Title

Predicting Structures and Thermodynamic Properties for Molecular Crystals at Finite Temperatures

Author

Heit, Yonaton N.

Publication Date

2016

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Predicting Structures and Thermodynamic Properties for Molecular Crystals at Finite Temperatures

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

Doctor of Philosophy

in

Chemistry

by

Yonaton N. Heit

March 2016

Dissertation Committee:

Dr. Gregory Beran, Chairperson
Dr. Chia-En Chang
Dr. Yadong Yin

The Dissertation of Yonaton N. Heit is approved by:

Committee Chairperson

University of California, Riverside

Acknowledgments

I would like to express my gratitude to my current and previous group members Dr. Shuhao Wen, Dr. Josh Hartman, Dr. Kelly Theel, Dr. Kaushik Nanda, Dr. Yuanhang Huang, Dr. Yin Liu, Josiah Jackson, Jessica McKinley, Dominique Nocito, and Watit Sontising for all their help and support. I enjoyed working with all of you and I wish you all the best in the future. I would also like to express my gratitude to my P.I. Dr. Gregory Beran for guidance throughout my graduate career. Thank you for your patience and understanding while I struggled with certain aspects of my projects. Without your help, none of this work would be possible.

In addition, I would like to express my gratitude National Science Foundation for funding Dr. Gregory Beran's and my research and therefore making my employment possible while I was at UCR. I would like to express my gratitude Extreme Science and Engineering Discovery Environment for allocating supercomputer time. Without this allocation, the calculations in chapters 4 and 5 have been possible.

Further, I would like to express my gratitude to Dr. Ali Sebetci, Dr. Shuhao Wen, Dr. Kaushik Nanda and Dr. Gregory Beran for writing of the underlining computer code which I modified to exploit space group symmetry and implement the quasi-harmonic approximation. Thank you Dr. Kaushik Nanda for helping me to understand the code.

Finally, I would like to thank my parents Gordon and Meryl Heit, my brother Tzvi Heit, and my partner Justin Kenward for all their help and support while I engaged in my research and worked on this thesis.

I would like to dedicate this dissertation to my parents, Gordon and Meryl Heit for supporting me through both college and graduate school.

ABSTRACT OF THE DISSERTATION

Predicting Structures and Thermodynamic Properties for Molecular Crystals at Finite Temperatures

by

Yonaton N. Heit

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, March 2016
Dr. Gregory Beran, Chairperson

Molecular crystals occur in a variety of chemically relevant problems, including pharmaceuticals and organic semi-conductor materials. There has been much interest in developing computational models which can predict crystal structures and properties accurately and with reasonable computational expense. One such model, hybrid many-body interaction (HMBI), fragments a chemical system into monomer, dimer, and many-body interactions, each of which may be handled using a different level of theory. HMBI has been used to predict crystal structures, lattice energies, and relative polymorph stability, particularly in cases where other methods such as periodic density functional theory (DFT) have struggled.

This dissertation extends the HMBI model in two important ways. First, the computational cost of these calculations is significantly reduced by the development and implementation of an algorithm to exploit space group symmetry. This algorithm reduces the number of monomer and dimers calculations that need to be performed by eliminating symmetrically equivalent ones. Exploitation of space group symmetry provides additional computational savings during a crystal geometry optimization by reducing the number of degrees of freedom that need to be optimized, which tends to decrease the number of optimization steps required to reach convergence.

Second, the ability to predict molecular crystal structures and properties at finite temperature is developed by coupling the HMBI model with the quasi-harmonic approximation. Traditional approaches either neglect temperature or approximate it with a harmonic vibrational model. However, molecular crystals expand appreciably with temperature and this expansion has significant impacts on crystal properties. Typically, as crystals expand, the lattice energy weakens and the phonon modes soften. Neglecting this expansion causes thermochemical properties such as enthalpy and entropy to be overestimated near room temperature. The quasi-harmonic HMBI model is demonstrated to predict temperature-dependent molar volumes, thermochemistry, and mechanical properties in excellent agreement with experiment for several small-molecule crystals—carbon dioxide, ice, acetic acid, and imidazole. These developments also pave the way toward computational prediction of molecular crystal phase diagrams as a function of temperature and pressure. Preliminary results examining the high-pressure phase diagram of carbon dioxide are presented.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Many molecules exhibit multiple distinct crystal packing motifs or polymorphs. Small changes in crystal packing can have profound and disastrous effects. For example the appearance of a previously unknown and insoluble polymorph of the anti-HIV drug, Ritonavir, caused havoc for its maker, Abbot Laboratories. The drug had to be withdrawn from the market and redesigned costing an estimated $250 million and preventing patients from receiving treatment.[1–3] Another example where a small change leads to large consequences is addition of a t-butyl group to rubrene. Rubrene normally exhibits extremely high charge carrier mobility. However, its t-butyl derivative adopts a different crystal packing motif which has no measurable carrier mobility.[4]

Predicting the structure and properties of molecular crystals is challenging, but it has applications ranging from pharmaceutical drugs to organic semiconductor materials. It can be difficult to explore all possible crystal structures and properties experimentally. Computational methods provide an excellent tool for predicting crystal structures and properties to supplement experimental work. Current applications of computational chemistry include predicting crystal structures,[5–13] determining the relative stability of polymorphs,[14–19] predicting thermochemistry,[14,20,21] interpreting vibrational spectroscopy experiments,[17,18,22] and assigning NMR chemical shifts.[23–25]

Evaluating the energetics reliably in order to rank the predicted structures properly is one of the key challenges in crystal structure prediction. Even when the polymorphs for a crystal are known, their relative stabilities can be difficult to determine. Glycine,[26] acetaminophen,[27–29] and oxalyl dihydrazide[16,19] are a few crystals that have contradictory polymorphic stability ordering from different studies and methods. A 2015 survey by Nyman and Day[30] found that of 508 polymorphic species with 1061 crystals found that over half of polymorphic pairs only differed by less about 2 kJ/mol and for over 95% by less than 7.2 kJ/mol. In order to distinguish these crystal, sub-kJ/mol accuracy or close to it is often required. On the other hand, highly accurate models tend to be computationally expensive, and one must balance between computational cost and accuracy.

Periodic density functional theory (DFT) is a widely used method[19,20,26–29,31–43] due to its computational affordability. Unfortunately, despite many successful DFT predictions, one often finds that different density functionals and dispersion corrections predict contradictory polymorph stability orderings.[16,26,39] It is not always obvious which functional is optimal for a given crystal. There is also no obvious means for systemically improving DFT, unlike wave function-based methods.

One of the simpler and useful wave function method, second-order Möller-Plesset perturbation theory (MP2), is computationally expensive for periodic crystals. There are various classical force field methods which are computationally cheap but not as accurate as the methods based on electronic structure theory. In response to this problem, our group has developed a fragment based QM/MM method, hybrid many-body interactions or HMBI to model molecular crystals. HMBI has had many successes such as predicting relative stability of polymorphs[16] and benchmarking the lattice energy of crystals,[44,45] which will be discussed in Section 1.8.

Not only does one want to predict the relative lattice energy stability of polymorphs, but one wants to predict the temperature and pressure under which a given phase will be stable. Mapping out phase diagrams for polymorphic crystals experimentally can be challenging because polymorph transformations can be sluggish[46,47] and/or path dependent,[48] meaning that a kinetically viable pathway to the thermodynamically stable phase may not be available. This is not an issue for theoretical studies, since structure predictions do not require knowledge of kinetic path taken to arrive at the crystals structure. Many computational studies have explored the pressure and temperature dependence of phase stability.[18,36,49–51] Pressure dependence can be modeled fairly trivially, while temperature dependence is more difficult because it involves contributions from phonon modes. Using the harmonic approximation, vibrational frequencies can be determined computationally. Crystals do expand with temperature and this expansion can be captured with the quasi-harmonic approximation by approximating how vibrational frequency changes with volume.[31,49,52,53] This dissertation will discuss the implementation of the quasi-harmonic approximation into the HMBI model in order to improve the prediction of temperature-dependent properties. The influence of thermal expansion of various thermodynamical properties will be examined. In addition, to make these quasi-harmonic studies feasible, an algorithm for exploiting space group symmetry to accelerate the calculations will be explored.

1.1 Born-Oppenheimer Approximation

The center of all quantum mechanical calculations is the Schrödinger Equation. Solving the Schrödinger Equation defines wave functions (Ψ) which are used to characterize all measurable properties of a system. The time independent Schrödinger Equation is more frequently used for chemical systems and describes a quantum system in a static state.

$$\mathbf{H}\Psi = E\Psi \tag{1.1}$$

The Hamiltonian ($\mathbf{H}$) for a chemical system is generally defined as

$$\mathbf{H} = -\sum_{i} \frac{1}{2}\nabla_i^2 - \sum_{k} \frac{1}{2m_k}\nabla_k^2 + \sum_{i}\sum_{k} \frac{Z_k}{r_{ik}} - \sum_{i<j} \frac{1}{r_{ij}} + \sum_{k<l} \frac{Z_k Z_l}{r_{kl}} \tag{1.2}$$

the i and j are indices for electrons and the k and l are the indices for the nuclei. The Laplacian Operator (∇^2) defined as

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \tag{1.3}$$

and when multiplied by $\frac{1}{2}$ or $\frac{1}{2m_k}$ is the kinetic energy operator which defines the kinetic portion of the energy. The Hamiltonian in Eq 1.2 characterizes the wave function for the electrons and nuclei. Due to the attraction and repulsion terms ($\frac{Z_k}{r_{ik}}$, $\frac{1}{r_{ij}}$, $\frac{Z_k}{r_{ik}}$) in the Hamiltonian, the motions of all particles are correlated. With the exception of the hydrogen-like system (one nucleus, one electron), solutions to the Schrödinger Equation are a many-body problem for which an analytical solution does not exist. They are various approximate methods to solve the Schrödinger Equation (see Section 1.5). Most of these methods simplify the Hamiltonian using the Born-Oppenheimer approximation. This approximation factorizes the wave function into a product of electronic and nuclear wave functions.

$$\Psi = \Psi_{el}\Psi_{nu} \tag{1.4}$$

Since electrons are lighter and faster than nuclei, the Born-Oppenheimer approximation also assumes that electrons relax in response to the position of fixed nuclei. The nuclear kinetic operator ($-\frac{1}{2m_k}\nabla_k^2$) disappears since the nuclei is fixed and the nuclei repulsion ($\frac{Z_k}{r_{ik}}$) is treated as a constant to form to electronic Hamililtonian.

$$\mathbf{H}_{el} = -\sum_i \frac{1}{2}\nabla_i^2 + \sum_i \sum_k \frac{Z_k}{r_{ik}} + \sum_{k<l} \frac{Z_k Z_l}{r_{kl}} - \sum_{i<j} \frac{1}{r_{ij}} + \sum_{k<l} \frac{Z_k Z_l}{r_{kl}} \tag{1.5}$$

The electronic Hamiltonian applied only to the electronic wave function defines the electronic energy.

$$\mathbf{H}_{el}\Psi_{el} = E_{el}\Psi_{el} \tag{1.6}$$

Since the position of the nuclei are fixed, the electronic energy is a function of the nuclei position. This function is called the potential energy surface (PES). The PES is a very powerful tool for computational chemistry. Using this tool, atomic structure can be determined. The nuclear geometry of a stable species is determined by minimizing the electronic energy. Transition states which determine the pathways for chemical and physical changes occur as saddle points in the PES. Bond length, bond angles and all position related parameters are all defined by these stationary points on the PES.

The nuclear Hamiltonian includes what was left out of the electronic Hamiltonian, the nuclear kinetic operator.

$$\mathbf{H}_{nu} = -\sum_{k} \frac{\hbar^2}{2m_k}\nabla_k^2 + \mathbf{H}_{el} \tag{1.7}$$

Because the nuclear and electron wave functions are separated, when this operator is applied to the complete wave function (nuclear and electron) the electronic Hamiltonian will operate only on the electronic wave function. This gives the electronic energy. Since the nuclear kinetic operator determines the kinetic energy of the nuclei, this means nuclear potential energy is equal to the only other energy term, the electronic energy.

$$\left(-\sum_{k} \frac{1}{2m_k}\nabla_k^2 + \mathbf{H}_{el}\right)\Psi_{el}\Psi_{nu} = \left(-\sum_{k} \frac{1}{2m_k}\nabla_k^2 + V(\mathbf{R})\right)\Psi_{el}\Psi_{nu} \tag{1.8}$$

$$\mathbf{H}_{nu} = -\sum_{k} \frac{\hbar^2}{2m_k}\nabla_k^2 + V(\mathbf{R}) \tag{1.9}$$

The fact that using the Born-Oppenheimer Approximation, the electronic energy and the nuclear potential are the same has implications for the next section. For more information of the Born-Oppenheimer Approximation see Refs 54 and 55.

1.2 Harmonic Approximation

In the Born-Oppenheimer approximation, the electronic energy omits nuclear vibrational motion due to neglecting the nuclei kinetic energy. The vibrational energy can be captured by solving the nuclear Schrödinger equation (Eq 1.9). As established in the last section, the potential $V(\mathbf{R})$ is equal to the electronic energy at $\mathbf{R}$. Solving the nuclear Schrödinger Equation proves to be difficult because the nuclear potential is generally only known at a given set of nuclear coordinates and not at all $\mathbf{R}$. Determining the potential at all $\mathbf{R}$ requires full characterization of the PES which is computationally unfeasible for complex systems. Instead, the potential is expanded by a Taylor Series.

$$V(\mathbf{R}) = V(\mathbf{R_0}) + (\mathbf{R} - \mathbf{R}_0)\frac{\partial V}{\partial R_0} + \frac{1}{2}(\mathbf{R} - \mathbf{R}_0)^T \frac{\partial^2 V}{\partial \mathbf{R}_0^2}(\mathbf{R} - \mathbf{R}_0) + \ldots \tag{1.10}$$

The potential at a equilibrium reference state $V(\mathbf{R}_0)$ is defined to zero. If the reference state is a stationary point such as a minimum or saddle point, the potential first derivative ($\frac{\partial V}{\partial R_0}$) is also zero. All terms beyond the second order are assumed to be negligible. Therefore the potential nuclear Hamiltonian can be replaced with the second order term of the Taylor expansion.

$$\mathbf{H}_{nu} = -\sum_k \frac{1}{2m_k}\nabla_k^2 + \frac{1}{2}(\mathbf{R} - \mathbf{R}_0)^T \frac{\partial^2 V}{\partial \mathbf{R}_0^2}(\mathbf{R} - \mathbf{R}_0) \tag{1.11}$$

The nuclear coordinates are then redefined as mass-dependent coordinates

$$Y_\alpha = \sqrt{m_\alpha}(R_\alpha - R_{0,\alpha}) \tag{1.12}$$

$$V_{\alpha,\beta} = \frac{1}{\sqrt{m_\alpha m_\beta}} \frac{\partial^2 V}{\partial R_{0,\alpha} \partial R_{0,\beta}} \tag{1.13}$$

$$\mathbf{H}_{nu} = -\sum_k \frac{\hbar^2}{2m_k} \nabla_k^2 + \frac{1}{2} \mathbf{Y}^T \mathbf{V} \mathbf{Y} \tag{1.14}$$

Since the mass-weighed Hessian matrix ($\mathbf{V}$) is symmetric

$$V_{\alpha,\beta} = V_{\beta,\alpha} \tag{1.15}$$

it is diagonalizable. There exists some unitary matrix $\mathbf{U}$, transforming $\mathbf{Y}$ into the some basis of $\mathbf{q}$

$$\mathbf{Y} = \mathbf{U}\mathbf{q} \tag{1.16}$$

which diagonalizes the weighed Hessian matrix ($\mathbf{V}$) by replacing $\mathbf{Y}$ with the right side of Eq 1.16

$$\begin{aligned} \mathbf{H}_{nu} &= -\sum_k \frac{\hbar^2}{2m_k} \nabla_k^2 + \frac{1}{2} \mathbf{q}^T \mathbf{U}^T \mathbf{V} \mathbf{U} \mathbf{q} && (1.17) \\ &= -\sum_k \frac{\hbar^2}{2m_k} \nabla_k^2 + \frac{1}{2} \sum_\alpha \epsilon_\alpha q_\alpha^2 && (1.18) \end{aligned}$$

where α are the degrees of freedom of the k nuclei. This is a powerful transformation. The nuclear potential is treated as a series of uncoupled one-dimensional harmonic oscillators

$$V(\mathbf{R}) = \frac{1}{2}\sum_{\alpha} \epsilon_{\alpha} q_{\alpha}^{2} \tag{1.19}$$

The mass-weighted force constant ϵ_{α} is used to determine the vibrational frequencies.

$$f_{\alpha} = \frac{1}{2\pi}\sqrt{\epsilon_{\alpha}} \tag{1.20}$$

These are the vibrational frequencies of the system vibrating about references geometry $\mathbf{R}_0$ which the Taylor Series in Eq 1.10 was expanded around. Since the first derivative in the Taylor expansion is assumed to be zero, the frequencies are only defined when the reference state is at a critical point such as a transition state or minimum. The mass weighted force constant (ϵ_{α}) and the eigenvectors which are the columns of $\mathbf{U}$ are determined by diagonalization of the the mass-weighed Hessian ($\mathbf{V}$).

The frequencies introduce a temperature dependence into the energy. Temperature is ill-defined in a quantum calculation and comes from statistical thermodynamics. In thermodynamics, temperature is defined by the average kinetic energy of a macroscopic system. Partition functions link the states allowed by quantum calculations to the bulk properties in thermodynamics. The vibrational energy determined by the standard vibrational partition function is.

$$E_{vib} = N\sum_{\alpha}\left(\frac{hf_{\alpha}}{2} + \frac{hf_{\alpha}}{exp\left(\frac{hf_{\alpha}}{kT}\right) - 1}\right) \tag{1.21}$$

Additional nuclear energy terms such as the translational and the rotational energies are also determined using standard partition functions. This terms are significant for gas phase molecules but condensed (liquid and solid) phase have no overall translation or rotational energy.

In addition to allowing one to determine vibrational energy, vibrational frequencies allows one to compute the entropy using the same standard partition functions.

$$S_{vib} = N\sum_{\alpha}\left(\frac{hf_{\alpha}}{T\left[exp\left(\frac{hf_{\alpha}}{kT}\right)-1\right]} - kln\left[1-exp\left(\frac{hf_{\alpha}}{kT}\right)\right]\right) \tag{1.22}$$

For gas phase molecules, an additional rotational and translational term in necessary to determine the total entropy. For disordered crystals, such as Ih ice, configurational entropy should also be included. If the system in question is the solid phase (no rotation or translation) and is ordered (no configurational entropy), the Gibbs free energy can be determined by

$$\begin{aligned} G &= E_{el} + PV + E_{vib} - TS_{vib} \\ &= E_{el} + PV + F_{vib} \end{aligned} \tag{1.23}$$

The Gibbs free energy is significant since it determines the stability of a phase or structure. The Helmholtz vibrational energy F_{vib} replaced the vibrational entropy and energy because it simplifies the vibrational terms.

$$\begin{aligned} F_{vib} &= E_{vib} - TS_{vib} \\ &= N\sum_{\alpha}\left(\frac{hf_{\alpha}}{2} + kTln\left[1-exp\left(\frac{hf_{\alpha}}{kT}\right)\right]\right) \end{aligned} \tag{1.24}$$

While the equations in this section were outlined for a quantum mechanical potential energy, molecular mechanics potentials such as those given in Section 1.6 may be used instead. For more information the harmonic approximation see Ref 55.

1.3 Unit Cell Parameters

Crystals contain a enormous number of atoms or ions packed close together. If these crystals are ordered, they can be represented by a series of identical unit cells. The unit cell gives the positions of a small set of atoms in a position of $\mathbf{R}$ within a parallelepiped. The unit cell's dimensions are defined by three lattice vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Using the periodic boundary conditions (PBC), the position of atoms in other unit cells can be determined by the translation vector $\mathbf{x}(\kappa)$

$$\mathbf{R}(\kappa) = \mathbf{R}(0) + \mathbf{x}(\kappa) \tag{1.25}$$

$$\mathbf{x}(\kappa) = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 \tag{1.26}$$

where a_1, a_2, and a_3 are integer indices of a unit cell. The cell where a_1,a_2,a_3 equal zero is the central unit cell ($\kappa = 0$). It is often more convenient to determine the dimensions of the unit cell using the six unit cell parameters (a, b, c, α, β, γ) rather than the three lattice vectors. The lattice lengths a, b, and c are the lengths of the sides of the unit cell or the lengths of three lattice vectors. The lattice angles α, β, and γ are the angles between the sides of the unit cell or the angle between the lengths of three lattice vectors. The relation between the lattice vectors and lattice parameters are:

$$\mathbf{v}_1 = \begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{v}_2 = \begin{pmatrix} b\cos\gamma \\ b\sin\gamma \\ 0 \end{pmatrix}$$

$$\mathbf{v}_3 = \begin{pmatrix} c\cos\beta \\ c\left(\frac{\cos\alpha - \cos\beta\cos\gamma}{\sin\gamma}\right) \\ c\left(\frac{\sqrt{\sin\alpha^2 - \cos\beta^2 - \cos\gamma^2 + 2\cos\alpha\cos\beta\cos\gamma}}{\sin\gamma}\right) \end{pmatrix} \tag{1.27}$$

Using the lattice parameters restricts the overall lattice vectors so that $\mathbf{v}_1$ must be along the global x-axis and $\mathbf{v}_2$ must be in the xy plane.

1.4 Lattice Dynamics

1.4.1 Phonon Dispersion

Since a unit cell is a complete representation of the crystal, one could naively determine the harmonic vibrational frequencies as described by Section 1.2, building a mass-weighed Hessian from all the nuclear positions. This mass-weighed Hessian will provide the vibrational frequencies of the atoms in the central unit cell vibrating in phase with all periodic image atoms in adjacent cells. To capture the out-of-phase vibrations between cells, a dynamical Hessian matrix is used,

$$\mathbf{D}_{\alpha,\beta}(l,l',\mathbf{k}) = \frac{1}{\sqrt{M_l M_{l'}}} \sum_{\kappa} \frac{\partial V}{\partial R_\alpha(0) \partial R_\beta(\kappa)} exp\left(-2\pi i \mathbf{k} \cdot \delta \mathbf{R}_{l,l'}(0,\kappa)\right) \tag{1.28}$$

where $\delta R_{k,l}(0,\kappa)$ is the distance vector between atom l in the central unit cell and atom l' in unit cell κ. The **k**-point vector **k** defines the propagation of phonons. Several sets of dynamical matrices are set at **k**-points each with different vibrational frequencies. The **k**-points are defined in reciprocal space or momentum space. The **k**-points in all subsequent chapters are selected using a Monkhorst-Pack grid.[56] The Monkhorst-Pack grid creates a set of **k**-points so that they are parallel with the reciprocal lattice vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$.

$$\mathbf{k} = k_1 \mathbf{b}_1 + k_2 \mathbf{b}_2 + k_3 \mathbf{b}_3 \tag{1.29}$$

where k_1, k_2, and k_3 are integers. The reciprocal lattice vectors define the dimensions of the Brillouin zone similar to how the lattice vectors define the dimensions of unit cell in direct space. The reciprocal lattice vectors can be determined from the lattice vectors in direct space by

$$\mathbf{b}_1 = \frac{2\pi\left(\mathbf{v}_2 \times \mathbf{v}_3\right)}{|\mathbf{v}_1 \cdot \mathbf{v}_2 \times \mathbf{v}_3|} \tag{1.30}$$

$$\mathbf{b}_2 = \frac{2\pi\left(\mathbf{v}_1 \times \mathbf{v}_3\right)}{|\mathbf{v}_1 \cdot \mathbf{v}_2 \times \mathbf{v}_3|} \tag{1.31}$$

$$\mathbf{b}_3 = \frac{2\pi\left(\mathbf{v}_1 \times \mathbf{v}_2\right)}{|\mathbf{v}_1 \cdot \mathbf{v}_2 \times \mathbf{v}_3|} \tag{1.32}$$

When the k_1, k_2, and k_3 all equal zero, the dynamical matrix reduces to the conventional mass-weighted Hessian for the central unit cell, diagonalization of which produces that standard in-phase γ-point phonon modes.

1.4.2 Quasi-Harmonic Approximation

It is well-known that crystals generally expand with increasing temperature. The volume of phase I carbon dioxide increases by nearly 10% (from 25.8 to 27.9 cm^3) as the temperature is increased from 15 to 190 K.[57] On the other extreme, Ih ice expands by about 2% as the temperature increases from 10 K to 265 K.[58] Ice Ih also presents an unusual case of negative thermal expansion. The volume decreases slightly as the temperature increases from 10 K to 70 K.

