. R. Saunders, and C. M. Zicovich-Wilson. CRYSTAL: A computational tool for the *ab initio* study of the electronic properties of crystals. *Z. Kristallogr.*, 220:571–573, 2005.

[129] R Dovesi, V R Saunders, C Roetti, R Orlando, C M Zicovich-Wilson, F Pascale, B Civalleri, K Doll, N M Harrison, I J Bush, Ph D'Arco, M Llunell, Computational Science, and Advanced Technologies. *CRYSTAL09 User's Manual*. University of Torino, Torino, 2009.

[130] Stefan Grimme. Density functional theory with London dispersion corrections. *WIRES: Comput. Mol. Sci.*, 1(2):211–228, March 2011.

[131] W. J. Hehre, R. Ditchfield, and J. A. Pople. Self-consistent molecular orbital methods. XII. Further extensions of Gaussian-type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.*, 56:2257–2261, 1972.

[132] A. Schafer, H. Horn, and R. Ahlrichs. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.*, 97:2571–, 1992.

[133] See supplementary material at http://dx.doi.org/10.1063/1.4764063 for fractional coordinates and lattice parameters for all optimized structures reported here.

[134] D. Zobel, P. Luger, W. Dreissig, and T. Koritsanszky. Charge density studies on small organic molecules around 20 K: oxalic acid dihydrate at 15 K and acetamide at 23 K. *Acta Cryst. B*, 48(6):837–848, December 1992.

[135] W I F David, R M Ibberson, G A Jeffrey, and J R Ruble. The crystal structure analysis of deuterated benzene and deuterated nitromethane by pulsed-neutron powder diffraction: a comparison with single crystal neutron diffraction analysis. *Physica B*, 181:597–600, 1992.

[136] E. D. Stevens. Low-temperature experimental electron density distribution of formamide. *Acta Cryst. B*, 34(2):544–551, February 1978.

[137] R K McMullan, Joel Epstein, John R Ruble, and B M Craven. The crystal structure of imidazole at 103 K by neutron diffraction. *Acta. Cryst. B*, 35:688–691, March 1979.

[138] G A Jeffrey, J R Ruble, R K Mcmullan, D J Defrees, J S Binkley, and J A Pople. Neutron Diffraction at 23 K and ab initio Molecular-Orbital Studies of the Molecular Structure of Acetamide. *Acta Cryst. B*, 36:2292–2299, 1980.

[139] I. Morrison, J.-C. Li, S. Jenkins, S. S. Xantheas, and M. C. Payne. *Ab initio* total energy studies of the static and dynamical properties of ice Ih. *J. Phys. Chem. B*, 101:6146–6150, 1997.

[140] W. F. Kuhs and M. S. Lehmann. The structure of ice-Ih. *Water Sci. Rev.*, 2:1–65, 1986.

[141] K. Röttger, A. Endriss, J. Ihringer, S. Doyle, and W. F. Kuhs. Lattice constants and thermal expansion of H_2O and D_2O ice Ih between 10 and 265 K. *Acta Cryst. B*, 50:644–648, 1994.

[142] T Beyer and S L Price. The errors in lattice energy minimisation studies: sensitivity to experimental variations in the molecular structure of paracetamol. *CrystEngComm*, 2(34):183, 2000.

[143] Muneaki Kamiya, So Hirata, and Marat Valiev. Fast electron correlation methods for molecular clusters without basis set superposition errors. *J. Chem. Phys.*, 128(7):074103, February 2008.

[144] M. O. Sinnokrot and C. D. Sherrill. High-accuracy quantum mechanical studies of π–π interactions in benzene dimers. *J. Phys. Chem. A*, 110:10656–10668, 2006.

[145] A. L. Ringer and C. D. Sherrill. First-principles computation of lattice energies of organic solids: The benzene crystal. *Chem. Eur. J.*, 14:2542–2547, 2008.

[146] James Alexander Chisholm and W D S Motherwell. COMPACK: A program for identifying crystal structure similarity using distances. *J. Appl. Crystall.*, 38(1):228–231, January 2005.

[147] C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek, and P. A. Wood. Mercury CSD 2.0—New features for the visualization and investigation of crystal structures. *J. Appl. Cryst.*, 41:455–470, 2008.

[148] A.S. Raw, M.S. Furness, D.S. Gill, R.C. Adams, F.O. Holcombe, and L.X. Yu. Regulatory considerations of pharmaceutical solid polymorphism in Abbreviated New Drug Applications (ANDAs). *Adv. Drug Deliv. Rev.*, 56(3):397, 2004.

[149] A. J. Fitzgerald, B. E. Cole, and P. F. Taday. Nondestructive analysis of tablet coating thicknesses using terahertz pulsed imaging. *J. Pharm. Sci.*, 94:177, 2004.