In the harmonic approximation, atoms vibrate around some equilibrium position. Increasing temperature shifts the phonon populations toward higher-energy states, but the equilibrium position does not change. Since there is no change in the equilibrium position when temperature increases, the harmonic approximation fails to capture the thermal expansion of crystals. One solution to this issue is the quasi-harmonic approximation. This approximation allows the crystal unit cell to expand (or compress). In doing so, the atom positions will relax and alter the harmonic phonon frequencies. The quasi-harmonic approximation estimates these changes in the phonon frequencies due to changes in the cell volume via Grünsensein parameters (γ):

$$\gamma_i = -\frac{\partial ln\left(f_i\right)}{\partial ln\left(V\right)} \tag{1.33}$$

Integrating Eq 1.33 allows for the determination of vibrational frequencie at a particular cell volume based on the frequencies.

$$f_i = f_{i,ref} \left(\frac{V}{V_{ref}} \right)^{-\gamma_i} \tag{1.34}$$

This equation is extremely powerful. As long as the frequencies at a reference volume are known, the frequencies at any given volume can be determined. Interestingly, this is without knowing positions of the atomic coordinates. The next question is the choice for the Grünsensein parameters. Is it a constant or a function of volume? Otero-de-la-Roza and Johnson performed linear, parabolic, and quartic fitting of the phase I carbon dioxide Grünsensein parameters.[31] They found that simpliest of the fitting, the linear fitting, reproduced the Helmholtz vibrational energy (see Eq 1.24) of the most complex fit, the quartic fitting, for about 200 cubic Bohrs around the reference volume. Since this is a longer range than would be expected from thermal expansion, a linear fit with constant Grünsensein parameters seem to be perfectly valid. The frequencies at only two volumes are necessary to fit Grünsensein parameters numerical.

$$\gamma_i = -\frac{\partial ln\left(f_i\right)}{\partial ln\left(V\right)} \approx -\frac{ln\left(f_{i,2}\right) - ln\left(f_{i,1}\right)}{ln\left(V_2\right) - ln\left(V_1\right)} \tag{1.35}$$

With well-defined Grünsensein parameters and frequencies at a reference volume, the Helmholtz vibrational energy is defined at any cell volume allowing for the Gibb's free energy to be minimized as a function of pressure and volume. With this one can obtain thermal expansion.

For more information on the quasi-harmonic approximation see Ref 59.

1.5 Electronic Structure Methods

Electronic structure methods solve the Schrödinger equation for the energy *ab initio* or "from the beginning". This means that values for the system are found without any *a priori* assumptions about strength or length of a bond, what angle between atoms should be, or any other empirical or experimental values. All that is assumed about the system are the position of the nuclei and the number of electrons. This contrasts force field methods which will be discussed in Section 1.6. The electronic energy and wave function of the system is determined using the Schrödinger Equation simplified by Born-Oppenheimer approximation as seen in Section 1.1 (see Eq 1.6). Due to the complexities of the Schrödinger Equation, electron structure methods start with the Hartree Fock (HF) method then include a series of corrections

1.5.1 Hartree Fock Method

In Hartree Fock (sometimes called the self-consistent-field method), each electron is treated as experiencing a mean field of the other electrons. The wave function of these N electrons in this mean field is a Slater Determinant.

$$\Psi_{el} = \frac{1}{\sqrt{N}} \begin{vmatrix} \Phi_1(1) & \Phi_2(1) & \Phi_3(1) & \cdots & \Phi_N(1) \\ \Phi_1(2) & \Phi_2(2) & \Phi_3(2 & \cdots & \Phi_N(2) \\ \Phi_1(3) & \Phi_2(3) & \Phi_3(3) & \cdots & \Phi_N(3) \\ \vdots & \vdots & \vdots & & \vdots \\ \Phi_1(N) & \Phi_2(N) & \Phi_3(N) & \cdots & \Phi_N(N) \end{vmatrix} \tag{1.36}$$

Each of the orbitals Φ_i are molecular orbitals. Only the ground state occupied orbitals are included. Besides simplifying the wave function, Slater Determinants also has the advantage that it observes the Pauli Exclusion Principle which require the wave function of fermions, such as electrons, to be antisymmetrical. That is the wave function changes sign when two electrons to switch orbitals.

$$\Psi_{el}(\Phi_1(1)\Phi_2(2)) = -\Psi_{el}(\Phi_1(2)\Phi_2(1)) \tag{1.37}$$

The Hartree-Fock ground state energy is determined using the variational principle,

$$E_0 = \frac{\int \Psi^*(r)\mathbf{H}_{el}\Psi(r)dr}{\int \Psi^*(r)\Psi(r)dr} \tag{1.38}$$

According to the variational principle, the energies produced from an approximate wave function will always be equal to or greater than the exact electronic energy.

$$E_{el} \leq E_0 \tag{1.39}$$

Initial guesses for the wave function of each orbital are chosen and through an iterative process, the orbitals are optimized to minimize the ground state energy. In general, the larger the number of orbitals or the basis set used, the better the results (see Section 1.5.4) A more complete review of Hartree-Fock can be found in Ref 60

1.5.2 Post-Hartree-Fock Methods

Hartree-Fock itself does not provide very useful predictions but it offers a good starting place for many more accurate methods. There are several post-Hartree-Fock methods that determine the correlation energy, the energy missed by Hartree-Fock.

$$E = E_{HF} + E_{corr} \tag{1.40}$$

One popular post Hartree-Fock method is second order Møller-Plesset Perturbation Theory or MP2. MP2 uses second-order perturbation theory to correct Hartree-Fock by adding the missing correlation energy based on double excitations of electrons from occupied ground state orbitals in the Stater Determinant to unoccupied or virtial orbitals.

$$\begin{aligned} E_{corr,MP2} = \quad & \sum_{j<i} \sum_{a<b} \left(\frac{\int \Phi_i^*(r_1)\Phi_j^*(r_2)\frac{1}{r_{1,2}}\Phi_a(r_1)\Phi_b(r_2)dr_1dr_2}{e_i + e_j - e_a - e_b} \right. \\ & \left. - \frac{\int \Phi_i^*(r_1)\Phi_j^*(r_2)\frac{1}{r_{1,2}}\Phi_b(r_1)\Phi_a(r_2)dr_1dr_2}{e_i + e_j - e_a - e_b} \right)^2 \end{aligned} \tag{1.41}$$

The i and j are the occupied orbitals in Stater Determinant and a and b are the virtual orbitals and the e's are the Hartree-Fock orbital energies. MP2 does have steeper computation scaling with system size than HF, scaling as N^5 versus N^4 respectfully, where N is the number of orbitals or size of the basis set (see Section 1.5.4). For most small systems, MP2 is affordable. The main issue with MP2 is that it overestimates the dispersion forces. Because of this, MP2 overbinds systems that have large dispersion such as pi-stacked benzene.[61,62] This deficiency in describing dispersion interactions can be corrected by going to higher-order perturbation theory, using the dispersion-corrected MP2C model,[63–65] or switching to coupled cluster theory.

Coupled cluster theory modifies the wave function through excitations of orbitals in the Slater Determent from ground state to viritual orbitals using the excitation operator **T**. Typically this operator is truncated to include only single and double excitations, in which case the model is called coupled cluster singles and doubles (CCSD).

$$
\begin{aligned}
\mathbf{T} &= \mathbf{T}_1 + \mathbf{T}_2 + \mathbf{T}_3 + \ldots + \mathbf{T}_n && (1.42) \\
&\approx \mathbf{T}_1 + \mathbf{T}_2 \\
\mathbf{T}_1\Psi_0 &= \sum_i \sum_a t_i^a \Psi_i^a && (1.43) \\
\mathbf{T}_2\Psi_0 &= \sum_{i<j} \sum_{a<b} t_{i,j}^{a,b} \Psi_{i,j}^{a,b} && (1.44)
\end{aligned}
$$

where Ψ_i^a is the Slater determinant of orbital i replaced by virtual orbital a, $\Psi_{i,j}^{a,b}$ is the Slater determine of orbital i and j replaced by virtual orbital a and b. The coupled cluster wave function uses the truncated excitation operator in exponential form to modify the Hartree Fock wave function. This exponential operator is expanded using a Taylor series

$$\begin{aligned}\Psi_{CCSD} &= e^{(\mathbf{T}_1+\mathbf{T}_2)}\Psi_0 \qquad (1.45)\\ &= \left(1+\mathbf{T}_1+(\mathbf{T}_2+\frac{1}{2}\mathbf{T}_1^2)+\ldots\right)\Psi_0\end{aligned}$$

The CCSD energy equations are solved projectively. Correlation beyond double excitations can be important, but full iterative inclusion of the triples is computationally impractical for all but very simple systems. Instead, the triples contribution is estimated perturbatively to create CCSD(T). CCSD(T) corrects the dispersion over binding of MP2[61,62] though the computational cost is high compared to MP2 (N^7 vs. N^5). Large-basis CCSD(T) results are often estimated by calculating the difference between MP2 and CCSD(T) in a small basis set then adding it to the MP2 energies of a larger basis set.

$$\Delta^{CCSD(T)} = E^{small}_{CCSD(T)} - E^{small}_{MP2} \qquad (1.46)$$

$$E^{large}_{CCSD(T)} \approx E^{large}_{MP2} + \Delta^{CCSD(T)} \qquad (1.47)$$

Despite its simplicity, this method of appoximating large basis set CCSD(T) has been found to be fairly reliable.[61,62]

1.5.3 Density Functional Theory

Density functional theory (DFT) provides a popular and computationally affordable alternative to wave function methods, since it provides reasonable accuracy at Hartee-

Fock-like cost. Instead of solving the Schrödinger equation to find the many-dimensional wave function, DFT computes the energy as a functional of the three-dimensional electron density (ρ). The challenge is that the exact form of this energy functional is unknown.

In standard Kohn-Sham DFT, the functional is written as a sum of a kinetic energy piece for a non-interacting system of electrons ($T_s[\rho]$), classical nuclear-electron ($E_{ne}[\rho]$) and electron-electron ($J[\rho]$) Coulombic terms, and an exchange-correlation functional ($E_{xc}[\rho]$):

$$E_{DFT} = T_s[\rho] + E_{ne}[\rho] + J[\rho] + E_{xc}[\rho] \tag{1.48}$$

The exchange-correlation functional captures the missing correlation from the kinetic energy term and the differences between classical and quantum mechanical Coulombic interactions (exchange). Different density functional methods vary based on the particular functional form adopted for the exchange-correlation functional.

In the context of molecular crystals, it should be noted that conventional semi-local density functionals fail to describe dispersion interactions,[66] since dispersion is a non-local phenomenon. It is common to correct these functionals with post-hoc dispersion corrections[67] such as D2,[68] D3,[69] TS,[40] or MBD.[29,41,70]

Even with these dispersion corrections, one often finds that different density functionals predict different polymorph stability orderings, and the correct result is not always obvious. Oxalyl dihydrazide has five known polymorphs and different functionals and dispersion corrections order the stability differently.[16,19] Other examples include the pharmaceuticals B5 and DB7.[39]

1.5.4 Basis Sets

In quantum computational chemistry, basis sets are used to represent orbitals. They are the equivalent of atomic orbitals in atomic theory. Each orbital is an atom-centered an Slater-type orbital (STO) is a linear combination of Gaussian type orbital (GTO).

$$\phi_i(STO) = \sum_a c_a \phi_a(GTO) \tag{1.49}$$

The STO can represent s, p, d, etc orbitals (called s-like, p-like, d-like, etc). The molecular orbitals are linear combination of the Slater-type orbitals.

$$\Phi = \sum_i \phi_i(STO) \tag{1.50}$$

In general, larger the basis sets (more STOs) provide better results because they provide more flexibility for the MOs. Dunning basis orbitals, the cc-pVXZ (n = D,T,Q,5,6,...) and the augmented basis set aug-cc-pVXZ (that include diffuse functionals), are commonly used because they provide smooth energy convergence with increasing basis set size. The X indicates the zeta level (double, triple, quadruple, etc) which as the level increases so does the size of the basis set.

Since the energy converges with increasing basis set size, the energy can be extrapolated to a basis set of infinite size also called the complete basis set limit (CBS). Typically, the triple zeta (TZ) and qudruple zeta (QZ) basis set are used for this extrapolation. When extrapolating the MP2 to the complete basis set, the HF and the correlation energy are extrapolated separately then added together to get the CBS.

$$E_{HF}^{CBS} = E_{HF}^{QZ} + \frac{E_{HF}^{QZ} - E_{HF}^{TZ}}{exp\,(1.54(4-3))} \tag{1.51}$$

$$E_{corr}^{CBS} = \frac{4^3 E_{corr}^{QZ} - 3^3 E_{corr}^{TZ}}{4^3 - 3^3} \tag{1.52}$$

$$E_{MP2}^{CBS} = E_{HF}^{CBS} + E_{corr}^{CBS} \tag{1.53}$$

This extrapolation is generally not performed for CCSD(T). Instead CCSD(T) correction is added to the MP2/CBS energies as shown by Eq 1.46 and 1.47. Typically the CCSD(T) correction is found using cc-pVDZ or aug-cc-pVDZ.

One of the side effects of using a finite basis set is basis set superposition error (BSSE). When determining the interaction energy of two molecules, the energy of them together in a dimer is subtracted from the energy of the individual molecules.

$$E(int) = E(AB) - E(A) - E(B) \tag{1.54}$$

For atom-centered basis sets like the ones discussed here, the number of basis functions increases with the numbers of atoms. The calculation of dimer of molecule A and B have more basis functions than either molecule by itself. This leads to artificial stabilization of the dimer which is not present in the monomer. This inconstancy in the treatment of the basis set increases error in the interaction energy, the BSSE. The most common correction to the BSSE is counterpoise correction. The basis set for the monomers is increased to include the basis functions of the full dimer. These extra functionals are included through ghost atoms. Ghost atoms are atoms that do not the nuclei and electrons in the calculation but basis functions centered around the nuclei are.

$$E(int) = E(AB)_{AB} - E(A)_{AB} - E(B)_{AB} \tag{1.55}$$

1.6 Classical Force Field Methods

Section 1.1 shows how the Born-Oppenheimer approximation can determine the nuclear potential from the electronic energy determined by quantum calculations (see Section 1.5). The nuclear potential can also be determined from models based on classical molecular mechanics. Molecular mechanics force fields tend to be less accurate than quantum mechanics ones, but they are also computationally cheaper. Classical models are particularly adept at modeling longer-range interactions (about > 10 Å), for which the differences between quantum and classical models are often small. Many different flavors of MM models exist. Three models are discussed here. The first one, AMBER, is presented because it is simple and easy to understand and the other two, AMOEBA and AIFF, are used in subsequent chapters.

1.6.1 AMBER

AMBER (Assisted Model Building with Energy Refinement) is simple force field model that is commonly used to simulate systems like proteins that are too large for more computationally expensive models. Like most MM models, energy can be separated into intramolecular and intermolecular terms.

$$E = E_{intra} + E_{inter} \tag{1.56}$$

The intramolecular terms involve bond, angle, and dihedral terms that occur within a single molecule,

$$E_{intra} = E_b + E_a + E_d \tag{1.57}$$

The first two terms, the bond (E_b) and the angle (E_a) are treated as harmonic potentials between two and three atoms respectively moving about an equilibrium position $r_{0,b}$ and $\theta_{0,a}$.

$$E_b = \sum_b k_b (r_b - r_{0,b})^2 \tag{1.58}$$

$$E_a = \sum_a k_a (\theta_a - \theta_{0,a})^2 \tag{1.59}$$

The dihedral term E_d describes the potential of a dihedral angle made of a four-atoms bond chain rotating around an equilibrium angle γ_d in a Fourier series.

$$E_d = \sum_{d,n} \frac{V_n}{2} \left[1 + \cos\left(n\phi_d - \gamma_d\right)\right] \tag{1.60}$$

The intermolecular portion includes electrostatic and van der Waals terms.

$$E_{inter} = E_{elect} + E_{vdW} \tag{1.61}$$

The electrostatic term is described using a point change Coulomb potential. Nuclei are commonly used as charge centers but other change groups such as lone pairs may also be used.

$$E_{elect} = \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{i,j}} \tag{1.62}$$

The last terms is the van der Waals. The van der Waals is an overarching term for non-covalent, non-electrostatic interactions. One of the more significant van der Waals forces is dispersion. The van der Waals interactions are described using the Lennard-Jones potential.

$$E_{vwd} = \sum_{i,j} 4\epsilon_{i,j} \left[\left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^6 - \left(\frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} \right] \tag{1.63}$$

CHARMM, OPLS, and TIPnP (n = 3-5) are examples of other models that other models are similar to AMBER. For more information on AMBER is REF 71.

1.6.2 AMOEBA

AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) is another MM force field. While AMOEBA and AMBER have the same intramolecular/intermolecular energy separation as shown in Eq 1.56, the intramolecular/intermolecular breakdown differs in a number of ways. The intermolecular interactions separates the electrostatic into permanent and induced electrostatic terms.

$$E_{inter} = E_{elect}^{perm} + E_{elect}^{ind} + E_{vdW} \tag{1.64}$$

The permanent electrostatic term treats electron density using a point multipole representation up to the rank of quadrupole,

$$\mathbf{M}_{perm}^{i} = [q^i, \mu_x^i, \mu_y^i, \mu_z^i, Q_{xy}^i, Q_{xz}^i.....]^T \tag{1.65}$$

where q^i is the charge of atom i, μ^i is the dipole, and Q^i is the quadrupole. The permanent multipoles are computed from distributed multipole analysis (DMA) [72] on the quantum mechanical electron density of the isolated molecule in its equilibrium geometry. The permanent electrostatic interactions between atoms/molecules are treated as a series of pair-wise interactions of multipoles at nuclear charge sites.

$$E_{elect}^{perm} = \sum_{i<j} (\mathbf{M}^i)^T \mathbf{T}^{ij} \mathbf{M}^j \tag{1.66}$$

where the interaction matrix between atom i and j is

$$\mathbf{T}^{ij} = \begin{pmatrix} 1 & \frac{\partial}{\partial x_j} & \frac{\partial}{\partial y_j} & \frac{\partial}{\partial z_j} & \frac{\partial}{\partial x_j^2} & \cdots \\ \frac{\partial}{\partial x_i} & \frac{\partial^2}{\partial x_i \partial x_j} & \frac{\partial^2}{\partial x_i \partial y_j} & \frac{\partial^2}{\partial x_i \partial z_j} & \frac{\partial^3}{\partial x_i \partial x_j^2} & \cdots \\ \frac{\partial}{\partial y_i} & \frac{\partial^2}{\partial y_i \partial x_j} & \frac{\partial^2}{\partial y_i \partial y_j} & \frac{\partial^2}{\partial y_i \partial z_j} & \frac{\partial^3}{\partial y_i \partial x_j^2} & \cdots \\ \frac{\partial}{\partial z_i} & \frac{\partial^2}{\partial z_i \partial x_j} & \frac{\partial^2}{\partial z_i \partial y_j} & \frac{\partial^2}{\partial z_i \partial z_j} & \frac{\partial^3}{\partial z_i \partial x_j^2} & \cdots \\ \frac{\partial^2}{\partial x_i^2} & \frac{\partial^3}{\partial x_i^2 \partial x_j} & \frac{\partial^3}{\partial x_i^2 \partial y_j} & \frac{\partial^3}{\partial x_i^2 \partial z_j} & \frac{\partial^4}{\partial x_i^2 \partial x_j^2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix} \frac{1}{r_{i,j}} \tag{1.67}$$

As a polarizable force field, Amoeba uses induced dipole moments to describe how the electron density (as represented by the permanent multipoles) are polarized by an external field.

$$\mu^i_{ind,l} = \alpha^i \left(\sum_j \sum_m T^{i,j}_{l,m} M^j_{perm,m} + \sum_{j'} \sum_m T^{i,j'}_{l,m} \mu^{j'}_{ind,m} \right) \tag{1.68}$$

where the α is the polarizability for atom i, j are atoms on other molecules and j′ are the other atoms. The polarizability is determined by atom type. Notice that Eq 1.68 implies that the induced dipoles of all the atoms are interdependency since the induced dipoles on one atom depends on the induced dipoles on all other atoms, which means they must be solved for iteratively. The polarization energy is determined using the Thole's damped interaction method.[73]

AMOEBA treats its van der Waals interaction using a buffered 14-7. While not as widely used as a Lennard-Jones Potential, this method better represents the the pair-wise interactions of noble gas.[74] All the parameters in this equation a determined by atom type.

$$E_{vdw} = \sum_{i<j} \epsilon_{ij} \left(\frac{1.07}{\rho_{ij} + 0.07} \right)^7 \left(\frac{1.12}{\rho_{ij}^7 + 0.12} - 2 \right) \qquad \rho_{ij} = \frac{r_{ij}}{r_{ij}^0} \tag{1.69}$$

AMOEBA has two extra terms in the intramolecular portion that are not in the AMBER model, the out-of-plane term E_{oop} and the bond-angle cross term $E_{b\theta}$.

$$E = E_b + E_a + E_d + E_{oop} + E_{b\theta} \tag{1.70}$$

The out-of-plane term uses a Wilson-Decius-Cross-function to restrain the sp^2 hybrid trigonal center from out-of-plane motions. The bond-angle cross terms allows the two bonds to stretch when the angle between them is reduced. The bond and angle terms are not treated harmonically. Instead the bond term is treated with a fourth-order Taylor expansion of a Morse potential and angles are treated with a sixth-order potential. For more details on AMOEBA see Ref 75.

$$E_b = \sum_b k_b (b - b_0)^2 \left[1 - 2.55(b - b_0) + 3.793125(b - b_0)^2 \right] \tag{1.71}$$

$$\begin{aligned} E_a \;=\; & \sum_a k_a (\theta_a - \theta_{0,a})^2 \Bigg[1 - 0.0014\,(\theta_a - \theta_{0,a}) + 5.6 * 10^{-5}\,(\theta_a - \theta_{0,a})^2 \\ & -7.0 * 10^{-7}\,(\theta_a - \theta_{0,a})^3 + 2.2 * 10^{-8}\,(\theta_a - \theta_{0,a})^4 \Bigg] \end{aligned} \tag{1.72}$$

$$E_{oop} = \sum_{\chi} K_{\chi}\chi \tag{1.73}$$

$$E_{b\theta} = \sum_{b\theta} K_{b\theta}\left[(b - b_0) + \left(b' - b'_0\right)\right](\theta - \theta_0) \tag{1.74}$$

1.6.3 AIFF

The ab initio force field (AIFF) is another polarizable force field. Unlike AMOEBA, all parameters used for the intermolecular terms are derived from quantum calculations, while AMOEBA only derives the multipoles using quantum calculations. Also AIFF parameters are determined at the current molecular geometry, rather than at the equilibrium geometry like in AMOEBA. This makes AIFF more reliable, but also more computationally expensive since the parameters have to be rederived at every geometry. AIFF does not include any intramolecular terms. As will be discussed in chapter 1.8, the fragment-based method which uses AIFF determines all intramolecular energy using quantum mechanics.

For the intermolecular terms, the electrostatic term is separated into permanent and induced electrostatics. Long-range van der Waals terms are described by a dispersion term. Short-range terms like exchange that are normally lumped in to the van der Waals contribution are captured by quantum calculations in the fragment-based method.

$$E = E_{elect}^{perm} + E_{elect}^{ind} + E_{disp} \tag{1.75}$$

The permanent electrostatic term is similar to the electrostatic term in AMOEBA.

$$E^{perm}_{elect} = \sum_{i<j} (\mathbf{M}^i_{perm})^T \mathbf{T}^{ij} \mathbf{M}^j_{perm} \tag{1.76}$$

The multipoles are truncated at quadruples for hydrogens while heavier atoms truncate at hexadecapoles. Unlike AMOEBA, all induced multipoles, not just the dipoles, are polarized. Similar to AMOEBA, the induced dipoles are interdependent and determined self-consistently through an iterative process to converge the induction energy.

$$M^i_{ind,l} = -\sum_j \sum_{l'} \sum_m \alpha^i_{ll'} f_n(r,\beta) T^{i,j}_{l',m} \left(M^j_{ind,m} + M^j_{perm,m} \right) \tag{1.77}$$

The induction energy is given by

$$E^{ind}_{elect} = \frac{1}{2} \sum_i \sum_j \sum_l \sum_m M^i_{ind,l} f_n(r,\beta) T^{i,j}_{l,m} M^j_{perm,m} \tag{1.78}$$

Atom i and j are any atom not on the same molecule. The polarizablity in AIFF's $\alpha^i_{ll'}$ have different values depending on which rank (e.g. change, dipole, quadropole) of the multipoles while AMOEBA has a single polarizablity for an atom. The Tang and Toennies damping function $f_n(R,\beta)$ was introduced to weaken induction in the short range to avoid the polarization catastrophe. The subscript n is the order of the electrostatic interaction $r^{-n}_{i,j}$ in the interaction matrix, $T^{i,j}_{l',m}$ (once the derivative is applied in Eq 1.67)

$$f_n(r,\beta) = 1 - \sum_{k=0}^{n} \left(\frac{\beta r^k}{k!} \right) exp(-\beta r) \tag{1.79}$$

The AIFF model includes both 2-body and 3-body dispersion terms.

$$E_{disp} = E_{disp}^{2-body} + E_{disp}^{3-body} \tag{1.80}$$

$$E_{disp}^{2-body} = -\sum_{i<j} \left(f_6(r_{ij},\beta)\frac{C_6}{r_{ij}^6} + f_8(r_{ij},\beta)\frac{C_8}{r_{ij}^8} \right) \tag{1.81}$$

$$E_{disp}^{3-body} = \sum_{i,j,k} f_9(r_{ij},\beta)C_9\frac{1 + 3\cos a \cos b \cos b}{r_{ij}^3 r_{ik}^3 r_{jk}^3} \tag{1.82}$$

where a, b, and c are the angles made between atoms i,j, and k not the same molecule. The dispersion coefficients (C_n) are derived using integration over the isotropic frequency-dependent polarizabilities. See citation 72 for more details. Three-body Axilrod-Teller-Muto dispersion term is often neglected in force fields, but it can be significant in crystals of non-polar molecules such as benzene.[76,77]

1.7 Symmetry

Molecular and crystal symmetry is significant for many molecular properties such as the selection rules for electronic transitions, IR/Raman spectroscopy, and chirality. The symmetry of a molecule determines its point group. Examples of point groups in the Schoenflies notation are C_{2v}, C_1, and D_{3d}. Point groups denote a set of symmetry operations that when applied to the member produces an identical structure.

$$\mathbf{q}_2 = \mathbf{R}\mathbf{q}_1 \tag{1.83}$$

The $\mathbf{q}$ are nuclear position of the atoms in the molecule and $\mathbf{R}$ is the matrix representation of the operator. While the coordinates of each atom may change due to the operation, the overall structure is unchanged as each atom whose position is altered is replaced by an identical atom.

Possible symmetry operators include rotation, reflection, inversion, improper rotations, and the identity. The rotation operator rotates around some axis by 360/n degrees where n is some integer. A rotation operator is denoted by C_n or n. A reflection operation or a mirror plane operator requires a plane which has identical atoms/positions on either side. The reflection operator moves the atom to the same placement on the other side of the plane. The reflection operation is usually denoted by σ or m. The inversion operator changes the sign of the x,y, and z coordinate in a Cartesian coordinate system with an origin at the inversion center. The inversion center is usually the molecule's center of mass and denoted by i. An improper rotation or roto-inversion operation is a C_n operation followed by reflection. The rotation axis and the reflection plane are perpendicular. The improper rotation is denoted by S_n or $\bar{\mathrm{n}}$. The identity operation, denoted by I or E, does not change the coordinates of any atoms. All point groups have the identity operator. Note the choice of origin and axes are significant for these operations to reproduce the same molecule.

Periodic crystals belong to space groups as well as point groups. Space groups are similar to point groups, except the symmetry involves not only the symmetry operator $\mathbf{R}$, but also a translational component $\mathbf{t}$.

$$\mathbf{q}_2 = \mathbf{R}\mathbf{q}_1 + \mathbf{t} \tag{1.84}$$

Because of the periodicity of the crystal and the space group operations, many atoms or ions are symmetrically equivalent. It is common to specify the coordinates of atoms of the crystal in terms of only the symmetrically unique atoms and to have the rest defined by the space group operations. The cell composed of only symmetrically unique atoms is called the asymmetric unit. Space group symmetry is significant for Chapter 2. For more information on point and space group symmetry see Ref 78.

1.8 HMBI

1.8.1 Formalism

Many fragment-based methods exist,[79–81] most of which fall under a common, unifying framework.[82,83] Fragment approaches for molecular crystals typically decompose the total energy of a crystal according to a many-body expansion,

$$E = E_{1-body} + E_{2-body} + E_{3-body} + ... \tag{1.85}$$

The first term in the many-body expansion describes the one-body or intramolecular interactions, which is just the sum of energy of the individual molecules.

$$E_{1-body} = \sum_{i} E_i \tag{1.86}$$

The next term is the two-body or pairwise intermolecular interactions. This term is the sum of energies of the all dimers subtracting their one-body contributions.

$$E_{2-body} = \sum_{i,j} (E_{i,j} - E_i - E_j) \tag{1.87}$$

The three-body intermolecular interactions are the sum of the energies of the trimers subtracting the two-body and one body interactions. The three body and higher-order terms are all combined into the many-body terms.

The one-body (intramolecular conformation) and short-range two-body terms contribute the most to the crystal lattice energy. Therefore, the HMBI model partitions the two-body terms into short-range and long-range dimers based on the intermolecular distance. The one-body and short range two-body interactions are modeled with quantum mechanics. The long-range two-body interaction and the many-body interactions are treated with molecular mechanics.

$$E^{HMBI} = E^{QM}_{1-body} + E^{QM}_{SR-2-body} + E^{MM}_{LR-2-body} + E^{MM}_{many-body} \tag{1.88}$$

where

$$E^{MM}_{many-body} = E^{MM}_{full} - E^{MM}_{SR-2-body} - E^{MM}_{LR-2-body} - E^{MM}_{1-body} \tag{1.89}$$

With subsequent manipulations, one arrives at the final, working HMBI equation:

$$\begin{aligned} E^{HMBI} &= E^{QM}_{1-body} + E^{QM}_{SR-2-body} + E^{MM}_{LR-2-body} \\ &\quad + E^{MM}_{full} - E^{MM}_{SR-2-body} - E^{MM}_{LR-2-body} - E^{MM}_{1-body} \\ &= E^{MM}_{full} + \left(E^{QM}_{1-body} - E^{MM}_{1-body}\right) + \left(E^{QM}_{SR-2-body} - E^{MM}_{SR-2-body}\right) \\ &= E^{MM}_{full} + \sum_{i}\left(E^{QM}_{i} - E^{MM}_{i}\right) + \sum_{i<j} d_{ij}(R)\left(\Delta^2 E^{QM}_{ij} - \Delta^2 E^{MM}_{ij}\right) \end{aligned} \tag{1.90}$$

where

$$\Delta^2 E_{ij} = E_{ij} - E_i - E_j \tag{1.91}$$

As noted in Section 1.5.4, one typically needs to correct for basis set superposition error using the counterpoise correction. When doing so, the molecule energies E_i and E_j are determined using the same “dimer” basis set as for the dimer energy E_{ij}.

The sigmoidal function $d_{ij}(R)$ in Eq 1.90 is used to transition smoothly between the short-range QM and the long-range MM treatment of the two-body interactions. It is defined by two cutoffs, c_1 and c_0. If the distance of the closest atoms of the molecules in the dimers R is less than c_1, $d_{ij}(R)$ is equal to 1. If R greater than c_0, $d_{ij}(R)$ is equal to 0.

If R is between c_0 and c_1,

$$d_{ij}(R) = \frac{1}{1 + exp\left(\frac{2|c_0 - c_1|}{(c_1 - R)} - \frac{|c_0 - c_1|}{R - c_0}\right)} \tag{1.92}$$

The cutoffs c_1 and c_0 usually only differ by 1 Å. Cutoffs should be carefully chosen since number of QM dimers increase quickly. Conservative cutoffs of 9 and 10 Å are often used.[44] When determining the unit cell energy of an infinite crystal with periodic boundary conditions (PBC), the energy of the full crystal is determined by Ewald summation[84] and two-body interactions between molecules inside and outside the unit cell are considered, Eq 1.90 becomes

$$\begin{aligned} E^{HMBI} &= E_{PBC}^{MM} + \sum_{i}\left(E_i^{QM} - E_i^{MM}\right) + \sum_{i<j} d_{ij}(R)\left(\Delta^2 E_{ij}^{QM} - \Delta^2 E_{ij}^{MM}\right) \\ &\quad + \frac{1}{2}\sum_{i,k} d_{ik}(R)\left(\Delta^2 E_{ik}^{QM} - \Delta^2 E_{ik}^{MM}\right) \end{aligned} \tag{1.93}$$

where i and j are molecules inside the central unit cell and k are the molecules outside the central unit cell. The two-body interaction between molecule i and molecule k is halved because only one of the molecules that make up the dimer is in the unit cell.

The user can choose which QM and MM methods to use for the different terms in HMBI. Typically, MP2 or CCSD(T) are used for the QM terms. Both Amoeba and AIFF have been successfully applied for the MM terms, with the latter generally giving more reliable results.[45,85]

1.8.2 Previous Successes of HMBI Model

HMBI has been demonstrated to be a reliable method. The errors in energy of clusters of small molecules relative to full quantum mechanics are small.[86] The lattice energy of small molecule molecular crystals were within 2-4 kJ/mol of the experimental results when the basis set of the qm portion was extrapolated to the complete basis set and CCSD(T) correction was used.[44] HMBI has determined the lattice parameters of a series of molecular crystals with an average RMS error of 1.6%.[87] HMBI also correctly ordered the stability of the polymorphed of oxalyl dihydrazide where other computational methods such as disperson-corrected periodic DFT failed.[16]

HMBI has gone beyond just reproducing experimental results, it has also been used to answer chemically interesting questions. HMBI was used to demonstrate the near-degeneracy of the two polymorphs of aspirin and to explain the energetics as an interplay between a more relaxed intramolecular conformation in form I and the improved intermolecular interactions arising from the long, catemeric chains of hydrogen bonds in form II.[15] HMBI was also used to explore the controversial proton-ordering in ice XV. Previous DFT calculations suggested a ferroelectric proton arrangement (i.e. a non-zero cell dipole moment) in the unit cell, while the structure inferred from experiment had an antiferroelectric arrangement (zero dipole moment). HMBI MP2 and CCSD(T) calculations predict that the anti-ferroelectric structure inferred from experiment is indeed more stable, apparently confirming the experimental results.[88] Interestingly, however, subsequent fully periodic MP2 calculations (without any fragment approximation) predict that the ferroelectric structure is more stable,[89] re-igniting the controversy. Finally, the HMBI code using a fragment-based DFT method has proved very successful in predicting nuclear magnetic resonsance chemical shifts.[90,91]

1.9 Outline of this Dissertation

Although HMBI has already proved useful for modeling clusters and periodic crystals, there are serveral ways in which it can be improved. First, the computational effort of HMBI is dominated by the evaluation of QM dimer contributions for the short-range two-body interactions. As mentioned in Section 1.8.1, the number of QM dimers in a calculation is determined by the QM/MM cutoffs. At the conservative 9 and 10 Å cutoffs, a typical small-molecule crystal requires require hundreds for QM dimer calculations. In an extreme case, the acetamide crystal requires approximately a thousand dimers at these cutoffs due to the unusually high 18 monomers in the unit cell.

However, large numbers of the monomer and dimer calculations in HMBI are redundant. Periodic crystals belong to space groups with specific symmetry operators. Using these operators to identify symmetry equivlent dimers, we can trim the number of expensive QM calculations required for the crystal. In the case of acetamide, exploiting symmetry leads to a 18 fold decrease in the number of dimer calculations required. In addition to reducing the number of dimers, employing space group symmetry reduces the number degrees of freedom for a unit cell geometry optimization, thereby often reducing the steps required to converge the optimization. The details of an efficient algorithm to exploit space group symmetry in HMBI calculations are discussed in Chapter 2.[92]

Second, previous research using HMBI neglected consideration of finite temperatures and pressures, despite their potentially significant impacts on crystal structures, stabilities, and other properties. Pressure dependence can be added to the model relatively easily, but accounting for thermal vibrational free energy is more challenging. In particular, the crystal structure expands upon heating, which weakens the lattice energy and alters the vibrational mode frequencies. To incorporate these finite-temperature effects, the HMBI model is coupled here to the quasi-harmonic approximation which accounts for the effects of the changing crystal structure on the phonon modes. Chapter 3 demonstrates that the combination of high-accuracy electronic structure calculations and quasi-harmonic expansion allows one to predict the structure, thermochemistry, and mechanical properties of carbon dioxide in excellent agreement with experiment.[93] Chapter 4 extends the analysis of quasi-harmonic thermal expansion to several other small-molecule crystals.[94] It particularly examines how sigificant thermal expansion is for predicting thermochemistry accurately. Finally, Chapter 5 provides preliminary results exploring the application of these techniques to predicting the solid phase diagram of carbon dioxide as a function of temperature and pressure. The work in Chapters 3–5 is feasible in large part due to the symmetry algorithms developed in Chapter 2.

Chapter 2

Exploiting space group symmetry in fragment-based molecular crystal calculations

2.1 Outline

The HMBI and other fragment-based methods enable high-level ab initio electronic structure calculations in for molecular crystals. Such studies remain computationally demanding, however. Here, we describe a straightforward algorithm for exploiting space group symmetry in fragment-based methods which often provides computational speed-ups of several fold or more. This algorithm does not require *a priori* specification of the space group or symmetry operators. Rather, the symmetrically equivalent fragments are identified automatically by aligning the individual fragments along their principle axes of inertia and testing for equivalence with other fragments. The symmetry operators relating equivalent fragments can then be worked out easily. Implementation of this algorithm for computing energies, nuclear gradients with respect to both atomic coordinates and lattice parameters, and the nuclear Hessiaxn is described. For a brief review of space group symmetry, see Section 1.7.

2.2 Introduction

Molecular crystal structure plays an important role in determining the physical and chemical properties of a wide variety of materials, ranging from pharmaceuticals[3,95,96] to organic semiconductors.[4,97–99] A given organic molecule may crystallize in several different forms, or polymorphs. Modeling the subtle energetic differences among these polymorphs often requires accurate quantum mechanical treatments such as those made possible through fragment-based electronic structure methods.[15–18,22,33,45,80,88,100–110]

Fragment-based methods such as HMBI were discussed in Section 1.8.1. Fragment approaches for molecular crystals typically decompose the energy of the total crystal according to a many-body expansion

$$E_{crystal} = E_{1-body} + E_{2-body} + E_{3-body} + \cdots \tag{2.1}$$

where a one-body term involves the energy of a single molecule, a two-body term involves the interaction between a pair of molecules, etc. Higher-order terms in this expansion grow increasingly expensive computationally both because they involve larger numbers of molecules in each fragment and because there are a combinatorial number of possible groupings of the fragments (e.g. many more trimers exist than dimers). For both physical reasons and practical computational efficiency, one frequently approximates the three-body and higher terms in some fashion. For example, electrostatic embedding is often used to capture many-body polarization effects.[17,102,111] Alternatively, our hybrid many-body interaction (HMBI) approach treats the one-body and short-range two-body interactions with quantum mechanics, while the long-range two-body and higher-order terms are approximated with a polarizable force field.[44,45,80]

Even with fragmentation, accurate molecular crystal calculations can be very computationally demanding,[44,64] so it is important to perform these calculations as efficiently as possible. In a fragment method like HMBI, the cost of evaluating the two-body interactions in the crystal forms the computational bottleneck in evaluating crystal energetics, optimizing structures, or calculating harmonic vibrational frequencies. Done naively, calculating the two-body contribution to the lattice energy from all pairs of molecules lying within 10 Å of each other in a crystal like formamide requires evaluating some 250 dimer interactions. However, many of these dimers are symmetrically equivalent due to the overall $P2_1/c$ space group symmetry of the crystal. Therefore, one only needs to calculate the symmetrically unique fraction of those dimers to obtain the correct crystal energy.

Some of the most common molecular crystal space groups exhibit four symmetrically equivalent molecules in the unit cell, and speed-ups of four- to eight-fold are often achievable by exploiting space-group symmetry. Other space groups have higher symmetry factors, which can lead to even greater savings. In acetamide ($R3c$ symmetry), for example, there are eighteen symmetrically equivalent monomers, and can one obtain eighteen-fold computational speed-ups by using symmetry to eliminate the redundant dimer calculations.

In principle, one could implement tables of symmetry operators for each of the 230 space groups[112] in the software handling the calculations, then specify the space group of the crystal in question, and use the symmetry operations found in those tables to identify symmetrically redundant calculations. Such an approach is tedious to implement, however. Of course, one could implement symmetry for only the most common space groups,[113] which would capture a large fraction of known organic molecular crystal structures, but not all. Sode and Hirata implemented specific space group operators for their study on hydrogen fluoride, for example.[100]

This paper describes a simple alternative approach for identifying and exploiting symmetrically equivalent monomers and dimers in a fragment-based molecular crystal calculation. This approach does not require one to implement space group tables or even to know the symmetry of the system *a priori.* Rather, it relies on rotating each monomer or dimer fragment to a common orientational frame based on the principle axes of inertia, and then testing for equivalence with other fragments. The symmetry operators relating the equivalent fragments can be extracted straightforwardly. Once the list of symmetrically unique monomers and dimers is known, one can compute the crystal energy readily. Exploiting the symmetry equivalences in nuclear gradients and Hessians requires slightly more programming effort, but it can also be done.

The discussion here assumes that the calculation of the entire crystal is being broken down into a series of monomer fragments as per the HMBI approach, and that those fragments are then combined to form dimers. However, the ideas employed herein can be generalized readily to incorporate higher-order terms in the many-body expansion (e.g. trimers) or to other fragmentation schemes.

2.3 Theory

2.3.1 Identification of symmetrically equivalent fragments

The first step of the symmetry algorithm is to generate a complete list of monomer and dimer fragments whose quantum mechanical energies will be used to determine the total energy of the crystal. In HMBI, this list includes all monomers in the central unit cell and all "short-range" dimers for which the molecules are separated by less than a user-defined cutoff radius. The dimer list contains both dimers involving two molecules in the central unit cell as well as those involving one periodic image molecule that lies outside the unit cell.

This list is subsequently pruned by identifying the symmetrically equivalent monomers and dimers. In addition, one determines the space group symmetry operators $\mathbf{R}$ and $\mathbf{t}$,

$$\mathbf{q}_2 = \mathbf{R}\mathbf{q}_1 + \mathbf{t} \tag{2.2}$$

that relate the Cartesian atomic coordinates of an atom in fragment 2, $\mathbf{q}_2$, to the symmetrically equivalent atom in fragment 1, $\mathbf{q}_1$ via a 3×3 rotation matrix $\mathbf{R}$ and a 3×1 translation vector $\mathbf{t}$.

To identify the symmetrically equivalent monomers in the central unit cell, each molecule is translated to center-of-mass coordinates and rotated into standard nuclear orientation. That is, the monomer's symmetric 3×3 inertia tensor is calculated according to

$$\mathbf{I} = \begin{pmatrix} \sum_i m_i(y_i^2 + z_i^2) & -\sum_i m_i x_i y_i & -\sum_i m_i x_i z_i \\ -\sum_i m_i y_i x_i & \sum_i m_i(x_i^2 + z_i^2) & -\sum_i m_i y_i z_i \\ -\sum_i m_i z_i x_i & -\sum_i m_i z_i y_i & \sum_i m_i(x_i^2 + y_i^2) \end{pmatrix} \tag{2.3}$$

where x_i, y_i, and z_i are the center-of-mass Cartesian coordinates of the i-th nucleus and m_i is its mass, and the sums run over all nuclei in the monomer. The three eigenvectors of the inertia tensor define the orthogonal principle axes of inertia, and the eigenvalues provide the corresponding moments of inertia. The monomer is then rotated so that its axes of inertia align with the coordinate axes.

For a general, asymmetric top with three non-degenerate moments of inertia, we adopt the convention that the largest moment is aligned along the z-axis, the middle moment is aligned along the y-axis, and the smallest moment is aligned along the x-axis. For a symmetric top, the axis of inertia corresponding to the non-degenerate eigenvalue is aligned along the z-axis. The monomer is then rotated about the z-axis until the atom with the lowest projection into the xy-plane (in center-of-mass coordinates) lies in the xz-plane. For a spherical top, the three moments of inertia are degenerate and cannot be used to uniquely specify the orientation. Instead, the molecule is rotated such that one of the multiple equivalent atoms closest to the origin lies along the z-axis, and an equivalent atom is rotated into the xz-plane. For a linear top, the molecule is aligned along the z-axis.

Once the molecules have been oriented in this manner, symmetrically equivalent monomers will have identical atomic coordinates to within mirror-plane reflections along the principle axes and some small, finite numerical tolerance (e.g. 10^{-4} Å). Therefore, one can easily test pairs of monomers for symmetry equivalence.

Once a pair of symmetrically equivalent monomers is identified, the space group operation relating them can be found trivially. The rotation operator mapping from monomer 1 to monomer 2, $\mathbf{R}^{2\leftarrow 1}$ is obtained as the appropriate product of the individual rotation operators $\mathbf{R_i}$ used to rotate the individual molecules into standard nuclear orientation and any mirror reflection operators $\mathbf{M}$ required to make the monomers equivalent.

$$\mathbf{R}^{2\leftarrow 1} = \mathbf{R}_2^T \mathbf{M} \mathbf{R}_1 \tag{2.4}$$

The translation operator $\mathbf{t}^{2\leftarrow 1}$ is simply given by the vector difference between the centers of mass of each monomer after the rotation of monomer 1 has been performed with respect to the origin of the global coordinate system.

The same procedure can be applied to the list of possible dimers. Degeneracies among the moments of inertia are rare for dimers, so most dimers will be asymmetric tops. In the end, one stores a list of symmetrically unique fragments. Exploiting space group symmetry in fragment-based molecular crystal calculations, the degeneracy factor σ of each symmetrically unique fragment, and the set of space group operations that generate the symmetrically equivalent fragments from the unique one.

For crystals with high symmetry point group fragment, the symmetrical equivalency between atoms on unique monomers is exploited using point group symmetry. See Appendix A for details.

2.3.2 Incorporating symmetry into the fragment-based HMBI model

The fragment-based quantum/classic hybrid many-body interaction approach treats the important one- and short-range two-body interactions with quantum mechanics (QM), while the long-range two-body and many-body interactions are treated with a classical polarizable molecular mechanics (MM) force field,

$$E_{total}^{HMBI} = E_{1\text{-}body}^{QM} + E_{short-range\ 2\text{-}body}^{QM} + E_{long-range\ 2\text{-}body}^{MM} + E_{many\text{-}body}^{MM} \tag{2.5}$$

To compute the MM terms for a periodic crystal in practice, one computes the full MM crystal energy lattice sum ($E_{crystal}^{MM}$), subtracts off the one-body and short-range two-body MM interactions, and replaces them with their QM counterparts:

$$\begin{aligned} E_{crystal}^{HMBI} &= E_{crystal}^{MM} + \sum_i \left(E_i^{QM} - E_i^{MM}\right) + \sum_{i<j} d_{ij}\left(\Delta^2 E_{ij}^{QM} - \Delta^2 E_{ij}^{MM}\right) \\ &\quad + \frac{1}{2}\sum_i \sum_k^{images} d_{ik}\left(\Delta^2 E_{ik}^{QM} - \Delta^2 E_{ik}^{MM}\right) \end{aligned} \tag{2.6}$$

In this expression, i and j correspond to monomers in the central unit cell, while k corresponds to periodic image monomers. The two-body interactions are given as the difference between the dimer energies E_{ij} and the monomer energies E_i and E_j: $\Delta^2 E_{ij} = E_{ij} - E_i - E_j$. The sigmoidal function $d_{ij}(R)$ is used to interpolate smoothly between the short-range QM and long-range MM treatments of the two-body interactions and is defined in Ref. 45.

Having identified the set of symmetrically unique monomer and dimer fragments as per Section 2.3.1, one needs only compute the energy for each unique fragment and multiply each term in the HMBI energy equation by a symmetry factor σ corresponding to the degeneracy of that particular monomer or dimer in the crystal:

$$\begin{aligned} E_{crystal}^{HMBI} &= E_{crystal}^{MM} + {\sum_{i}}' \sigma_i \left(E_i^{QM} - E_i^{MM} \right) + {\sum_{i<j}}' \sigma_{ij} d_{ij} \left(\Delta^2 E_{ij}^{QM} - \Delta^2 E_{ij}^{MM} \right) \\ &\quad + \frac{1}{2} {\sum_{i}}' {\sum_{k}^{images}}' \sigma_{ik} d_{ik} \left(\Delta^2 E_{ik}^{QM} - \Delta^2 E_{ik}^{MM} \right) \end{aligned} \tag{2.7}$$

The primes on the sums indicate that they run over only symmetrically unique fragments.