[150] C. J. Strachan, P. F. Taday, D. A. Newnham, K. C. Gordon, J. A. Zeitler, M. Pepper, and T. Rades. Using terahertz pulsed spectroscopy to quantify pharmaceutical polymorphism and crystallinity. *J. Pharm. Sci.*, 94:837, 2005.

[151] J. A. Zeitler, P. F. Taday, D. A. Newnham, M. Pepper, K. C. Gordon, and T. Rades. Terahertz pulsed spectroscopy and imaging in the pharmaceutical setting–a review. *J. Pharm. Pharmacol.*, 59:209, 2007.

[152] M. D. King, W. D. Buchanan, and T. M. Korter. Identification and Quantification of Polymorphism in the Pharmaceutical Compound Diclofenac Acid by Terahertz Spectroscopy and Solid-State Density Functional Theory. *Anal. Chem.*, 83:3786, 2011.

[153] E. M. Witko, W. D. Buchanan, , and T. M. Korter. Terahertz Spectroscopy and Solid-State Density Functional Theory Simulations of the Improvised Explosive Oxidizers Potassium Nitrate and Ammonium Nitrate. *J. Phys. Chem. A*, 115:1241012418, 2011.

[154] C. M. Freeman, J. W. Andzelm, C. S. Ewig, J.-R. Hill, and B. Delley. The structure and energetics of glycine polymorphs based on first principles simulation using density functional theory . *Chem. Commun.*, page 24552456, 1999.

[155] S. A. Rivera, D. G. Allis, and B. S. Hudson. Importance of Vibrational Zero-Point Energy Contribution to the Relative Polymorph Energies of Hydrogen-Bonded Species. *Cryst. Growth Des.*, 8:39053907, 2008.

[156] M. A. Neumann, F. J. J. Leusen, and J. Kendrick. A major advance in crystal structure prediction. *Angew. Chem.,Int. Ed.*, 47:24272430, 2008.

[157] M. Born and H. Kun. Dynamical Theory of Crystal Lattices, 1954.

[158] Th. Strassle, A. M. Saitta, S. Klotz, and M. Braden. Phonon Dispersion of Ice under Pressure. *Phys. Rev. Lett.*, 93:225901, 2004.

[159] O. Sode and S. Hirata. Second-order many-body perturbation study of solid hydrogen fluoride under pressure. *Phys. Chem. Chem. Phys.*, 14:7765–7779, 2012.

[160] X. He, O. Sode, S. S. Xantheas, and S. Hirata. Second-order many-body perturbation study of ice Ih. *J. Chem. Phys.*, 137:204505, 2012.

[161] A. A. Reilly and A. Tkatchenko. Seamless and Accurate Modeling of Organic Molecular Materials. *J. Phys. Chem. Lett.*, 4(6):1028–1033, 2013.

[162] A. Otero de-la Roza and Erin R. Johnson. A benchmark for non-covalent interactions in solids. *J. Chem. Phys.*, 137:054103, 2012.

[163] H. J. Monkhorst and J. D. Pack. Special points for Brillouin-zone integrations. *Phys. Rev. B*, 13:5188, 1976.

[164] R. Feistel and W. Wagner. A New Equation of State for H_2O Ice Ih. *J.Phys. Chem. Ref. Data*, 35(2):1021, 2006.

[165] R. Feistel and W. Wagner. Sublimation pressure and sublimation enthalpy of H_2O ice Ih between 0 and 273.16 K. *Geochimica et Cosmochimica Acta*, 71:36–45, 2007.

[166] V. A. Popav, V. G. Manzhelii, , and M. I. Bagatskii. Heat Capacity of Solid NH_3. *J. Low Temp. Phys.*, 5(4):427, 1971.

[167] L. Haar. Thermodynamic Properties of Ammonia as an ideal Gas, 1968.

[168] J. C. Li and D. K. Ross. Evidence for two kinds of hydrogen bond in ice. *Nature(London)*, 365:327, 1993.

[169] J. C. Li, S.M.Bennington, and D.K.Ross. Further evidence for the existence of two kinds of H-bonds in ice Ih. *Phys. Lett. A*, 192:295, 1994.

[170] W. A. Adeagbo, A. Zayak, and P. Entel. Ab initio study of the structure and dynamical properties of crystalline ice. *Phase Transitions: A Multinational Journal*, 78:1-3:179–196, 2005.

[171] J. Li. Inelastic neutron scattering studies of hydrogen bonding in ices. *J. Chem. Phys.* , 105(16):6733, 1996.

[172] J. S. Tse and D. D. Klug. Comments on "Further evidence for the existence of two kinds of H-bonds in ice Ih" by Li et al. *Phys. Lett. A*, 198:464, 1995.