In practice, the number of symmetrically unique monomers that need to be calculated is reduced from the number of molecules in the full unit cell (Z) to the number of molecules in the asymmetric unit cell (Z'). The number of dimer calculations needed for the energy is reduced by a factor of up to two times the number of symmetry operations in the space group, as will be discussed in Section 2.6.

2.3.3 Crystal structure optimization

Optimizing a crystal structure with N atoms in the unit cell requires minimizing the energy with respect to both the $3N$ nuclear coordinates describing the atomic positions and with respect to the six unit cell lattice parameters (a, b, c, α, β, and γ). Imposing symmetry on the crystal simplifies geometry optimization in two ways. First, because many of the monomer and dimer fragments are symmetrically equivalent, one need only compute the forces for the symmetrically unique subset. Second, symmetry imposes constraints on the motions of the atoms—the overall space group symmetry must not be disrupted during the optimization. In practice, this means that the length of the nuclear gradient is reduced from $3N$ to $3N(Z'/Z)$ degrees of freedom.

For example, crystalline formamide ($HCONH_2$) adopts $P2_1/c$ symmetry, with four molecules (24 atoms total) in the unit cell and $3N = 72$ components in the nuclear gradient (plus six lattice parameter terms). However, all four molecules are equivalent by symmetry and the asymmetric cell contains only one molecule (6 atoms). Imposing space group symmetry during the optimization reduces the length of the gradient vector four-fold, to only 18 atomic coordinate terms (plus six lattice parameter terms). Note that if the point group symmetry is exploited for crystals with high symmetry monomers, the gradient vector could further be reduced. (See Appendix A) In the end, exploiting symmetry (1) reduces the number of fragment calculations required for each force calculation and (2) reduces the number of degrees of freedom that need to be optimized, which often reduces the number of geometry optimization cycles required to reach convergence.

2.3.4 Symmetry in the nuclear gradient

Expressions for the HMBI nuclear gradient have been presented previously.[87] Here, those expressions are modified to include the symmetry factors σ_i and σ_{ij}. The nuclear gradient of the HMBI energy with respect to the l-th Cartesian coordinate of atom q on symmetrically unique monomer z given by,

$$\begin{aligned}\frac{\partial E^{HMBI}_{crystal}}{\partial q^z_l} &= \frac{E^{MM}_{crystal}}{\partial q^z_l} + \sigma_z \left(\frac{\partial E^{QM}_z}{\partial q^z_l} - \frac{\partial E^{MM}_z}{\partial q^z_l} \right) \\ &+ \sum_{i<j}{}' \sigma_{ij} d_{ij} \left(\frac{\partial \Delta^2 E^{QM}_{ij}}{\partial q^z_l} - \frac{\partial \Delta^2 E^{MM}_{ij}}{\partial q^z_l} \right) \\ &+ \sum_{i<j}{}' \sigma_{ij} \frac{\partial d_{ij}}{\partial q^z_l} \left(\Delta^2 E^{QM}_{ij} - \Delta^2 E^{MM}_{ij} \right) \\ &+ \frac{1}{2} \sum_{ik}{}' \sigma_{ik} d_{ik} \left(\frac{\partial \Delta^2 E^{QM}_{ik}}{\partial q^z_l} - \frac{\partial \Delta^2 E^{MM}_{ik}}{\partial q^z_l} \right) \\ &+ \frac{1}{2} \sum_{ik}{}' \sigma_{ik} \frac{\partial d_{ik}}{\partial q^z_l} \left(\Delta^2 E^{QM}_{ik} - \Delta^2 E^{MM}_{ik} \right) \qquad (2.8)\end{aligned}$$

The primed sums run over only symmetrically unique fragments. A detailed expression for the gradient of the switching function d_{ij} has been provided previously.[87]

The two-body gradient terms in Eq 2.8 contribute only if either monomer i and/or j (or i or k in the periodic image terms) is symmetrically equivalent to monomer z,

$$\frac{\partial \Delta^2 E_{ij}}{\partial q^z_l} = \frac{\partial E_{ij}}{\partial q^z_l} - \frac{\partial E_i}{\partial q^z_l} - \frac{\partial E_j}{\partial q^z_l} \qquad (2.9)$$

However, an additional subtlety arises in the two-body contributions to the gradient. The final gradient only has elements corresponding to the atoms of monomer(s) z which lie(s) in the asymmetric unit cell. However, some of the symmetrically unique two-body contributions may be calculated in terms of monomers that are nominally different from z, but which are equivalent to it by symmetry. In those cases, one must rotate the gradient contributions for those atoms onto the atoms of monomer z via the rotation operator $\mathbf{R}^{z\leftarrow i}$ computed in Eq 2.4. One or both monomers in the dimer may be symmetry equivalent to monomer z in the asymmetric cell, so contributions with respect to the atoms in each monomer within the dimer are rotated independently (see Figure 2.1):

$$\begin{aligned}\frac{\partial \Delta^2 E_{ij}}{\partial q_l^z} &= \sum_{l'} \left(\frac{\partial \Delta^2 E_{ij}}{\partial p_{l'}^i}\right)\left(\frac{\partial p_{l'}^i}{\partial q_l^z}\right) + \sum_{l'} \left(\frac{\partial \Delta^2 E_{ij}}{\partial p_{l'}^j}\right)\left(\frac{\partial p_{l'}^j}{\partial q_l^z}\right) \\ &= \sum_{l'} R_{ll'}^{z\leftarrow i}\left(\frac{\partial \Delta^2 E_{ij}}{\partial p_{l'}^i}\right) + \sum_{l'} R_{ll'}^{z\leftarrow j}\left(\frac{\partial \Delta^2 E_{ij}}{\partial p_{l'}^j}\right) \qquad (2.10)\end{aligned}$$

If i or j equals z, then the rotation matrix is the identity matrix. If either monomer i or j is not symmetrically equivalent to monomer z, then the corresponding rotation matrix is a 3×3 zero matrix.

Note that one often employs a counterpoise correction to compensate for basis set superposition error when optimizing molecular crystals. This implies that the monomer contributions in Eq 2.9 involve ghost basis functions and are already in the same orientation as the dimer term. If no counterpoise correction is applied, however, one needs to ensure the monomer and dimer contributions are rotated to the same frame before evaluating Eq 2.9.

In addition, the evaluation of Eq 2.8 also requires the analogous rotation of contributions involving the gradient of the damping function d_{ij},

$$\frac{\partial d_{ij}}{\partial q_l^z} = \sum_{l'} R_{ll'}^{z\leftarrow i}\left(\frac{\partial d_{ij}}{\partial p_{l'}^i}\right) + \sum_{l'} R_{ll'}^{z\leftarrow j}\left(\frac{\partial d_{ij}}{\partial p_{l'}^j}\right) \qquad (2.11)$$

2.3.5 Symmetry in the lattice parameter gradient

Fully optimizing a crystal structure also requires minimizing the energy with respect to the six unit cell lattice parameters (a, b, c, α, β, γ). Ref 87 expresses the gradient

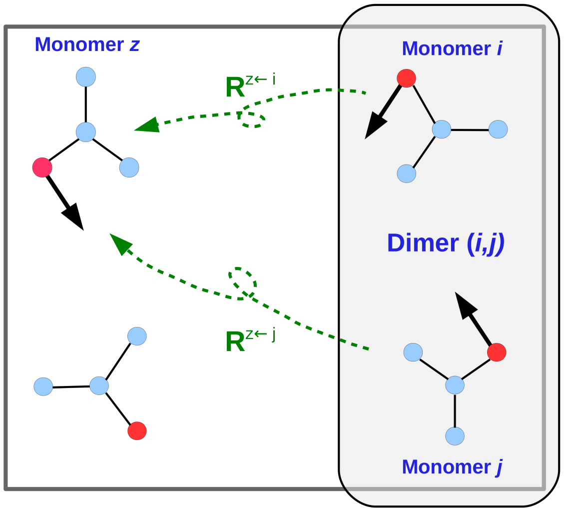

Figure 2.1: If monomers i and j are equivalent by symmetry to monomer z in the asymmetric unit cell, the two-body contributions to the gradient (black arrows) from the highlighted red atom on each monomer in dimer (i, j) must rotated onto the corresponding red atom on monomer z.

of the HMBI energy with respect to lattice parameter u as,

$$\frac{\partial E^{HMBI}_{crystal}}{\partial u} = \frac{\partial E^{MM}_{crystal}}{\partial u} + \frac{\partial E'}{\partial u} \tag{2.12}$$

where $\frac{\partial E'}{\partial u}$ corresponds to the gradient of the last five terms in Eq 2.8 and u is one of the six lattice parameters. Because there are only a handful of lattice parameters and because the force field calculations are computationally inexpensive, we generally obtain the gradient of the full MM crystal energy with respect to the lattice parameters, $\frac{\partial E^{MM}_{crystal}}{\partial u}$, via finite difference, while $\frac{\partial E'}{\partial u}$ is determined analytically.

Evaluating these gradients $\frac{\partial E'}{\partial u}$ is easier in terms of the Cartesian components of the lattice column vectors $\mathbf{v}_1$,$\mathbf{v}_2$, and $\mathbf{v}_3$:

$$(\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3) = \begin{pmatrix} a & b\cos\gamma & c\cos\beta \\ 0 & b\sin\gamma & c\frac{\cos\alpha-\cos\beta\cos\gamma}{\sin\gamma} \\ 0 & 0 & c\frac{\sqrt{\sin^2\gamma-\cos^2\alpha-\cos^2\beta+2\cos\alpha\cos\beta\cos\gamma}}{\sin\gamma} \end{pmatrix} \tag{2.13}$$

Note that $\mathbf{v}_1$ lies along the global x-axis, $\mathbf{v}_2$ lies in the xy plane, and $\mathbf{v}_3$ can lie anywhere in xyz space. Using the chain rule of differentiation, the energy gradient with respect to the lattice parameters can be related to the gradients with respect to the unit cell vector components $\mathrm{v}_{\epsilon l}$,

$$\frac{\partial E'}{\partial u} = \sum_{\epsilon=1}^{3} \sum_{l=x,y,z} \frac{\partial E'}{\partial \mathrm{v}_{\epsilon l}} \frac{\partial \mathrm{v}_{\epsilon l}}{\partial u} \tag{2.14}$$

By definition (see Eq 2.13), three of the partial derivatives of $\mathrm{v}_{\epsilon l}$ are zero $\left(\frac{\partial \mathrm{v}_{1y}}{\partial u} = \frac{\partial \mathrm{v}_{1z}}{\partial u} = \frac{\partial \mathrm{v}_{2z}}{\partial u} = 0\right)$ and can be neglected. The remaining six terms are given by (*cf.* Eq 2.6),

$$\begin{aligned} \frac{\partial E'}{\partial \mathrm{v}_{\epsilon l}} \quad = \quad & \sum_{i<j} d_{ij} \sum_{q} \frac{\partial q_l^{ij}}{\partial \mathrm{v}_{\epsilon l}} \left(\frac{\partial \Delta^2 E_{ij}^{QM}}{\partial q_l^{ij}} - \frac{\partial \Delta^2 E_{ij}^{QM}}{\partial q_l^{ij}} \right) \\ & + \sum_{i<j} \left(\Delta^2 E_{ij}^{QM} - \Delta^2 E_{ij}^{MM} \right) \sum_{q} \frac{\partial q_l^{ij}}{\partial \mathrm{v}_{\epsilon l}} \frac{\partial d_{ij}}{\partial q_l^{ij}} \\ & + \sum_{ik} d_{ik} \sum_{q} \frac{\partial q_l^{ik}}{\partial \mathrm{v}_{\epsilon l}} \left(\frac{\partial \Delta^2 E_{ik}^{QM}}{\partial q_l^{ik}} - \frac{\partial \Delta^2 E_{ik}^{QM}}{\partial q_l^{ik}} \right) \\ & + \sum_{ik} \left(\Delta^2 E_{ik}^{QM} - \Delta^2 E_{ik}^{MM} \right) \sum_{q} \frac{\partial q_l^{ik}}{\partial \mathrm{v}_{\epsilon l}} \frac{\partial d_{ik}}{\partial q_l^{ik}} \end{aligned} \tag{2.15}$$

As before, i and j refer to monomers in the central unit cell, while k refers to a periodic image monomer. The index q^{ij} sums over atoms in dimer (i, j), and l refers to the Cartesian coordinates x, y, or z. The $\frac{\partial q_l^{ij}}{\partial \mathrm{v}_{\epsilon l}}$ terms come from the chain rule,

$$\frac{\partial E'}{\partial \mathrm{v}_{\epsilon l}} = \sum_q \frac{\partial E'}{\partial q_l} \frac{\partial q_l}{\partial \mathrm{v}_{\epsilon l}} \tag{2.16}$$

As described previously,[87] expressing the gradients in terms of Cartesian coordinates instead of fractional ones means that the one-body terms do not contribute to the lattice parameter gradient, Eq 2.15. The exception are monomers belonging to high symmetry point groups. In these cases, the coordiates of atoms of symmetrial unique monomers may adjust as lattice parameters relax to preserve symmetry. These special cases are addressed in Appendix A.

Note that Eq 2.15 does not include any symmetry factors σ_{ij}, and that the sums run over all possible dimers instead of just the symmetrically unique ones. Employing symmetry in the lattice parameter gradient involves capturing how the individual monomers translate to preserve symmetry as the lattice parameters change. To do so, we first compute the gradient contributions for the the symmetrically unique dimers (y, z) and rotate them using the space group operators to generate the complete set of gradient contributions,

$$\frac{\partial E_{ij}}{\partial q_l^{ij}} = \sum_{l'} \left(\frac{\partial E_{yz}}{\partial p_{l'}^{yz}} \right) \left(\frac{\partial p_{l'}^{yz}}{\partial q_l^{ij}} \right) = \sum_{l'} R_{ll'}^{ij \leftarrow yz} \left(\frac{\partial E_{yz}}{\partial p_{l'}^{yz}} \right) \tag{2.17}$$

where $R^{ij \leftarrow yz}$ is the rotation operator that transforms dimer (y, z) into dimer (i, j), and q_l^{ij} refers to the l-th Cartesian coordinate on atom q in dimer (i, j), and $p_{l'}^{yz}$ is defined analogously.

In addition, to preserve symmetry while changing the lattice vectors, one must translate the monomers appropriately (Figure 2.2). The amount of translation is determined

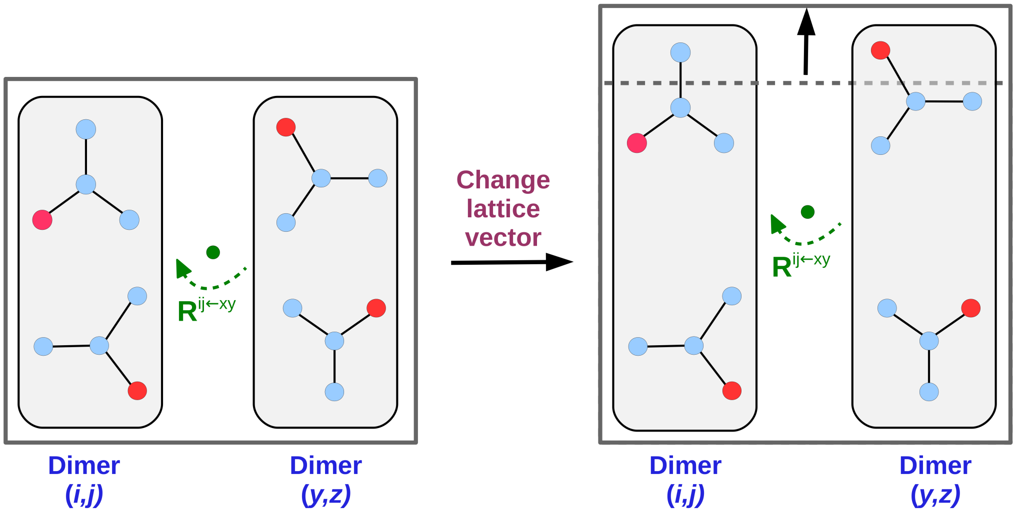

Figure 2.2: Changing the unit cell lattice vectors requires translating the monomers to preserve the symmetry equivalence of dimers (i, j) and (x, y) via rotation operator $\mathbf{R}^{ij\leftarrow xy}$.

by the translation operator $\mathbf{t}$ in the space group operator (Eq 2.2). Accordingly, the term $\frac{\partial q_l}{\partial \mathrm{v}_{\epsilon l}}$ in Eq 2.16 is given by,

$$\frac{\partial q_l}{\partial \mathrm{v}_{\epsilon l}} = \frac{\mathrm{t}_l}{\partial \mathrm{v}_{\epsilon l}} = \hat{l}_t \tag{2.18}$$

where $\hat{l}_t$ are the elements of the fractional coordinate representation of the translational operator $\mathbf{t} = \hat{x}_t \mathbf{v}_1 + \hat{y}_t \mathbf{v}_2 + \hat{z}_t \mathbf{v}_2$.

With all of these ingredients, the final expressions for the gradient with respect to each lattice parameter (Eq 2.12) are computed as,

$$\frac{\partial E^{HMBI}_{crystal}}{\partial a} = \frac{\partial E^{MM}_{crystal}}{\partial a} + \frac{\partial E'}{\partial \mathrm{v}_{1x}} \tag{2.19}$$

$$\frac{\partial E^{HMBI}_{crystal}}{\partial b} = \frac{\partial E^{MM}_{crystal}}{\partial b} + \frac{\partial E'}{\partial \mathrm{v}_{2x}} \cos\gamma + \frac{\partial E'}{\partial \mathrm{v}_{2x}} \sin\gamma \tag{2.20}$$

$$\begin{aligned}\frac{\partial E^{HMBI}_{crystal}}{\partial c} &= \frac{\partial E^{MM}_{crystal}}{\partial c} + \frac{\partial E'}{\partial \mathrm{v}_{3x}}\cos\beta + \frac{\partial E'}{\partial \mathrm{v}_{3y}}\frac{\cos\alpha - \cos\beta\cos\gamma}{\sin\gamma} \\ &\quad + \frac{\partial E'}{\partial \mathrm{v}_{3z}}\frac{\sqrt{\sin^2\gamma - \cos^2\alpha - \cos^2\beta + 2\cos\alpha\cos\beta\cos\gamma}}{\sin\gamma}\end{aligned} \tag{2.21}$$

$$\begin{aligned}\frac{\partial E^{HMBI}_{crystal}}{\partial \alpha} &= \frac{\partial E^{MM}_{crystal}}{\partial \alpha} + c\frac{\sin\alpha}{\sin\gamma}\Bigg(- \frac{\partial E'}{\partial \mathrm{v}_{3y}} \\ &\quad + \frac{\partial E'}{\partial \mathrm{v}_{3z}}\frac{\cos\alpha - \cos\beta\cos\gamma}{\sqrt{\sin^2\gamma - \cos^2\alpha - \cos^2\beta + 2\cos\alpha\cos\beta\cos\gamma}}\Bigg)\end{aligned} \tag{2.22}$$

$$\begin{aligned}\frac{\partial E^{HMBI}_{crystal}}{\partial \beta} &= \frac{\partial E^{MM}_{crystal}}{\partial \beta} + c\sin\beta\Bigg(- \frac{\partial E'}{\partial \mathrm{v}_{3x}} - \frac{\partial E'}{\partial \mathrm{v}_{3y}}\frac{1}{\tan\gamma} \\ &\quad + \frac{\partial E'}{\partial \mathrm{v}_{3z}}\frac{\cos\beta - \cos\alpha\cos\gamma}{\sin\gamma\sqrt{\sin^2\gamma - \cos^2\alpha - \cos^2\beta + 2\cos\alpha\cos\beta\cos\gamma}}\Bigg)\end{aligned} \tag{2.23}$$

$$\begin{aligned}\frac{\partial E^{HMBI}_{crystal}}{\partial \gamma} &= \frac{\partial E^{MM}_{crystal}}{\partial \gamma} + b\left(- \frac{\partial E'}{\partial \mathrm{v}_{2x}}\sin\gamma + \frac{\partial E'}{\partial \mathrm{v}_{2y}}\cos\gamma\right) \\ &\quad + c\Bigg(\frac{\partial E'}{\partial \mathrm{v}_{3y}}\frac{\cos\alpha - \cos\beta\cos\gamma}{\sin^2\gamma} \\ &\quad + \frac{\partial E'}{\partial \mathrm{v}_{3z}}\frac{\cos^2\alpha\cos\gamma + \cos^2\beta\cos\gamma - \frac{1}{2}\cos\alpha\cos\beta(3 - \cos 2\gamma)}{\sin^2\gamma\sqrt{\sin^2\gamma - \cos^2\alpha - \cos^2\beta + 2\cos\alpha\cos\beta\cos\gamma}}\Bigg)\end{aligned} \tag{2.24}$$

To preserve the space group symmetry, gradients with respect to any 90° or 120° lattice angle are constrained to be zero, and gradients with respect to any equivalent lattice constants are combined and treated as a single entry in the gradient during optimization. The steps for constructing the lattice parameter gradient are summarize in Table 2.1.

Table 2.1: Algorithm for computing the energy gradient with respect to lattice parameter u. Note that the loop over dimers in step 2 includes both dimers in the central unit cell and those involving periodic image monomers.

```
1) Compute $\frac{\partial E^{MM}_{crystal}}{\partial u}$ via finite difference (while preserving symmetry)
2) Construct $\frac{\partial E'}{\partial \mathrm{v}_{\epsilon l}}$ for each lattice vector component $\mathrm{v}_{\epsilon l}$
    loop over all dimers $(i, j)$
        generate $\frac{\partial \Delta^2 E^{QM}_{ij}}{\partial q^{ij}_l}$ and $\frac{\partial d_{ij}}{\partial q^{ij}_l}$ from symmetrically unique dimer $(x, y)$ via $R^{ij \leftarrow yz}_{ll'}$ (Eq 2.17)
        compute $\frac{\partial q_l}{\partial \mathrm{v}_{\epsilon l}}$ (Eq 2.18)
        increment $\frac{\partial E'}{\partial \mathrm{v}_{\epsilon l}}$ (Eq 2.15)
    end loop over dimers
3) Transform $\frac{\partial E'}{\partial \mathrm{v}_{\epsilon l}}$ into $\frac{\partial E'}{\partial u}$ (Eq 2.19–2.24)
4) Compute $\frac{\partial E^{HMBI}_{crystal}}{\partial u} = \frac{\partial E^{MM}_{crystal}}{\partial u} + \frac{\partial E'}{\partial u}$
```

2.4 Symmetry in the nuclear Hessian and finding crystal phonons

Computing crystal phonon modes and frequencies in the harmonic approximation requires constructing the nuclear Hessian. Here we focus on frequencies at the Γ-point, though one can generalize this to obtain frequencies at other points in reciprocal space via lattice dynamics.[114] The Hessian is a $3N \times 3N$ matrix, which is diagonalized to produce three translational modes and $3N-3$ phonon modes. In contrast to the gradient, one should not use symmetry to reduce the size of the overall Hessian, as this will eliminate all but the symmetric vibrational modes. However, one can use still symmetry to generate the full set of one- and two-body contributions inexpensively from the symmetrically reduced set.

The second derivative of the energy with respect to the l-th coordinate on atom q

and the m-th coordinate on atom p is given by,

$$
\begin{aligned}
\frac{\partial^2 E^{HMBI}_{crystal}}{\partial q_l \partial p_m} &= \frac{\partial^2 E^{MM}_{crystal}}{\partial q_l \partial p_m} + \left(\frac{\partial^2 \Delta E^{QM}_{i}}{\partial q_l \partial p_m} - \frac{\partial^2 \Delta E^{MM}_{i}}{\partial q_l \partial p_m} \right) \\
&+ d_{ij} \left(\frac{\partial^2 \Delta^2 E^{QM}_{ij}}{\partial q_l \partial p_m} - \frac{\partial^2 \Delta^2 E^{MM}_{ij}}{\partial q_l \partial p_m} \right) \\
&+ \frac{\partial^2 d_{ij}}{\partial q_l \partial p_m} \left(\Delta^2 E^{QM}_{ij} - \Delta^2 E^{MM}_{ij} \right) \\
&+ \frac{\partial d_{ij}}{\partial q_l} \left(\frac{\partial \Delta^2 E^{QM}_{ij}}{\partial p_m} - \frac{\partial \Delta^2 E^{MM}_{ij}}{\partial p_m} \right) \\
&+ \frac{\partial d_{ij}}{\partial p_m} \left(\frac{\partial \Delta^2 E^{QM}_{ij}}{\partial q_l} - \frac{\partial \Delta^2 E^{MM}_{ij}}{\partial q_l} \right) \\
&+ \sum_{n=i,j} \sum_{k} d_{nk} \left(\frac{\partial^2 \Delta^2 E^{QM}_{nk}}{\partial q_l \partial p_m} - \frac{\partial^2 \Delta^2 E^{QM}_{nk}}{\partial q_l \partial p_m} \right) \\
&+ \sum_{n=i,j} \sum_{k} \frac{\partial^2 d_{nk}}{\partial q_l \partial p_m} \left(\Delta^2 E^{QM}_{nk} - \Delta^2 E^{MM}_{nk} \right) \\
&+ \sum_{n=i,j} \sum_{k} \frac{\partial d_{nk}}{\partial q_l} \left(\frac{\partial \Delta^2 E^{QM}_{nk}}{\partial p_m} - \frac{\partial \Delta^2 E^{MM}_{nk}}{\partial p_m} \right) \\
&+ \sum_{n=i,j} \sum_{k} \frac{\partial d_{nk}}{\partial p_m} \left(\frac{\partial \Delta^2 E^{QM}_{nk}}{\partial q_l} - \frac{\partial \Delta^2 E^{MM}_{nk}}{\partial q_l} \right) \qquad (2.25)
\end{aligned}
$$

where

$$
\frac{\partial^2 \Delta^2 E_{ij}}{\partial q_l \partial q_m} = \frac{\partial^2 E_{ij}}{\partial q_l \partial q_m} - \frac{\partial^2 E_{i}}{\partial q_l \partial q_m} - \frac{\partial^2 E_{j}}{\partial q_l \partial q_m} \qquad (2.26)
$$

Once again, rotation matrices are used to rotate the one-body and two-body contributions from the symmetrically unique dimers to the various symmetrically equivalent contributions to generate a complete set of one- and two-body contributions. In this case, two rotation matrices are needed—one for each atomic coordinate used to differentiate the energy. One obtains:

$$\begin{aligned}
\frac{\partial^2 E_i}{\partial q_l^i \partial p_m^i} &= \sum_{l'} \sum_{m'} \left(\frac{\partial r_{l'}^z}{\partial q_l^i} \right) \left(\frac{\partial^2 E_n}{\partial r_{l'}^z \partial s_{m'}^z} \right) \left(\frac{\partial s_{m'}^z}{\partial p_m^i} \right) \\
&= \sum_{l'} \sum_{m'} R_{ll'}^{i \leftarrow z} \left(\frac{\partial^2 E_n}{\partial r_{l'}^z \partial s_{m'}^z} \right) R_{mm'}^{i \leftarrow z} \qquad (2.27)
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 \Delta^2 E_{ij}}{\partial q_l^i \partial p_m^j} &= \sum_{l'} \sum_{m'} \left(\frac{\partial r_{l'}^z}{\partial q_l^i} \right) \left(\frac{\partial^2 \Delta^2 E_{nk}}{\partial r_{l'}^z \partial s_{m'}^{z'}} \right) \left(\frac{\partial s_{m'}^{z'}}{\partial p_m^j} \right) \\
&= \sum_{l'} \sum_{m'} R_{ll'}^{i \leftarrow z} \left(\frac{\partial^2 \Delta^2 E_{nk}}{\partial r_{l'}^z \partial s_{m'}^{z'}} \right) R_{mm'}^{j \leftarrow z'} \qquad (2.28)
\end{aligned}$$

The one-body term (Eq 2.27) only contributes if both atoms q and p lie on monomer i. These contributions are obtained by rotating the symmetrically equivalent contributions from atoms r and s on symmetrically unique monomer z. The two-body term contributes only if atoms q and p lie on monomers i and/or j, and they are equivalent to the symmetrically unique two-body contributions from atoms r and s on monomers z and z'. Once all the elements of the Hessian are computed, the Hessian can be mass-weighted and diagonalized to obtain the vibrational modes.

2.5 Computational Methods

Quantum mechanical calculations were performed using a development version of the Q-Chem software package[115] with dual-basis RI-MP2[116,117,117–119] and the aug-cc-pVDZ basis set (and the corresponding density-fitting and dual basis sets).[120–122] Only valence electrons were correlated in the MP2 calculations, and counterpoise corrections were applied for each dimer calculation to compensate for basis set superposition errors. The molecular mechanics calculations were performed using the Tinker 6.2 software package.[123] and the AMOEBA force field.[75]

In the HMBI fragment calculations, the transition between QM and MM treatments of the pairwise interaction energies was performed between 9–10 Å.[44] Geometry optimizations were performed using the L-BFGS algorithm[124] as implemented in the open-source DL-FIND geometry optimizer[125] using hybrid delocalized internal coordinates. Employing space group symmetry increases the magnitude of individual nuclear gradient elements considerably, since contributions are effectively multiplied by the number of symmetrically equivalent monomers in the unit cell. Therefore, the default convergence criteria for the root-mean-square (3.0×10^{-4} hartrees/bohr) and maximum gradient element sizes (4.5×10^{-4} hartrees/bohr) were scaled up by the ratio Z/Z' in the space-group symmetry optimizations.

Initial experimental crystal structures were obtained from the Cambridge Crystallographic Data Center for formamide(RefCode FORMAM01),[126] acetic acid(RefCode ACETAC01),[127] acetamide (RefCode ACEMID05),[128] and imidazole (RefCode IMAZOL06).[129] See Figure 2.3.

2.6 Results and Discussion

The new symmetry-exploiting algorithm was utilized to optimize the structures, compute lattice energies, and obtain the harmonic phonons at the Γ point for crystals of

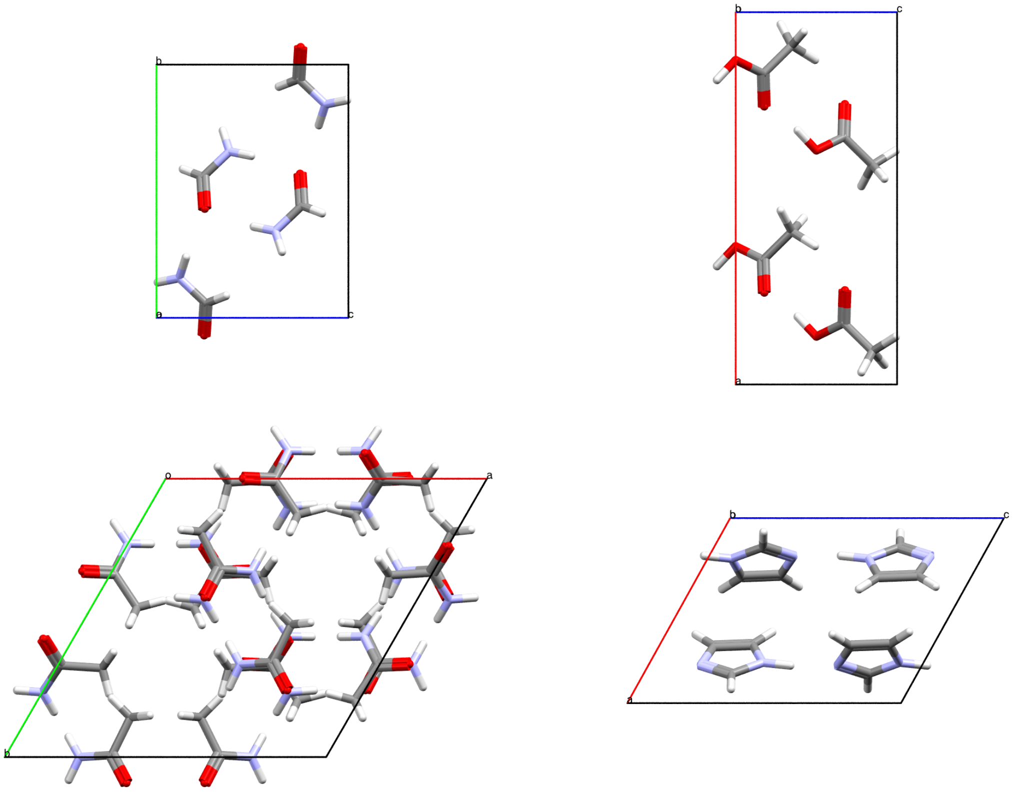

Figure 2.3: Crystals considered here (clockwise from top left): formamide (space group $P2_1/c$), acetic acid ($Pna2_1$), imidazole ($P2_1/c$), and acetamide ($R3c$).

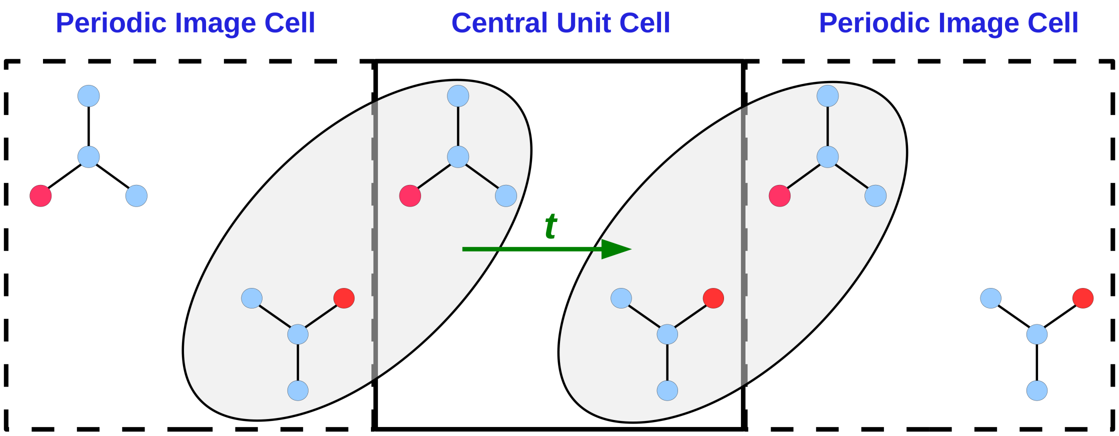

Figure 2.4: Simple translational symmetry reduces the number of necessary dimer calculations involving one molecule in the central unit cell and one periodic image monomer by a factor of two.

formamide, acetic acid, acetamide, and imidazole crystals. These results are compared against those from calculations in which only simple translational symmetry in the dimers arising from the use of periodic boundary conditions is exploited. This provides a nice point of comparison, since an algorithm that exploits translational symmetry in the dimers is trivial to implement, and it reduces the number of dimers involving one molecule in the central unit cell and one periodic image monomer by a factor of two (Figure 2.4).

The calculation of the two-body (dimer) contributions forms the bottleneck in our crystal calculations. Calculation of the one-body and force-field terms are computationally negligible in comparison. Therefore, the discussion here focuses only on the dimer calculations. As expected, exploiting space group symmetry produces appreciable computational savings in computing the lattice energies compared to the translational-symmetry-only algorithm.

For the crystals with four molecules ($Z = 4$) in the unit cell and only one molecule in the asymmetric unit cell ($Z' = 1$), space group symmetry reduces the number of necessary dimer calculations by three- to four-fold. For acetamide ($Z = 18$ and $Z' = 1$), however, the savings are 18-fold. The same reductions in the number of dimers is obtained when computing the Hessian. In all cases, the computational time is reduced by essentially the same factor due to the elimination of the redundant dimer calculations.

The reason for non-integer savings factors Table 2.2 for cases like formamide and imidazole (both $P2_1/c$) merits further discussion. As noted above, any dimer composed of one molecule in the central unit cell and one molecule outside it exhibits one translationally equivalent dimer due to the use of periodic boundary conditions. Each of these dimers can be transformed according to the symmetry operators in the space group. Therefore, the total number of symmetrically equivalent dimers will be up to two times the number of space group operations in the crystal.

Table 2.2: Number of individual dimer calculations required to obtain the crystal energy or harmonic phonon modes.

	Space Group	Z / Z'	w/ Translational Symmetry only	w/ Space Group Symmetry	Savings Factor
Formamide	$P2_1/c$	4 / 1	248	80	3.1
Acetic Acid	$Pna2_1$	4 / 1	212	53	4.0
Acetamide	$R3c$	18 / 1	990	55	18.0
Imidazole	$P2_1/c$	4 / 1	212	66	3.2

In the imidazole crystal with four space group operations in $P2_1/c$, for example, this means each symmetrically unique dimer could occur up to eight times. However, for certain sets of dimers in the crystal, multiple distinct symmetry operations produce the same dimer, as shown in Figure 2.5. For instance, applying either an inversion operator i

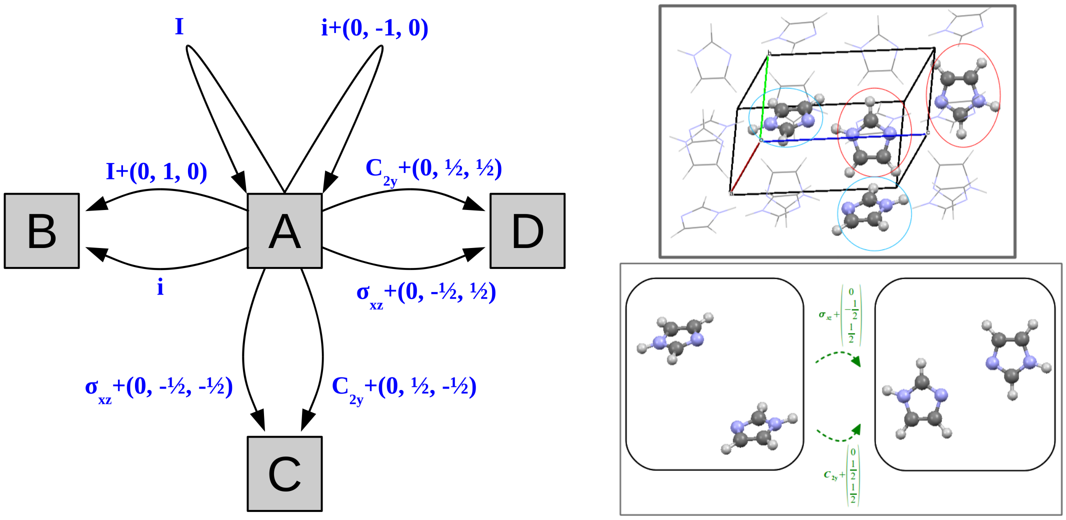

Figure 2.5: A given imidazole dimer can exhibit up to eight-fold degeneracy, with symmetry equivalent structures arising from the four different space group operators and two translationally equivalent versions of each operator. However, for the particular dimer circled in blue, half the operators are redundant, and only four distinct dimers exist in the crystal (labeled A, B, C, and D above). For example, the figure on the right shows that two different operations map from dimer A (blue circles) to dimer D (red circles). Therefore, one can obtain only 4-fold savings instead of 8-fold savings for this set of dimers. Furthermore, since simple translational symmetry alone captures half those savings, exploiting space group symmetry provides only 2-fold additional savings.

or an identity operator I plus a translation of one unit cell length in y to dimer "A" both produce the same dimer "B." This cuts the potential savings from symmetry in half, to four-fold. Translational symmetry alone would capture half of those four-fold savings. For example, reflecting dimer "A" through the xz plan and translating it by (0, -1/2, -1/2) in fractional coordinates produces dimer "C." However, the same reflection combined with a different (0, -1/2, 1/2) translation produces dimer "D." So compared to an algorithm that just exploits translational symmetry, using space group symmetry adds only an additional two-fold savings instead of the four-fold savings seen for many other dimers in the crystal. In practice, then, the non-integer savings factors listed in Table 2.2 result from having some dimers with 4-fold savings while others have only 2-fold savings beyond translational symmetry alone.

While the computational savings provided by space group symmetry in computing crystals energies and Hessians are useful, exploitation of symmetry can play an even larger role in molecular crystal structure optimization by reducing both the cost of each gradient evaluation and by facilitating convergence of the optimization by reducing the dimensionality of the optimization problem, as demonstrated by greater savings on Table 2.3 compared to Table 2.2.

As shown in Tables 2.4 and 2.5, the structures optimized with and without space group symmetry are very similar. The root-mean-square deviations (rmsd_{15}[130]) between the structures obtained with and without space group symmetry range 0.03–0.07 Å, while the differences in energy and root-mean-square deviations in the harmonic vibrational frequen-

Table 2.3: Number of geometry optimization steps and time required to optimize crystal structures with and without space group symmetry on 40 processor cores.

	w/ Translational Symmetry only		*w/ Space Group Symmetry*		Savings
	# of Steps	Days	# of Steps	Days	Factor
Formamide	60	1.8	44	0.4	4.2
Acetic Acid	59	3.9	34	0.6	6.9
Acetamide	166[a]	57.3[a]	52	1.0	57.5
Imidazole	128	11.6	62	1.7	6.7

[a] Optimization aborted due to slow convergence. The resulting structure lies 2 kJ/mol per monomer higher in energy than the converged structure optimized with symmetry.

cies are small (Table 2.5). Note that the implementation of the gradients and Hessian terms have been validated via finite difference. The small variations in the structures optimized with and without space group symmetry arise from minor differences in the optimization procedure and from the flatness of the potential energy surfaces. In particular, when space group symmetry is employed, the monomers are translated at each optimization step to preserve symmetry upon changes in the lattice parameters. This translation is unnecessary in the symmetry-free algorithm. This means that the two optimizations take slightly different routes toward the potential energy minimum. Furthermore, reduction in length of the gradient with space group symmetry magnifies individual gradient elements, which affects the optimization step size and the point at which the root-mean-square gradient change meets the convergence criterion. Finally, the crystal energy surfaces tend to be very flat, making it possible to converge to slightly different structures with very similar energies, as in the cases of acetic acid and imidazole.

For acetamide, the optimization with translational symmetry converged extremely slowly and was aborted after 166 steps and almost two months on 40 processor cores. The resulting structure lies ~2 kJ/mol above that of the converged structure obtained with space group symmetry.

Table 2.4: Comparison of crystals structures optimized at the dual-basis RI-MP2/aug-cc-pVDZ level with and without space group symmetry. Low-temperature experimental lattice parameters are also provided for reference.

Crystal	Symmetry	a (Å)	b (Å)	c(Å)	α (°)	β (°)	γ (°)
Formamide	Translational	3.596	9.315	8.556	90.39	125.59	89.99
$P2_1/c$	Space Group	3.555	9.220	8.471	90	124.53	90
	Expt @ 108 K^a	3.613	9.053	8.419	90	125.39	90
Acetic Acid	Translational	13.340	3.909	5.898	89.64	90.03	90.19
$Pna2_1$	Space Group	13.384	3.907	5.903	90	90	90
	Expt @ 40 K^b	13.151	3.923	5.762	90	90	90
Acetamide	Translational	11.647	11.642	12.936	90.29	89.77	119.70
$R3c$	Space Group	11.676	11.676	12.957	90	90	120
	Expt @ 23 K^c	11.492	11.492	12.892	90	90	120
Imidazole	Translational	7.362	5.296	9.821	89.84	120.03	89.98
$P2_1/c$	Space Group	7.453	5.272	9.832	90	121.00	90
	Expt @ 103 K^d	7.569	5.366	9.785	90	119.08	90

[a] Ref 126 [b] Ref 127 [c] Ref 128 [d] Ref 129

Table 2.5: Difference between the lattice energies ($\Delta E = E_{trans} - E_{space}$), optimized structures (in rmsd$_{15}$ values for all atoms), and vibrational frequencies (rmsd) for the structures optimized with space group symmetry relative to those using translational symmetry only.

Crystal	ΔE (kJ/mol)	Structure rmsd$_{15}$ (Å)	Vib. Freq. rmsd (cm^{-1})
Formamide	-0.8	0.067	6.4
Acetic Acid	0.2	0.032	1.0
Acetamide	1.9	0.037	[a]
Imidazole	0.1	0.052	1.9

[a] Vibrational frequencies for acetamide were not computed without symmetry because the structure optimization failed to converge.

For formamide, the combination of MP2 and the AMOEBA force field appears to prefer a slightly symmetry broken structure, leading to a larger discrepancy between the structures obtained with and without symmetry. In particular, it appears that the the AMOEBA force field used for the long-range and many-body terms appears to have some reliability issues. A complete crystal structure relaxation starting from the experimental structure converges to a very different structure than the experimental one. However, partially relaxing the atomic positions in a fixed unit cell for a 5–10 cycles before relaxing the cell parameters produces the experimentally observed structures. At the HMBI MP2/aug-cc-pVDZ + AMOEBA level, the experimental-like structures lie 3-4 kJ/mol above the seemingly artifactual one. For the comparison structure optimization without space group symmetry, we also observe that the AMOEBA force field contributions lead to a 0.8 kJ/mol preference for a symmetry-broken version of the experimental structure over the symmetry-preserving one. However, single-point energy calculations on the optimized geometries using the more reliable ab initio force field (AIFF)[45] instead of AMOEBA for the MM terms reverse the ordering, making the correct structure optimized with symmetry the most stable structure by more than 1 kJ/mol. In any case, these issues do not affect the performance of the space group symmetry algorithm, which is the primary focus here.

2.7 Conclusions

A straightforward algorithm for implementing space group symmetry in fragment-based electronic structure calculations has been described. This algorithm does not require one to specify the symmetry operations in advance. Rather, symmetry equivalence is determined by orienting each monomer or dimer fragment in a common reference frame and comparing their structures. The space group operators relating symmetrically equivalent fragments can be extracted readily.

The space group symmetry algorithm has been implemented for energy, nuclear gradient, and harmonic phonon frequency calculations. Exploiting space group symmetry typically accelerates the evaluation of all of these quantities by a factor of several times or more, depending on the space group. The reduction in the degrees of freedom of the symmetry-preserving gradient often provides additional computational savings during a structure optimization.

Chapter 3

Predicting finite-temperature properties of crystalline carbon dioxide from first principles with quantitative accuracy

3.1 Outline

In the previous work, no attempt has been made to incorperate the effects of temperature or thermal expansion into the HMBI model. In fact, most electronic structure predictions neglect temperature and/or thermal expansion. Molecular crystal structures, thermodynamics, and mechanical properties can vary substantially with temperature, and predicting these temperature-dependencies correctly is important for many practical applications in the pharmaceutical industry and other fields. Neglecting temperature could potentially lead to erroneous results. Here, we demonstrate that by combining HMBI, using large basis set MP2 and even CCSD(T), with a quasi-harmonic treatment of thermal expansion, experimentally observable properties such as the unit cell volume, heat capacity, enthalpy, entropy, sublimation point and bulk modulus of phase I crystalline carbon dioxide can be predicted in excellent agreement with experiment over a broad range of temperatures. These results point toward a promising future for *ab initio* prediction of molecular crystal properties at real-world temperatures and pressures. Review Section 1.4.2 for an overview of the quasi-harmonic approximation.

3.2 Introduction

Moving beyond 0 K to predict molecular crystal structures and properties at finite temperatures represents the next frontier in modeling organic materials. Instead of predicting a multitude of potential crystal polymorphs at 0 K, can we tell a pharmaceutical chemist if the desired drug formulation is thermodynamically favored at room temperature? Or can we predict the polymorphic phase diagram over a broad range of temperatures and pressures? The unexpected appearance of a new, more stable polymorph of a drug can have dire consequences for its stability or bioavailability, sometimes even forcing the drug's removal from the market until a new formulation can be developed.[1–3]

When manufacturing a drug tablet, the rapid decompression which occurs immediately after compaction of the drug powder can lead to a variety of defects or even catastrophic failure of the tablet.[131] Mechanical properties like the Young's modulus and the bulk modulus of a molecular crystal provide important insights into the bulk behavior of pharmaceutical powders during the tabletting process.[95,132,133] Given that such mechanical properties can exhibit sizable temperature dependence, can we predict them at the relevant temperatures?