[173] P. Zhang, L. Tian, Z. P. Zhang, G. Shao, , and J. C. Li. Investigation of the hydrogen bonding in ice Ih by first-principles density function methods. *J .Chem. Phys.*, 137:044504, 2012.

[174] R. G. Della Valle, P. F. Fracassi, R. Righini, S. Califano, and S. H. Walmsley. Phonon dispersion curves and phonon lifetimes in crystalline ammonia. *Chem. Phys.*, 44:189–196, 1979.

[175] R. Righini, N. Neto, S. Califano, and S. H. Walmsley. Lattice dynamics of crystalline ammonia and deutero-ammonia. *Chem. Phys.*, 33:345–353, 1978.

[176] M. M. Siddick and G. J. Acklandand C. A. Morrison. Constrained dynamics and extraction of normal modes from ab initio molecular dynamics: Application to ammonia. *J. Chem. Phys.*, 125:064707, 2006.

[177] O. S. Brinbrek and A. Anderson. Raman spectra of molecular crystals. Ammonia and 3-deutero-ammonia. *chem. Phys. Lett.*, 15(3):421, 1972.

[178] L. L. Shipman, A. W. Burgess, and H. A. Scheraga. Lattice Energies and Heats of Sublimation at 0 K for n-Pentane, n-Hexane, n-Octane, and Ammonia. *J. Phys. Chem.*, 80:52, 1976.

[179] * P. G. Radaelli C. G. Salzmann, 1, E. Mayer, and J. L. Finney. Ice XV: A new thermodynamically stable phase of ice. *Phys. Rev. Lett.*, 103:105701, 2009.

[180] C. Knight and S. J. Singer. Prediction of a Phase Transition to a Hydrogen Bond Ordered Form of Ice VI . *J. Phys. Chem. B* , 109:2104021046, 2005.

[181] J.-L. Kuo and W. F. Kuhs. A First Principles Study on the Structure of Ice-VI: Static Distortion, Molecular Geometry and Proton Ordering. *J. Phys. Chem. B*, 110:36973703, 2006.

[182] C. G. Salzmann, P. G. Radaelli, B. Slaterc, and J. L. Finney. The polymorphism of ice: five unresolved questions. *Phys. Chem. Chem. Phys.*, 13:1846818480, 2011.

[183] S. F. Boys and F. Bernardi. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.*, 19:553–566, 1970.

[184] A. J. Sadlej. Medium-size polarized basis sets for high-level correlated calculations of molecular electric properties. *Collect. Czech. Chem. Commun.*, 53:1995–2016, 1988.

[185] A. J. Sadlej. Medium-size polarized basis sets for high-level-correlated calculations of molecular electric properties. *Theor. Chim. Acta*, 79:123–140, 1991.

[186] J. F. Nye. Physical properties of crystals, their representation by tensors and matrices, Oxford, Clarendon Press, 1957.

[187] K. Doll. Analytical stress tensor and pressure calculations with the CRYSTAL code . *Mol. Phys.*, 108:223–227, 2010.

[188] A. bond, R. Boese, and G. R. Desiraju. On the polymorphism of aspirin: crystalline aspirin as intergrowths of two 'polymorphic' domains. *Angew. Chem. and Int. Ed.*, 46:618–22, 2007.

[189] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B*, 114:2549, 2010.

[190] T. H. Dunning. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, 90:1007–1023, 1989.

[191] B. I. Dunlap. Fitting the Coulomb potential variationally in $X\alpha$ molecular calculations. *J. Chem. Phys.*, 78:3140–3142, 1983.

[192] M. W. Feyereisen, G. Fitzgerald, and A. Komornicki. Use of approximate integrals in *ab initio* theory. An application in MP2 energy calculations. *Chem. Phys. Lett.*, 208:359–363, 1993.

[193] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs. Auxiliary basis sets to approximate Coulomb potentials. *Chem. Phys. Lett.*, 240:283–289, 1995.

[194] C. Pisani, L. Maschio, S. Casassa, M. Halo, M. Schütz, and D. Usvyat. Periodic local MP2 method for the study of electronic correlation in crystals: Theory and preliminary applications. *J. Comput. Chem.*, 29:2113–2124, 2008.

[195] K. Hongo, M. A. Watson, R. S. Sanchez-Carrera, T. Iitaka, and A. Aspuru-Guzik. Failure of conventional density functionals for the prediction of molecular crystal polymorphism: A quantum Monte Carlo study. *J. Phys. Chem. Lett.*, 1:1789–1794, 2010.

[196] Dan C. Sorescu and Betsy M Rice. Theoretical Predictions of Energetic Molecular Crystals at Ambient and Hydrostatic Compression Conditions Using Dispersion Corrections to Conventional Density Functionals (DFT-D). *J. Phys. Chem. C*, 114(14):6734–6748, April 2010.