Crystal structure prediction has undergone rapid advances over the past several decades, with a number of successful predictions in the Blind tests of Crystal Structure Prediction,[7–12] improved optimization algorithms for identifying stable crystal packing motifs,[134–142] and major advances in dispersion-corrected density functional theory (DFT)[20,26,29,31,32,38,42,43,69] and fragment-based electronic structure methods using the many-body expansion[15–18,21,33,44,45,50,51,77,88,104,143,144] that enable the routine application of high-accuracy quantum mechanical methods to organic crystals. Molecular crystal lattice energies can now be predicted to within sub-kJ/mol accuracy in certain cases,[77] or within a few kJ/mol more routinely.[20,31,37,38,43,45]

The next generation of molecular crystal modeling needs to move beyond 0 K lattice energies and structures, and predict crystal structures and properties at the finite temperatures and pressures where most real-world experimental applications occur. Progress in this direction has already been made. For example, DFT studies of high-pressure molecular crystal phases have become routine, and Hirata and co-workers have recently predicted the phase boundary for phase I and phase III carbon dioxide with second-order Møller-Plesset perturbation theory (MP2).[18,50,51] They have also used similar calculations to simulate various properties and spectroscopic features in ice,[22,145] carbon dioxide,[21,146] and other systems.[18] Reilly and Tkatchenko used harmonic free energy estimates with many-body dispersion-corrected density functional theory to rationalize the experimental preference for form I aspirin over form II.[14] However, most such studies, particularly those based on *ab initio* methods beyond DFT, rely on structures optimized without consideration of temperature. The volume of a molecular crystal unit cell often expands by several percent between 0 K and room temperature, with substantial impacts on many crystal properties.

Capturing these finite temperature effects is challenging. Molecular dynamics and Monte Carlo simulations provide a conceptually straightforward means of accessing these finite-temperature properties that has proved effective for studying organic crystal free energies/phase diagrams,[139,140,147–150] and nucleation/growth[151–154] at the force field level. However, achieving the requisite accuracy in larger, non-rigid molecules with force fields remains a major challenge. On the other hand, the comparatively high computational cost of more accurate electronic structure methods makes extensive configurational sampling infeasible in most cases.

Instead, we demonstrate here that coupling large-basis second-order MP2 and coupled cluster singles, doubles and perturbative triples (CCSD(T)) electronic structure calculations with the quasi-harmonic approximation enables one to predict a wide variety of properties of crystalline carbon dioxide (phase I) with unprecedented accuracy. The quasi-harmonic approximation has a long-history in materials modeling, but to our knowledge, this study represents the first time it has been combined with electronic structure calculations that approach the *ab initio* limit for molecular crystals.

Carbon dioxide is much smaller than typical organic compounds, of course. It also exhibits weaker many-body interactions than many larger and/or polar molecules.

Nevertheless, it makes an excellent test case for several reasons: a wealth of experimental data exists against which the predictions can be tested, its small molecular size makes it feasible to assess the accuracy that can be obtained with calculations which approach the *ab initio* limit, and it has also been the subject of many earlier DFT [34–36,155] and smaller-basis MP2 studies. [21,50,51,146]

We show that extrapolated complete basis set MP2 and CCSD(T) calculations predict the crystal volume within 2%, the heat capacity within $0.2R$ ($< 5\%$ for $T = 50$–190 K), the sublimation enthalpy within 1.5 kJ/mol, and the sublimation entropy within 2 J/mol K (2%), all over a temperature range spanning 200 K. CCSD(T) predicts the sublimation point of dry ice (194.7 K) to within 6 K. In contrast to previous difficulties in modeling the bulk modulus of crystalline CO_2, [21] we predict both its magnitude and temperature dependence in excellent agreement with experiment. Overall, the ability to achieve quantitative accuracy for a broad spectrum of molecular crystal properties in phase I carbon dioxide provides much cause for optimism in the future extension of finite-temperature predictions to larger, more chemically interesting species.

3.3 Theory

To determine the structure of phase I carbon dioxide at a given temperature T and pressure P, one minimizes the Gibbs free energy G with respect to both the atomic positions in the unit cell and the unit cell parameters (a, b, c, α, β, γ).

$$G(T, P) = U_{el} + PV + F_{vib}(T) \tag{3.1}$$

where U_{el} is the internal electronic energy, PV is the pressure-volume contribution, and F_{vib} represents the Helmholtz vibrational free energy contribution. This latter contribution is expressed in terms of the harmonic phonons $\omega_{k,i}$ evaluated at multiple **k**-points in reciprocal space:

$$F_{vib}(T) = \frac{N_A}{n} \sum_{k} \sum_{i} \left(\frac{\hbar\omega_{k,i}}{2} + k_b T \ln \left[1 - exp\left(-\frac{\hbar\omega_{k,i}}{k_b T} \right) \right] \right) \tag{3.2}$$

where n is the number of unit cells in the supercell approximation and N_A is Avogadro's number.

Both U_{el} and F_{vib} are computed using the fragment-based hybrid many-body interaction (HMBI) model,[44,45,86,87] which allows one to perform high-level MP2 and even coupled cluster calculations on periodic systems like molecular crystals with reasonable computational cost. The phonons at a given **k**-point are evaluated by constructing and then diagonalizing the mass-weighted dynamical matrix using the supercell approach,

$$D_{\alpha,\beta}\left(l, l'; \mathbf{k}\right) = \frac{1}{\sqrt{M_l M_{l'}}} \sum_{\kappa} \frac{\partial^2 U_{El}}{\partial R_\alpha(0) \partial R_\beta(\kappa)} exp\left(-2\pi i \mathbf{k} \cdot \delta\mathbf{R}_{l,l'}(0,\kappa)\right) \tag{3.3}$$

where $\frac{\partial^2 U_{El}}{\partial R_\alpha(0) \partial R_\beta(\kappa)}$ are individual elements in the supercell Hessian[92,156] involving the R_α coordinate of atom l in the central unit cell (index 0) and the R_β coordinate of atom l' in the unit cell with index κ, and $\delta\mathbf{R}_{l,l'}(0,\kappa)$ is the displacement vector between these two atoms. The M's are the atomic masses. The exponential introduces the phase shift in the harmonic motions of the periodic image atoms relative to the atoms in the central unit cell.

Unlike conventional periodic boundary condition models, the fragment approach used in HMBI enables these equations to be evaluated for the supercell Hessian with minimal additional effort beyond the normal unit cell Hessian.[17,156] All the necessary supercell QM contributions to the force constants in Eq 3.3 can be constructed from contributions already available from the standard unit cell Hessian. The only new contribution required is the full supercell Hessian at the MM level. This allows one to capture phonon dispersion by using a large supercell and sampling many **k**-points with comparatively low additional computation cost.

The treatment of F_{vib} is potentially very computational demanding, since it would normally require many cycles of geometry optimization and a phonon calculation for each update in the unit cell parameters. Instead, we approximate the phonons for a given crystal volume using the quasi-harmonic approximation (QHA). The QHA relates the i-th phonon frequency ω_i at a given volume V to a reference frequency $\omega_{i,ref}$ obtained at some reference volume V_{ref} via the Grüneisen parameter for that phonon mode γ_i,

$$\omega_i = \omega_{i,ref} \left(\frac{V}{V_{ref}} \right)^{-\gamma_i} \tag{3.4}$$

where

$$\gamma_i = - \left(\frac{\partial \ln \omega_i}{\partial \ln V} \right) \tag{3.5}$$

Overall, then, a reference crystal structure, unit cell volume, and phonon frequencies are obtained at zero temperature and pressure by minimizing U_{el}, and the harmonic phonon modes are computed using lattice dynamics. Two additional rounds of geometry optimization and phonon calculation are performed with fixed unit cell parameters at cell volumes which are slightly larger and smaller than V_{ref}. Here, the cell volume was expanded/contracted by 10 Å^3, or roughly 5%. The Grüneisen parameter for each of the 3N vibrational modes is computed via finite difference using a linear model.[31] The phonon frequencies at the different volumes were matched via maximum overlap of the phonon eigenmodes.

With the reference frequencies, reference volume, and Grüneisen parameters obtained from these three optimizations and lattice dynamics phonon calculations, one can evaluate F_{vib} at any given temperature and unit cell volume. That allows one to find the crystal structure at any temperature and pressure by minimizing the Gibbs free energy (Eq 3.1).

The sublimation enthalpy and entropy were computed as the difference between the value found for gas and 1/4 the value found for the crystal (because there are 4 molecules in the unit cell).

$$\Delta H_{sub} = H_{gas} - \frac{1}{4} H_{crystal} \tag{3.6}$$

$$\Delta S_{sub} = S_{gas} - \frac{1}{4} S_{vib,crystal} \tag{3.7}$$

Once the crystal structures were obtained as a function of temperature and pressure, its enthalpy was computed as the sum of the electronic energy of the crystal, $U_{el,crystal}$, plus a pressure-volume term PV and a vibrational internal energy contribution, $E_{vib,crystal}$.

$$H_{crystal} = U_{el,crystal} + PV + E_{vib,crystal} \tag{3.8}$$

The vibrational internal energy was determined using standard harmonic oscillator expression.

$$E_{vib,crystal} = \frac{N_A}{n} \sum_k^n \sum_i \left(\frac{\hbar\omega_{k,i}}{2} + \frac{\hbar\omega_{k,i}}{exp\left(\frac{\hbar\omega_{k,i}}{k_b T}\right) - 1} \right) \tag{3.9}$$

Since phase I carbon dioxide is an ordered crystal, its configurational entropy can be neglected. Its entropy can be computed from the standard vibrational expression alone.

$$S_{vib,crystal} = \frac{N_A}{n} \sum_k^n \sum_i \left(\frac{\hbar\omega_{k,i}}{T\left(exp\left(\frac{\hbar\omega_{k,i}}{k_b T}\right) - 1\right)} - k_b \ln\left[1 - exp\left(-\frac{\hbar\omega_{k,i}}{k_b T}\right)\right] \right) \tag{3.10}$$

The gas phase was modeled using standard ideal gas partition functions. Of course, the gas may exhibit deviations from ideality at the low temperatures considered here, but they are hopefully not too large. The enthalpy of the gas is given as the sum of electronic ($U_{el,gas}$), translational ($3/2RT$), rotational (RT), vibrational (E_{vib}) terms plus an additional factor of RT from the PV term.

$$H_{gas} = U_{el,gas} + \frac{3}{2}RT + RT + E_{vib} + RT \tag{3.11}$$

where E_{vib} is defined as:

$$E_{vib,gas} = N_A \sum_i^4 \left(\frac{\hbar\omega_i}{2} + \frac{\hbar\omega_i}{exp\left(\frac{\hbar\omega_i}{k_b T}\right) - 1} \right) \tag{3.12}$$

The entropy of the gas were computed as a sum of standard rotation, translation and, vibrational partition functions.

$$S_{gas} = S_{trans,gas} + S_{rot,gas} + S_{vib,gas} \tag{3.13}$$

where the terms are defined as:

$$S_{trans,gas} = R\ln\left[\left(\frac{2\pi m k_b T}{h^2}\right)^{3/2}\left(\frac{k_b T e^{5/2}}{p}\right)\right] \tag{3.14}$$

$$S_{rot,gas} = R\ln\left(\frac{8\pi^2 eTIk_b}{\sigma h^2}\right) \tag{3.15}$$

$$S_{vib,gas} = N_A\sum_{i}^{4}\left(\frac{\hbar\omega_i}{T\left(exp\left(\frac{\hbar\omega_i}{k_bT}\right)-1\right)} - k_b\ln\left[1-exp\left(-\frac{\hbar\omega_i}{k_bT}\right)\right]\right) \tag{3.16}$$

Note that the symmetry factor σ in S_{rot} is 2 for carbon dioxide.

3.4 Methods

Crystal energies were computed using HMBI. (see EQ 1.90) All QM contributions were calculated with either density-fitted MP2[116,117,157,158] or CCSD(T)[159,160] in the Dunning aug-cc-pVXZ basis sets (abbreviated as aXZ here)[120,121] using Molpro 2012.[161,162] A counterpoise correction for basis set superposition error[163] was employed for each two-body dimer calculation. The energies, gradients, and Hessian elements were all extrapolated to the complete basis set (CBS) limit using a two-point TQ extrapolation of both the Hartree-Fock[164] and correlation energy contributions.[165] Energies and gradients at the CCSD(T)/CBS limit were estimated by correcting the MP2/CBS limit values with the difference between CCSD(T) and MP2, ΔCCSD(T) $\approx$ CCSD(T) - MP2, computed in the aug-cc-pVDZ basis set. MP2 phonons were used to evaluate F_{vib} in the CCSD(T) calculations. The MM contributions were computed using the Amoeba force field and Tinker 6.3.[166] Intermolecular force field parameters for CO_2 were generated using Poltype version

1.1.3.[167]

The relative rigidity and lack of many-body polarization effects makes carbon dioxide a good candidate for simple, fixed charge force field models, though the importance of many-body dispersion effects has been noted.[168] For comparison with the HMBI electronic structure results, the predictions here were repeated using the empirical CO_2 potential of Cygan and co-workers.[169] This flexible, three-point model includes standard harmonic stretch and bend terms, point-charge electrostatics, and Lennard-Jones dispersion/repulsion terms. It was particularly parameterized to reproduce vibrational spectra, which should help it capture the phonon contributions. The carbon dioxide quadrupole moment also proves important for modeling its solid state,[170] and the point charges in this force field generate a molecular quadrupole of -4.22 D·Å, in good agreement with the experimental value of -4.27±0.18 D·Å.[171] Additional test calculations with the TraPPE force field,[172] which uses the same functional form but slightly different empirical parameters, produced similar results (not presented here). Of course, many other CO_2 potentials exist, and a more elaborate or physical potential (e.g. Ref 173) might.

Substantial computational savings were obtained by exploiting the $Pa\bar{3}$ space group symmetry of phase I CO_2 throughout.[92] Symmetry reduces the number of two-body dimer calculations required from ∼100 to 5–9 (depending on the pressure). As discussed in Chapter 2, symmetry also reduces the number of degrees of freedom that must be optimized to only the atomic coordinates present in the symmetrically unique monomer(s). The symmetry algrithm in Chapter 2 deminstrates how the optimization's the degrees of freedom of a crystal could be reduced to the coordinates of atoms on the symmetrical unique monomers. For crystals with molecules that belong high symmetry point groups, such as phase I CO_2, the number of degrees of freedom can be reduced further. These special cases are outlined in the Appendix A. For carbon dioxide the number of degrees of freedom in the geometry optimization were reduced from 42 to two: the lattice constant a and the C=O bond length.

3.5 Results and Discussion

The next sections compare the predicted and experimental values for thermal expansion, thermodynamic properties, and the bulk modulus. All predicted values plotted in Figures here.

3.5.1 Thermal expansion

To begin, we predict the thermal expansion of the CO_2 lattice at atmospheric pressure by optimizing the quasi-harmonic Gibbs free energy at a series of different temperatures. At 1 atm, the PV term only contributes ∼0.01 kJ/mol to the overall energy, so it was neglected here. Figure 4.2 compares these predictions against experimental results from Manzhelii et al,[174] Krupskii et al,[57] and the low-temperature fit (20–114 K) of Keesom and Köhler.[175,176] In a small aug-cc-pVDZ basis set, MP2 substantially underbinds the crystal, leading to a substantial over-estimation of the unit cell volume. As we approach the complete-basis-set (CBS) limit, however the MP2 prediction improves dramatically, with MP2/CBS underestimating the cell volume by only 2–3%. Fortuitously, the slightly smaller aug-cc-pVQZ basis performs even better, with predicted volumes lying within ∼0.5% of

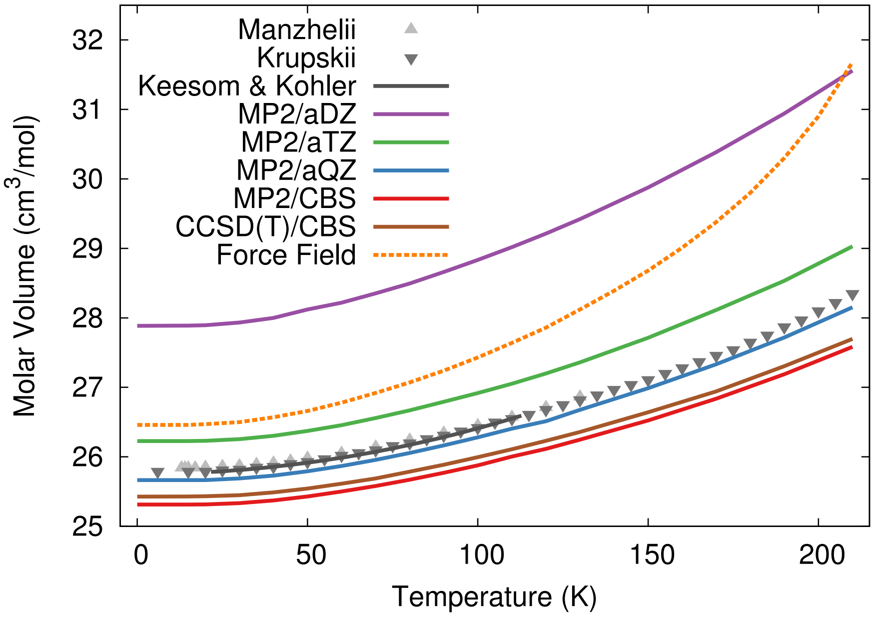

Figure 3.1: Predicted thermal expansion of the CO_2(s) unit cell compared to the experimental values[57,174–176] in gray.

experiment.

The errors are nearly constant across the entire temperature range. For instance, MP2/CBS underestimates the volume by 0.5 cm^3/mol (2%) at low temperatures, and this error increases to only 0.7 cm^3/mol (3%) at the sublimation point (194.7 K). Most of the error is present already in the lowest temperature results, which suggests it largely stems from the underlying fragment-based electronic structure treatment, rather than from the quasi-harmonic approximation. The treatment of phonon dispersion via lattice dynamics is also important here. Using Γ-point frequencies only causes the model to underestimate the rate of thermal expansion noticeably as shown in Figure 3.2.

One might hope to obtain further improvements by moving beyond second-order perturbation theory to the CCSD(T) level. However, previous work indicates that correla-

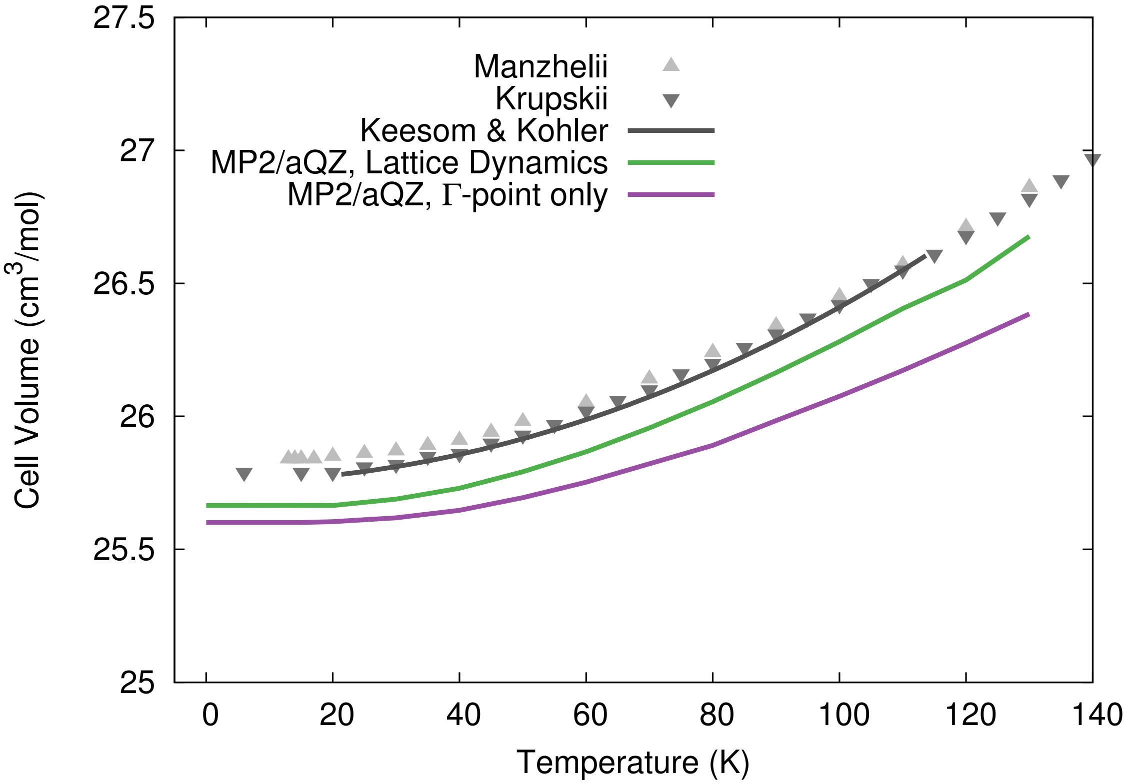

Figure 3.2: Comparison of the MP2/aQZ thermal expansion with Γ-point only phonons versus including phonon dispersion via lattice dynamics.

tion energy contributions beyond second-order perturbation theory are small in crystalline CO_2, with the lattice energy shifting by only ~0.3 kJ/mol between MP2 and CCSD(T).[45] Here, refining the thermal expansion predictions at the CCSD(T)/CBS level (with the free energy computed as the sum of CCSD(T) internal energies and MP2 vibrational free energy contributions) reduces the errors by only 0.1 cm^3/mol. Nevertheless, these results show that large-basis electronic structure calculations plus the quasi-harmonic approximation model the temperature dependence of the carbon dioxide unit cell volume very reliably all the way up to the sublimation point.

For comparison, the force field potential performs quite well at low temperature,

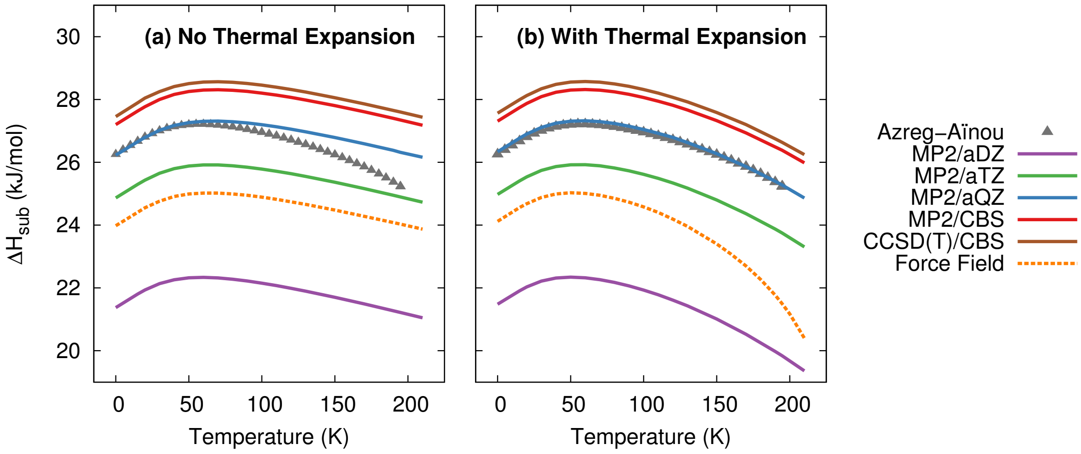

Figure 3.3: Predicted enthalpies of sublimation at 1 atm (a) neglecting thermal expansion and (b) with quasiharmonic thermal expansion, relative to the empirical data of Azreg-Aïnou.[177]

predicting a cell volume that is roughly on par with the MP2/aug-cc-pVTZ calculation with orders of magnitude lower computational cost. However, as the temperature increases, the force field model expands the crystal volume much too rapidly.

3.5.2 Thermodynamic properties

Given the excellent treatment of thermal expansion, we next investigate the model's ability to predict thermodynamic properties such as the heat capacity and the enthalpies and entropies of sublimation. Such properties are critical to determining polymorph stability at finite temperatures. For each of these properties, predictions were made with and without the thermal expansion provided by the quasi-harmonic approximation.

Figure 3.3 plots the enthalpy of sublimation at 1 atm relative to the experimentally-derived ΔH_{sub} determined by Azreg-Aïnou.[177] Azreg-Aïnou derived these values using fits to the experimentally observed heat capacity and vapor pressure data, ideal gas partition functions, various small corrections for gas imperfection, and other details.

The sublimation enthalpy is dominated by the crystal lattice energy. The zero-point and thermal enthalpy corrections account for only ~10% (at low temperature) to ~25% or more (at the sublimation point) of the total sublimation enthalpy. Accordingly, the sublimation enthalpy should behave similarly to the lattice energy with regard to the basis set: Small-basis MP2 underestimates the CO_2 lattice energy significantly, but using large basis sets mostly corrects this error.[44,45] As expected, small basis sets predict a sublimation enthalpy that is too small, while MP2/aug-cc-pVQZ fortuitously predicts a sublimation enthalpy in almost perfect agreement with experiment. Extrapolating to the complete-basis-set limit produces a sublimation enthalpy that overestimates the experimental value by only 1.0–1.1 kJ/mol. CCSD(T)/CBS binds crystalline CO_2 slightly more,[45] which increases the sublimation enthalpy further, to a value 1.3–1.4 kJ/mol too large. This accuracy is near the limit of what is achievable with modern electronic structure theory. Errors in the lattice energy of 1–2 kJ/mol represent a best-case scenario for practical molecular crystal calculations,[45,77] while errors of several kJ/mol are more typical.[20,31,37,38,43]

Figure 3.3 also highlights how the approximate treatment of anharmonicity and thermal expansion via the quasi-harmonic approximation proves essential to capturing the proper temperature dependence above 50 K. Without the quasi-harmonic approximation, the theoretical calculations substantially overestimate the sublimation enthalpy at higher temperatures. When the quasi-harmonic approximation is employed, however, the calculations obtain the correct curvature across a 200 K temperature range. Both the MP2/CBS and CCSD(T)/CBS results predict the maximum in the sublimation enthalpy at 59 K, in

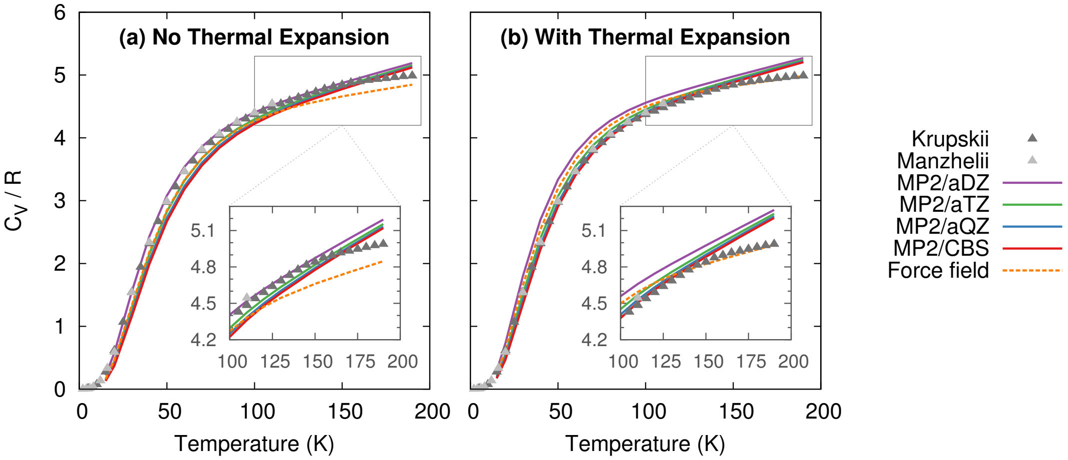

Figure 3.4: Predicted isochoric heat capacity (a) neglecting thermal expansion and (b) with quasi-harmonic thermal expansion relative to the experimental data of Krupskii et al[57] and Manzhelii et al.[174]

excellent agreement with the 58.829 K reported by Azreg-Aïnou.[177]

Once again, the force field model used here performs almost as well as the MP2/aug-cc-pVTZ results at low temperatures or when thermal expansion is neglected. However, the exaggerated thermal expansion seen in Figure 4.2 is reflected in poor prediction of the sublimation enthalpy at warmer temperatures.

Given the high accuracy of the MP2 and CCSD(T) sublimation enthalpy predictions as a function of temperature, it is not surprising that the isochoric heat capacity, C_V, is also predicted reliably (Figure 3.4). The isochoric heat capacity of the crystal was computed using the standard harmonic oscillator expression,

$$C_v = \frac{R}{n}\sum_{k}^{n}\sum_{i}\left[\left(\frac{\hbar\omega_{k,i}}{k_bT\left(exp\left(\frac{\hbar\omega_{k,i}}{k_bT}\right)-1\right)}\right)^2 exp\left(\frac{\hbar\omega_{k,i}}{k_bT}\right)\right] \tag{3.17}$$

Note that CCSD(T) results are not provided because CCSD(T) phonons are unavailable.

For the heat capacity, all models perform fairly well. Neglecting thermal expansion fortuitously causes small aug-cc-pVDZ basis MP2 to out-performs what should be the more accurate large basis calculations relative to the experimental values of Krupskii et al[57] and Manzhelii et al.[174] When thermal expansion is included, on the other hand, the accuracy of the predictions does improve with increasing basis set, as one generally expects.

Similar to previously published small-basis MP2 results,[21] we find that MP2 underestimates the heat capacity slightly at low temperature. Errors of 1–1.5 J/mol K (0.1–0.2R) are observed below 50 K. However, the results here perform better than the earlier MP2 ones at moderate temperatures (e.g. ∼50–150 K), with errors typically well below 1 J/mol K (0.1R) in the range 50–150 K. At higher temperatures, the predictions begin to deviate more noticeably from the experimental data, probably due to increased anharmonicity in the phonons. This suggests that one might expect larger deviations from the correct temperature-dependence of the sublimation enthalpy at higher temperatures. Nevertheless, on the whole, MP2 predicts the heat capacity accurately across a fairly wide temperature range. For comparison, the force field model behaves similarly to MP2/aug-cc-pVDZ and aug-cc-pVTZ at low and intermediate temperatures, but it asymptotes more quickly than the MP2 heat capacities at higher temperatures. This actually leads to a slightly better prediction of the heat capacity near 200 K when thermal expansion is included, though the result is somewhat fortuitous, given the problems seen earlier in the volume and sublimation enthalpy.

Entropy also plays a critical role in phase stability. The entropy of sublimation at the sublimation point (T=194.7 K) is well-known,[178] but we are not aware of any ex-

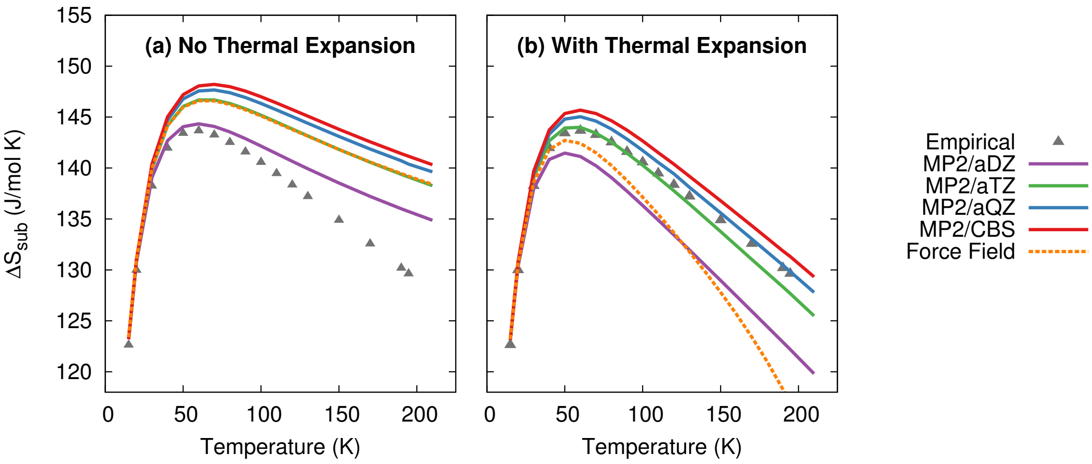

Figure 3.5: Predicted entropies of sublimation at 1 atm (a) neglecting thermal expansion and (b) with quasi-harmonic thermal expansion, relative to the data empirically derived from experiment.

isting tabulation of the experimental sublimation entropy as a function of temperature. Accordingly, we derived an empirical sublimation entropy from existing experimental data according to:

$$\begin{aligned} \Delta S_{sub}^{emp}(T) &= \Delta S_{sub}^{expt}(194.7K) - \int_{194.7K}^{T} \frac{C_{p,crystal}^{expt}(T)}{T} dT \\ &\quad + (S_{gas}(T) - S_{gas}(194.7K)) \end{aligned} \tag{3.18}$$

This expression relates the sublimation entropy at a given temperature to the experimental value at 194.7 K plus corrections for how the entropies of the crystal and the gas change as a function of temperature. The changes in the entropy of the crystal were computed via integration of the experimental isobaric heat capacities,[178] while the gas contributions were evaluated using ideal gas partition functions and the experimentally determined rotational constant[179] and vibrational frequencies.[180] The highest temperature for which the experi-

Table 3.1: Predicted sublimation temperatures T_{sub} at 1 atm, and the corresponding enthalpies and entropies of sublimation at the experimental sublimation point of 194.7 K

	With Thermal Expansion		
	T_{sub} (K)	ΔH_{sub}(194.7 K) (kJ/mol)	ΔS_{sub}(194.7 K) (J/mol K)
Force Field	183.4	21.5	116.9
MP2/aug-cc-pVDZ	163.6	19.8	122.2
MP2/aug-cc-pVTZ	185.4	23.7	127.7
MP2/aug-cc-pVQZ	193.7	25.3	129.8
MP2/CBS	199.5	26.4	131.3
CCSD(T)/CBS[a]	201.4	26.6	131.3
Giauque and Egan[178]	194.7	25.2	129.6
	No Thermal Expansion		
	T_{sub}	ΔH_{sub}(194.7 K)	ΔS_{sub}(194.7 K)
Force Field	172.9	24.0	139.2
MP2/aug-cc-pVDZ	157.1	21.2	135.7
MP2/aug-cc-pVTZ	179.0	24.9	139.1
MP2/aug-cc-pVQZ	187.4	26.3	140.3
MP2/CBS	193.4	27.3	141.2
CCSD(T)/CBS[a]	195.2	27.6	141.1
Giauque and Egan[178]	194.7	25.2	129.6

[a] Using MP2/CBS frequencies and thermal contributions.

mental isobaric heat capacities has been reported is 189.78 K. The isobaric heat capacities at 194.7 K were linearly extrapolated to 55.51 J/mol K from the five highest temperature values.

As shown in Figure 3.5, the quasi-harmonic treatment of thermal expansion proves critical to obtaining the correct temperature dependence of the entropy. Without thermal expansion, MP2/CBS overestimates the sublimation entropy above 50 K by up to 9%. In contrast, including thermal expansion dramatically reduces the errors, predicting the sublimation entropy to within 1–2% throughout the 200 K temperature range. For comparison, without thermal expansion, the force field mimics MP2/aug-cc-pVTZ. However, once thermal expansion is included, the force field predicts an entropy of sublimation that decreases much too rapidly at higher temperatures.

Finally, the sublimation point can be predicted by combining the enthalpy and entropy of sublimation to determine the temperature where $\Delta G_{sub} = 0$. As shown in Table 4.3, small aug-cc-pVDZ basis MP2 calculations underestimate the sublimation temperature by 30 K. Increasing the basis set, however, allows one to predict the experimental sublimation temperature of 194.7 K within 5 K (MP2/CBS) or 7 K (CCSD(T)/CBS) when thermal expansion is included. The CCSD(T) enthalpy and entropy of sublimation at 194.7 K are predicted to within 1.2 kJ/mol (4%) and 1.7 J/mol K (1%), respectively.

If one neglects thermal expansion, CCSD(T)/CBS predicts a sublimation temperature of 195.2 K, which agrees almost perfectly with the experimental temperature. However, this accuracy results from fortuitous error cancellation—the ΔH_{sub} and ΔS_{sub} values at 194.7 K are both 9–10% too large. The force field predicts sublimation temperature of 172.9 K without thermal expansion, or 183.4 K with thermal expansion. As before, these values are similar to those obtained from MP2/aug-cc-pVTZ. One should note, however, that in the case where thermal expansion is included, the enthalpy and entropy of sublimation are both underestimated considerably to produce the relatively good estimate for the sublimation temperature.

Once again, these sublimation point predictions reiterate the importance of modeling thermal expansion. More importantly, they hint toward a future where high-quality *ab initio* prediction of phase diagrams as a function of both temperature and pressure may be routine.

3.5.3 Bulk Modulus

Mechanical properties like the bulk modulus are also of considerable interest for many applications. To obtain the bulk modulus, one typically measures the crystal volume as a function of pressure, and then fits the resulting data to an equation of state, treating the isothermal bulk modulus at zero pressure (B_0), its first pressure derivative (B_0'), and the unit cell volume at zero pressure (V_0) as adjustable parameters. Many equations of state exist, including the third-order Birch-Murnaghan[181] and Vinet[182] equations. Non-linear least squares fits to these equations of state can be problematic, with the resulting fit parameters being ill-constrained (i.e. a wide range of parameters produce comparably good fits) and highly correlated.[183,184] The resulting parameters depend strongly on the reference volume at zero pressure (V_0), especially when using the Birch-Murnaghan equation of state.[183] This challenge is particularly acute at room temperature, where crystalline carbon dioxide does not exist at zero pressure, and V_0 must be obtained via extrapolation from finite-pressure volumes. Hence, considerable uncertainty surrounds the experimental bulk modulus parameters for CO_2.[57,174,183–189]

Theory can predict the pressure-volume data at a given temperature to fit the equation of state, and it can predict the zero-pressure unit cell volume V_0 via direct geometry optimization. This latter feature enables one to validate the V_0 obtained in a fit or even constrain V_0, if necessary, in order to extract B_0 and B_0'. Previous theoretical studies have predicted a variety of bulk modulus values,[21,34–36] though the difficulty in computing these parameters reliably has been noted.[21] These earlier studies either neglected thermal expansion[21,35,36] or omitted van der Waals dispersion,[34,36] which is significant for CO_2.[35,146]

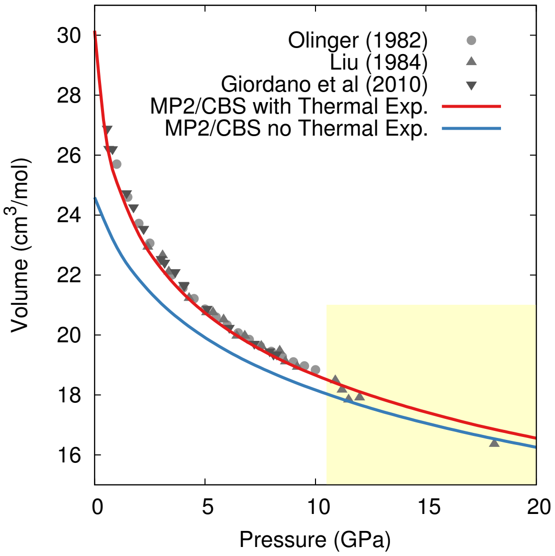

Figure 3.6: Comparison of the experimental and predicted MP2/CBS pressure versus volume curves at 296 K, with and without quasi-harmonic thermal expansion. Note that the drop in the experimental volumes above 10 GPa (shaded region) is believed to reflect a transition to phase III,[190] while the calculations presented are for phase I throughout.

Here, we demonstrate that the combination of high-level electronic structure calculations and a quasi-harmonic treatment predicts B_0 and B_0' in excellent agreement with the best experimental values across a wide range of temperatures.

Pressure versus volume curves were calculated by optimizing the crystal geometry under a series of external pressures ranging from 0–10 GPa (0–20 GPa for 296 K) at 0 K, 130 K, 190 K, and 296 K under the quasi-harmonic approximation. Analogous calculations were also performed at 0 K without the quasi-harmonic vibrational contribution F_{vib}. As a representative example, Figure 3.6 compares the experimental and room-temperature MP2/CBS predicted pressure versus volume curves with and without the inclusion of quasi-harmonic thermal expansion. Inclusion of thermal expansion proves critical to reproducing the experimental pressure/volume data. Differences between the curves with and without thermal expansion persist even at 20 GPa, where one might have hoped that the high external pressure would obviate the need to treat thermal expansion.

For each temperature and level of theory, the values of V_0, B_0, and B_0' were extracted via non-linear least squares fitting to the Vinet equation of state,

$$P = 3B_0 \frac{(1-\tilde{V})}{\tilde{V}^2} \exp\left[\frac{3}{2}(B_0'-1)(1-\tilde{V})\right] \tag{3.19}$$

where $\tilde{V} = (V/V_0)^{1/3}$. The Vinet equation of state fits prove much more robust than the Birch-Murnaghan one for the CO_2 P-V curves. The fits to the predicted P-V curves were validated by performing a second set of fits in which V_0 was fixed at the molar volume obtained directly by optimizing the crystal at a given temperature and zero pressure. Both sets of fits produced very similar volumes and bulk moduli.

Figure 3.7 compares the predicted values of B_0 and B_0' obtained here to previously reported theoretical and experimental values. Without the quasi-harmonic approximation, the bulk modulus parameters obtained here are similar to earlier predictions using MP2/aug-cc-pVTZ by Li and co-workers[21] and various dispersion-corrected density functional calculations.[35] However, the bulk modulus shrinks several-fold upon heating to room temperature, and the treatment of thermal expansion provided by the quasi-harmonic approximation is required to capture that.

Basis set effects are also fairly important for the bulk modulus—the MP2 B_0 value increases by 30–130% (depending on temperature) from a small aug-cc-pVDZ basis to the complete basis set limit. The pressure derivative B_0' is less sensitive to basis set. Correlation beyond second-order perturbation theory proves relatively unimportant here.

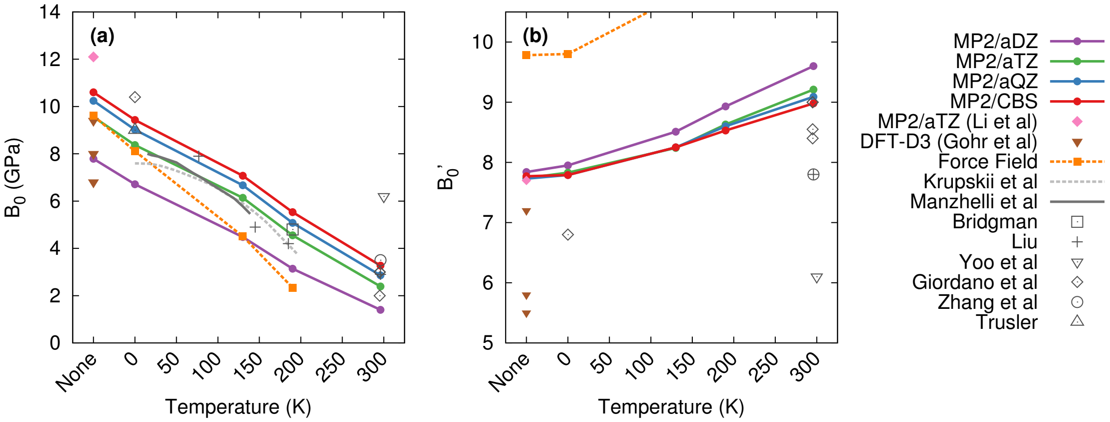

Figure 3.7: Experimental (gray) and predicted (colored) values of the (a) bulk modulus B_0 and (b) its first pressure derivative B_0'.The label "None" in the figures refers to calculations which neglect temperature and the quasi-harmonic approximation entirely.

At 190 K, switching from MP2 to CCSD(T) increases V_0 by 0.1 cm^3/mol, increases B_0 by 0.2 GPa, and does not alter B_0' (see Table S2 in the ESI†).

The experimental bulk modulus data exhibits considerable scatter, but the bulk moduli B_0 predicted here are consistent with most of the literature data across all temperatures (Figure 3.7). Less experimental data exists for the first-pressure derivative B_0', but values predicted here are in good agreement with the available experimental ones. MP2/CBS overestimates the reported room temperature experimental values of B_0' by 5–15%, but the predicted value lies within the typical experimental error bars. For instance, the MP2/CBS predictions of $B_0 = 3.3$ GPa and $B_0' = 9.0$ at 296 K are in excellent agreement with the Vinet equation of state fit by Giordano et al,[184] which found $B_0 = 3 \pm 1$ GPa and $B_0' = 8.4 \pm 0.8$. The MP2 predictions are also consistent with the Vinet fits to the Olinger[186] and Liu[183] experimental P-V curves reported by Giordano et al,[184] which exhibit even larger uncertainties. Moreover, the MP2 predictions compare well with experimental bulk modulus values at other temperatures, including those from Krupskii et al,[57] Manzhelii et al,[174] Bridgman,[185] Liu,[183] and Trusler.[188]

The experimentally obtained $B_0 = 6.2$ GPa and $B_0' = 6.1$ values at 300 K reported by Yoo et al[187] are considerable outliers with respect to both our theoretical predictions and the other experimental values. Ref 187 provides few details of the data or fitting procedure used for phase I, but their reported zero-pressure volume $V_0 = 25.1$ cm^3/mol is substantially smaller than the values of $\sim$30 $\pm$2 cm^3/mol found by Giordano et al,[184] 31.4 cm^3/mol inferred by Liu[183], and 30.1 cm^3/mol predicted by MP2/CBS geometry optimization. In fact, their room-temperature V_0 is smaller than the experimental volume of 25.8 cm^3/mol at 6 K.[57] Therefore, these bulk modulus values probably reflect a spurious fit to the experimental data.

For comparison, the force field predicts a reasonable bulk modulus without temperature or at 0 K (where only zero-point effects are included), but it exaggerates the thermal expansion and predicts that the bulk modulus decreases much more rapidly with temperature than experiments or the MP2 calculations indicate. Similarly, the first pressure derivative of the bulk modulus is overestimated and increases too quickly with temperature in the force field model. Note too that at 296 K, the CO_2 crystal proved unbound with the force field model, and no reasonable fit could be found to the Vinet equation of state.

In the end, the electronic structure results here demonstrate that theory can provide a powerful tool for predicting properties such as the bulk modulus, which can be difficult to extract reliably from experiment. The calculations here provide support for the room-temperature bulk moduli obtained by Giordano et al and others, while simultaneously suggesting that some reported values are unlikely. Furthermore, theory can be used to identify a plausible experimental zero-pressure volume, which is often a key step in extracting bulk modulus parameters from experiment. Finally, the treatment of thermal expansion proves critical to predicting the correct the temperature-dependence of the bulk modulus parameters.

3.6 Conclusions

In summary, we are rapidly transitioning into an era where electronic structure theory can directly predict a wide range of experimentally observable molecular crystal properties under practical temperature and pressure conditions. As shown here, the combination of accurate electronic structure theory calculations and a quasi-harmonic treatment of thermal expansion enables one to predict crystal structures, thermodynamics, and mechanical properties for phase I carbon dioxide in excellent agreement with experiment. While the simple force field considered here behaves very well at low temperatures and predicts results on roughly par with those from MP2/aug-cc-pVTZ, the electronic structure calculations provide substantially improved agreement with experiment at higher temperatures.

The performance of the quasi-harmonic approximation seen here does start to degrade at higher temperatures, so it remains to be seen how well it performs in larger crystals which are stable at room temperature and above. Still, the excellent performance seen here up to 200 K (or room temperature for the bulk modulus) for carbon dioxide provides considerable cause for optimism. Of course, the increased anharmonicity found in larger, more flexible organic molecules will also create new challenges for the simple quasi-harmonic approximation used here.

The quantum mechanical calculations here are made feasible by fragment-based electronic structure methods, which make MP2 and even coupled cluster calculations computationally affordable for molecular crystals. Although the extrapolated complete-basis MP2 and CCSD(T) calculations employed on CO_2 here would be much more computationally challenging for a pharmaceutical crystal, in many cases one can probably obtain useful predictions using a lower level of theory. MP2/aug-cc-pVTZ already predicts many of the properties in reasonable agreement with experiment, albeit with several-fold less computational effort than the larger-basis results. It may provide a useful level of theory for modeling crystals of larger molecules. Continuing algorithmic developments and decreasing costs of computer hardware will hopefully make finite-temperature predictions on chemically interesting organic molecular crystals routine in the near future.

Chapter 4

How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures?

4.1 Outline

Molecular crystals expand appreciably upon heating due to both zero-point and thermal vibrational motion, yet this expansion is often neglected in molecular crystal modeling studies. Here, a quasi-harmonic approximation is coupled with fragment-based hybrid many-body interaction calculations to predict thermal expansion and finite-temperature thermochemical properties in crystalline carbon dioxide, ice Ih, acetic acid, and imidazole. Fragment-based second-order Møller-Plesset perturbation theory (MP2) and coupled cluster theory with singles, doubles and perturbative triples (CCSD(T)) predict the thermal expansion and the temperature dependence of the enthalpies, entropies, and Gibbs free energies of sublimation in good agreement with experiment. The errors introduced by neglecting thermal expansion in the enthalpy and entropy cancel somewhat in the Gibbs free energy. The resulting ~1–2 kJ/mol errors in the free energy near room temperature are comparable to or smaller than the errors expected from the electronic structure treatment, but they may be sufficiently large to affect free energy rankings among energetically close polymorphs.

4.2 Introduction

Molecular crystals occur in pharmaceuticals, organic semiconductor materials, and many other areas of chemistry. The molecular packing in a crystal has substantial impacts on its properties. Polymorphism, or the tendency for a given molecule to adopt multiple distinct packing motifs, provides excellent examples of this phenomenon. Estimates suggest that at least half of organic crystals exhibit multiple polymorphs.[191] Undesirable pharmaceutical polymorphs can exhibit reduced bioavailability,[1–3,192] while at other times alternative crystal forms may be targeted for their improved physical properties.

Computational chemistry plays an increasingly important role in predicting crystal structures, phase diagrams, spectroscopic observables, mechanical properties, and other molecular crystal properties that can help characterize crystals or identify potential new forms. Particular attention in recent years has been focused on crystal structure prediction.[6–8,12,13,193] The most stable crystal structures exhibit the lowest free energies. However, rankings based on lattice energies, which neglect both thermal and vibrational zero-point energy effects, are often used as proxy for free energy. Twenty years ago, Gavezzotti and Filippini[194] argued that the free energy contributions arising from room-temperature lattice vibrational entropy are generally smaller than the enthalpic differences between polymorphs.

More recently, Nyman and Day[30] surveyed 508 sets of polymorphic crystals and found that the harmonic vibrational free energy contributions at 300 K often contribute ~1 kJ/mol or less to the relative stabilities among polymorphs. Though the vibrational free energy contribution is small, they observed that it frequently opposes the lattice energy difference, which suggests that it will eventually lead to an enantiotropic phase transition at some temperature (unless the crystal melts first). Indeed, in almost 10% of the cases they considered, free energy rankings at room temperature predict a different polymorph stability than the one inferred from the lattice energy. Such observations are consistent with the frequency with which temperature-dependent transitions between polymorphs occur experimentally.

Many examples where vibrational zero-point and free energy contributions feature in the context of polymorphism can be found in the literature, including glycol and glycerol,[195] pyridine,[196] glycine,[197] co-crystals of urea and acetic acid,[198] and aspirin.[14] In aspirin, for instance, the two polymorphs are predicted to be virtually degenerate in lattice energy, but the free energy appears to favor form I.[14] A couple groups considered the impact of free energy ranking in the fourth blind test of crystal structure prediction, though the free energy contributions did not significantly revise the lattice energy rankings in those particular crystals.[8]

Temperature also plays an important dynamical role in molecular crystals. Thermal averaging over lattice energy minima often effectively reduces the number of minima on the free energy surface.[199] Metadynamics studies on benzene,[139] 5-fluorouracil,[140] and pigment red 179[200] each demonstrate a reduction from many lattice energy minima to a smaller number of free energy minima, though the extent of reduction varies widely with the nature of the system.[199]

Finite temperature effects on crystal properties are not limited to thermochemistry. For example, crystals typically expand upon heating, which affects the electronic coupling and non-local electron-phonon coupling in organic semi-conductors. Shifting the intermolecular separation between two adjacent anthracene molecules taken from the crystal can alter the transfer integral by ∼30%, for example.[201] Thermal expansion also narrows the valence and conduction bandwidths in organic semiconductor materials like pentacene and rubrene.[202]

All of these examples demonstrate the potential importance of accounting for finite temperature and computing free energies instead of lattice energies when modeling molecular crystals. However, even studies that do estimate finite-temperature free energies often do so using a fixed-cell harmonic approximation based on the minimum electronic energy structure, ignoring thermal expansion of the crystal. This is especially true when the molecular crystals are modeled using computationally expensive electronic structure methods instead of classical force fields. Thermal expansion alters many crystal properties. For example, expansion-induced softening of the low-frequency intermolecular lattice phonon modes will increase the magnitude of the entropic contributions. This raises the question: How significant are thermal expansion effects in predicting molecular crystal properties at finite temperatures?

Modeling thermal expansion requires minimizing the crystal structure at a given temperature with respect to the free energy instead of the more readily computed electronic energy. Using conventional molecular dynamics and/or free energy sampling techniques is generally computationally prohibitive when using quantum mechanical techniques. Instead, the quasi-harmonic approximation provides a computationally practical alternative to free energy sampling techniques.[52,203–205] It assumes that anharmonicity in the crystal arises primarily from the intermolecular expansion. It approximates the crystal volume dependence of the harmonic phonon frequencies in terms of a reference set of phonons and a set of Grüneisen parameters which describe how those phonon frequencies change with volume.

Despite its simplicity, the quasi-harmonic approximation provides a useful tool for investigating how the unit cell volume and other properties of small-molecule crystals vary as a function of temperature (though it does not address the dynamical thermal averaging aspects mentioned earlier). In a recent study on crystalline carbon dioxide (phase I),[93] we demonstrated that a quasi-harmonic treatment of thermal expansion at the complete-basis-set (CBS) limit second-order Møller-Plesset perturbation theory (MP2) or even coupled cluster singles, doubles, and perturbative triples (CCSD(T)) level of theory accurately captures the ~10% volume expansion that occurs between the minimum electronic energy structure and the structure near the 194.7 K sublimation point. The same model also allows one to predict the sublimation enthalpy and entropy and the room-temperature bulk modulus in excellent agreement with experiment. In contrast, neglecting thermal expansion introduces appreciable errors in the predicted thermochemistry and mechanical properties. These high-level calculations in a periodic crystal are made feasible using the fragment-based hybrid many-body interaction (HMBI) model,[44,86,144] which combines a QM treatment of the intra- and dominant intermolecular interactions with a classical molecular mechanics (MM) treatment of the weaker interactions.

Here, we extend the previous study by comparing predictions of finite-temperature properties in several different small-molecule crystals: carbon dioxide, ice Ih, the orthorhombic polymorph of acetic acid, and the α polymorph of imidazole. These relatively simple crystals were chosen because they exhibit a variety of intermolecular packing interactions, they have a experimental data available at various different temperatures, and they are small enough to enable relatively high-level electronic structure methods to be employed. The absence of appreciable conformational flexibility in these small molecules provides a best-case scenario for the quasi-harmonic approximation, since any changes in the intramolecular structures with temperature will be small.

In the following sections, we first examine the extent of thermal expansion that occurs in each crystal due to zero-point and thermal contributions. The degree of expansion observed varies with the types of intermolecular interactions found in the different crystals. Second, we evaluate the performance of fragment-based electronic structure models with and without quasi-harmonic expansion for predicting the enthalpies, entropies, and Gibbs free

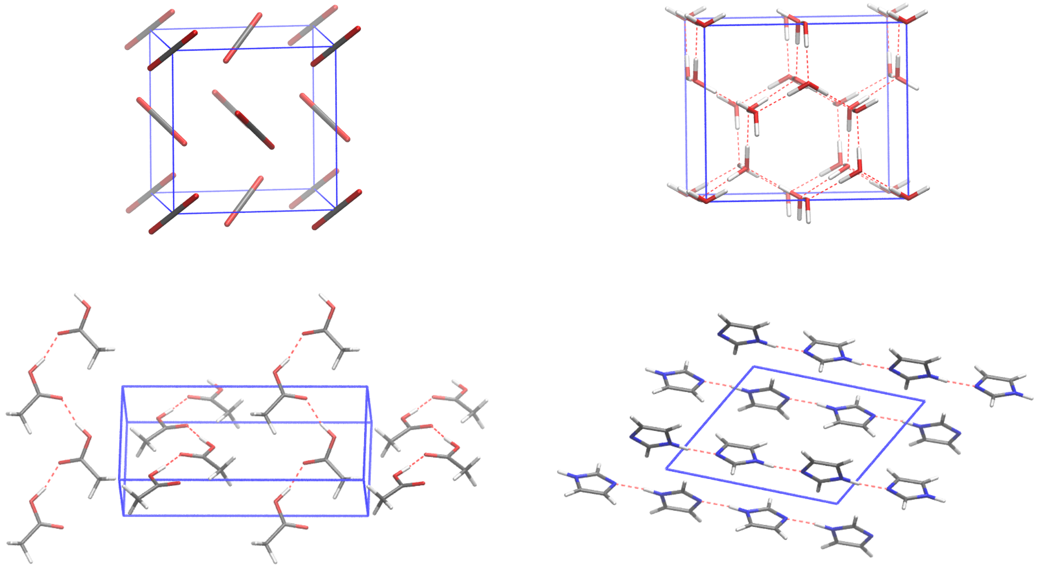

Figure 4.1: *Clockwise from top left:* Structures of phase I carbon dioxide, ice Ih, α imidazole, and orthorhombic acetic acid.

energies of sublimation for these crystals over a range of temperatures. The predictions are assessed against experimental data or empirical results derived from experiment. Finally, we investigate the importance of electronic structure method and basis set on predicting these properties correctly, and we attempt to decouple effects of the model chemistry on the structure optimization/phonon calculation from those on the lattice energy evaluation.

4.3 Theory

4.3.1 Quasi-harmonic structure optimization

The structure of a crystal at a given pressure and temperature is determined by minimizing the Gibbs free energy,

$$G(T, P) = U_{el} + PV + F_{vib}(T) \tag{4.1}$$

with respect to the atomic coordinates and unit cell lattice parameters. In this equation, U_{el} is the internal (electronic) energy, PV is the pressure/volume contribution, and F_{vib} is the Helmholtz vibrational free energy.

The electronic energy U_{el} is determined using the fragment-based hybrid many-body interaction (HMBI) model.[44,45,86,87] The HMBI model combines a QM treatment of the individual molecules in the unit cell (1-body terms) and their short-ranged pairwise interactions (SR 2-body terms) with an MM polarizable force field treatment of longer-range (LR 2-body terms) and many-body intermolecular interactions,

$$U_{el} = E^{\text{QM}}_{\text{1-body}} + E^{\text{QM}}_{\text{SR 2-body}} + E^{\text{MM}}_{\text{LR 2-body}} + E^{\text{MM}}_{\text{many-body}} \tag{4.2}$$

In practice, the short-range two-body QM terms capture interactions between molecules inside the unit cell and the nearby periodic image cells, while the long-range two-body MM terms are handled via Ewald summation.

The Helmholtz vibrational free energy $F_{vib}(T)$ is derived from standard harmonic oscillator vibrational partition function expressions obtained from statistical mechanics. It includes both zero-point energy and thermal contributions to the free energy:

$$F_{vib}(T) = \frac{N_a}{k_{total}} \sum_{k}^{k_{total}} \sum_{i} \frac{\hbar\omega_{k,i}}{2} + k_B T \ln\left[1 - \exp\left(\frac{\hbar\omega_{k,i}}{k_B T}\right)\right] \tag{4.3}$$

where k_{total} the total number of **k**-point. Evaluating $F_{vib}(T)$ requires knowledge of the harmonic vibrational frequencies for the current unit cell. In principle, this requires optimizing the atomic positions with fixed lattice parameters followed by computing the phonon frequencies via lattice dynamics. Unfortunately, that repeating that process for each step in a free-energy minimization is very computationally demanding.

Instead, the i-th vibrational frequency $\omega_{k,i}$ at a particular unit cell volume V and **k**-point k is estimated relative to a reference harmonic frequency $\omega_{k,i}^{ref}$ computed at a reference unit cell volume V_{ref} according to the quasi-harmonic approximation. The quasi-harmonic approximation defines the change of the i-th vibrational frequency with respect to unit cell volume according to a Grüneisen parameter $\gamma_{k,i}$,

$$\gamma_{k,i} = \frac{\partial\omega_{k,i}}{\partial V} \tag{4.4}$$

Integrating Eq 4.4 gives,

$$\omega_{k,i} = \omega_{k,i}^{ref}\left(\frac{V}{V^{ref}}\right)^{-\gamma_{k,i}} \tag{4.5}$$

Here, the reference volume and frequencies are obtained via optimizing the crystal unit cell with respect to the electronic energy U_{el}. The Grüneisen parameters are determined via finite difference, using two additional structure optimizations and vibrational frequency calculations performed in fixed unit cells that have been expanded or compressed by a small amount.

The vibrational modes computed for a particular unit correspond to the zone-center ($\mathbf{k} \neq 0$) phonons. However, phonon dispersion at $\mathbf{k} \neq 0$ can play an important role. The phonon modes at a given $\mathbf{k}$-point are evaluated via lattice dynamics,[114] which involves the construction and diagonalization of the mass-weighted supercell dynamical matrix,

$$\hat{D}_{\alpha,\beta}(l,l',\mathbf{k}) = \frac{1}{\sqrt{M_l M_{l'}}} \sum_{\kappa} \frac{\partial V}{\partial R_\alpha(0) \partial R_\beta(\kappa)} \exp\left(-2\pi i \mathbf{k} \cdot \delta \mathbf{R}_{l,l'}(0,\kappa)\right) \tag{4.6}$$

where $\frac{\partial V}{\partial R_\alpha(0) \partial R_\beta(\kappa)}$ is an element of the supercell Hessian between coordinate $\alpha(0)$ of atom l in the central unit cell and coordinate $\beta(\kappa)$ of atom l' in periodic image cell κ. $\delta \mathbf{R}_{l,l'}(0,\kappa)$ is the distance between atom l and l'.

A major advantage a fragment-based methods like HMBI or the binary interaction model[17] have over more traditional models like periodic DFT or periodic MP2 is that the construction of the lattice dynamics supercell Hessian requires little additional computational cost compared to the standard unit cell Hessian. For HMBI, all the necessary QM contributions to the supercell Hessian are already available in the standard unit cell Hessian. The Hessian contributions arising from a monomer or a short range dimer two-body interactions in the standard Hessian can be transposed onto the translationally equivalent dimer in the supercell Hessian according to the periodic symmetry of the lattice. The only additional HMBI contribution needed is the MM supercell Hessian, which requires minimal additional computational effort compared to the cost of evaluating the QM two-body Hessian contributions.

4.3.2 Thermochemistry

Once the crystal structures and vibrational frequencies are known at a given temperature and pressure, one can compute other thermodynamic quantities. Here, we predict enthalpies, entropies, and Gibbs free energies of sublimation for comparison with experiment. These are computed as the enthalpy, entropy, or free energy difference between the gas and solid, with the solid contribution normalized according to the number of molecules n in the unit cell:

$$\Delta H_{sub} = H_{gas} - \frac{1}{n} H_{solid} \tag{4.7}$$

$$\Delta S_{sub} = S_{gas} - \frac{1}{n} S_{solid} \tag{4.8}$$

$$\Delta G_{sub} = H_{sub} - T \Delta S_{sub} \tag{4.9}$$

The enthalpy of the solid is computed from the electronic energy plus the PV term and the vibrational energy U_{vib}

$$H_{solid} = U_{el,solid} + PV + U_{vib,solid} \tag{4.10}$$

The harmonic vibrational energy is determined from the standard vibrational statistical mechanics equation,

$$U_{vib,solid} = \frac{N_A}{k_{total}} \sum_{k}^{k_{total}} \sum_{i} \left(\frac{\hbar\omega_{k,i}}{2} + \frac{\hbar\omega_{k,i}}{\exp\left(\frac{\hbar\omega_{k,i}}{k_B T}\right) - 1} \right) \tag{4.11}$$

The entropy of the solid is determined from its standard vibrational statistical mechanics equation plus the configurational entropy

$$S_{solid} = S_{solid,vib} + S_{conf} \tag{4.12}$$

$$S_{solid,vib} = \frac{N_A}{k_{total}} \sum_{k}^{k_{total}} \sum_{i} \left(\frac{\hbar\omega_{k,i}}{T\left(\exp\left(\frac{\hbar\omega_{k,i}}{k_B T}\right) - 1\right)} - k_B \ln\left[1 - \exp\left(\frac{\hbar\omega_{k,i}}{k_B T}\right)\right] \right) \tag{4.13}$$

For most crystals considered here, the configurational entropy S_{conf} is set to zero. However, the intrinsic proton disorder in ice Ih produces a non-zero configurational entropy of $R\ \ln\left(\frac{3}{2}\right)$ according to the Pauling model.[206]

The gas phase is modeled as an ideal gas, with the enthalpy written as the sum of the electronic energy, the translational and rotational, and vibrational energy plus a factor of RT from the PV term.

$$H_{gas} = U_{el,gas} + U_{trans,gas} + U_{rot,gas} + U_{vib,gas} + RT \tag{4.14}$$

The translational energy ($U_{trans,gas}$) is equal to $\frac{3}{2}RT$. The rotational energy ($U_{trans,gas}$) is equal to RT for carbon dioxide (linear molecule) and $\frac{3}{2}RT$ for all other compounds considered here. The vibrational energy contribution ($U_{vib,gas}$) is given by,

$$U_{vib,gas} = N_A \sum_{i} \left(\frac{\hbar\omega_i}{2} + \frac{\hbar\omega_i}{\exp\left(\frac{\hbar\omega_i}{k_B T}\right) - 1} \right) \tag{4.15}$$

The gas phase entropy is the sum of the vibrational entropy ($S_{rot,gas}$), translation entropy

($S_{trans,gas}$), the rotational entropy ($S_{rot,gas}$)

$$S_{gas} = S_{vib,gas} + S_{trans,gas} + S_{rot,gas} \tag{4.16}$$

Like the solid, the gas phase vibrational entropy is determined from the standard harmonic oscillator model,

$$S_{gas,vib} = N_A \sum_i \left(\frac{\hbar\omega_i}{T\left(\exp\left(\frac{\hbar\omega_i}{k_B T}\right) - 1\right)} - k_B \ln\left[1 - \exp\left(\frac{\hbar\omega_i}{k_B T}\right)\right] \right) \tag{4.17}$$

The translation entropy is based on the ideal gas model,

$$S_{gas,tran} = R\ \ln\left[\left(\frac{2\pi M k_B T}{h^2}\right)^{\frac{3}{2}} \left(\frac{k_B T \exp\left(\frac{5}{2}\right)}{P}\right)\right] \tag{4.18}$$

The rotation of carbon dioxide is modeled as a linear ideal gas molecule,

$$S_{gas,rot,CO_2} = R\ \ln\left(\frac{8\pi^2 eTIk_B}{\sigma h^2}\right) \tag{4.19}$$

For all other molecules considered here, the rotational entropy was modeled according to the standard ideal gas rotational entropy expression for a non-linear polyatomic molecule,

$$S_{gas,rot} = R\ ln\left[\frac{\left(512\pi^7 T^3 e^3 k_B^3 I_A I_B I_C\right)^{\frac{1}{2}}}{\sigma h^3}\right] \tag{4.20}$$

where the I's are the moment of inertia. The number of symmetrical rotations (σ) is 2 for carbon dioxide and water and 1 for acetic acid and imidazole.

4.4 Computational Methods

Four crystals are considered here: ice Ih, carbon dioxide (phase I), acetic acid (ambient-pressure orthorhombic phase), and imidazole (α polymorph). The initial acetic acid and imidazole structures were taken from the Cambridge Structure Database (RefCodes ACETAC01[127] and IMAZOL06,[129] respectively). The initial ice structure used the dipole-free 16-molecule supercell obtained from Ref 207. All structures reported for carbon dioxide here were obtained previously[93] using the same quasi-harmonic techniques described here.

Electronic energies were calculated using the HMBI fragment approach. Substantial computational savings were achieved for acetic acid ($Pna2_1$ symmetry), imidazole ($P2_1/c$) and carbon dioxide ($Pa\bar{3}$) by exploiting space group symmetry for the energy, gradient, and Hessian evaluations.[92] Specific analysis of the symmetry for these crystals have been presented previously.[92,93] Ice exhibits $P1$ symmetry due to its disordered proton arrangement, which means that only basic translational symmetry arising from the periodic supercell can be exploited.

For structure optimizations and phonon frequency calculations, the QM contributions in HMBI were evaluated using the counterpoise-corrected[163] density-fitted MP2[116,117,157,158] and Dunning aug-cc-pVXZ (abbreviated to aXZ) basis sets[120,121] as implemented in Molpro 2012.[161,162] Molpro uses analytical MP2 nuclear gradients, while the Hessian elements were computed via finite difference of the gradients. Complete-basis-set (CBS) limit MP2 results were obtained using the standard two-point extrapolation models.[164,165] In some cases, calculations at the estimated CCSD(T)/CBS limit were also performed. For single-point energies or geometry optimizations, this was done using a focal point method which combined MP2/CBS results with a correction for the difference between MP2/aDZ and CCSD(T)/aDZ. For CCSD(T)-level quasi-harmonic optimizations in carbon dioxide, the reference phonon frequencies and Grüneisen parameters were approximated using the values obtained at the MP2/CBS limit. See Ref 93 for details.

During the structure optimization and lattice dynamics calculations, the HMBI MM contributions were computed with the Amoeba polarizable force field[75,208] and Tinker 6.3.[166] Force field parameters for carbon dioxide were computed using Poltype[167] as described in Ref 93. A subsequent set of single-point energy refinements reported here replace the MM contribution with one evaluated using our *ab initio* force field (AIFF).[45,72,85] The AIFF includes multipolar two-body electrostatics (up to hexadecapole), many-body polarization, two- and three-body dispersion. These contributions are represented in terms of distributed multipoles, distributed polarizabilities, and distributed dispersion coefficients which are computed using asymptotically corrected PBE0 density functional theory and aug-cc-pVTZ basis with CamCasp 5.6.[209]

For each crystal, the following steps were repeated with each electronic structure method/basis set combination:

1. The crystal structure was first optimized with respect to the HMBI electronic energy using a given method/basis set and Amoeba MM terms. Phonons were computed at the same level of theory using lattice dynamics with a 3×3×3 supercell and a 3×3×3 Monkhorst-Pack k-point grid. These calculations provide the initial structure, reference unit cell volume V^{ref}, and reference frequencies $\omega_{k,i}^{ref}$ used in the quasi-harmonic approximation. Gas-phase molecules were optimized using the same electronic struc-

ture method and basis set as the crystal.

2. Grüneisen parameters (Eq 4.4) were obtained via finite difference between phonon frequencies obtained at two distinct unit cell volumes. Specifically, the optimized unit cells from Step 1 were separately expanded and contracted isotropically by 10 Å^3. For each of these new unit cell dimensions, the atomic coordinates were relaxed subject to fixed lattice parameters. Lattice dynamics phonon calculations were performed on each optimized structure. Overlap of the normal mode eigenvectors was used to ensure proper assignment of the frequencies between the two structures in the finite difference.

3. Quasi-harmonic Gibbs free energy optimizations were then performed for each species at each of several different temperatures. These calculations allow the unit cell to relax due to zero-point energy and thermal expansion. The vibrational free energy contribution to the crystal was estimated as a function of temperature and unit cell volume according to Eqs 4.3 and 4.5. A pressure of 1 atm was used in all cases. The PV term is negligible in the solid at this pressure and was omitted, but it was included in the gas-phase free energies.

4. For each temperature, the enthalpies, entropies, and Gibbs free energies of sublimation were computed at the same level of theory using the structures and quasi-harmonic frequencies obtained from Step 3.

5. A second set of enthalpies of sublimation was computed for each geometry from Step 3 in which the single-point electronic energy contributions (e.g. U_{el} in Eq 4.1)

is computed using CCSD(T)/CBS QM and AIFF MM. Comparing these single-point enthalpies at the same, high level of theory using geometries obtained with different basis sets helps decouple the effects of crystal structure and electronic structure method/basis set on the predicted enthalpies of sublimation.

6. Finally, the volumes, enthalpies, entropies, and Gibbs free energies were also evaluated without the quasi-harmonic approximation (denoted "No QHA" in figures and tables). In this case, the electronic energy optimized structures and the reference phonon frequencies from Step 1 are assumed not to change with temperature when evaluating the various statistical mechanics expressions in Section 4.3.2.

The specific electronic structure methods and basis sets applied to each of the four crystals depends on the computational cost. For carbon dioxide, the small size of the molecule and high symmetry in the phase I crystal enabled geometry optimizations using up to MP2/CBS and even the estimated CCSD(T)/CBS limit. Despite the absence of useful symmetry in ice Ih, the small size of water molecules enabled optimizations up to the MP2/CBS limit. Post-MP2 corrections to the ice lattice energy and lattice constants are small,[44,107] so coupled cluster calculations were not performed on that crystal. Due to their larger molecular sizes, the acetic acid and imidazole crystal optimizations were performed using only the smaller aDZ and aTZ basis sets.

4.5 Results and Discussion

The following sections examine the convergence of the predicted molar volume, enthalpy of sublimation, and entropy of sublimation for each of the four crystals both with respect to the method/basis set and relative to experiment. Experimental molar volumes were obtained directly from the literature. The enthalpies and entropies of sublimation were mostly derived empirically using experimental data found in the literature, as described in the Appendix B.

Experimental uncertainties were not reported for many of the individual contributions used to derive the enthalpies and entropies of sublimation. For enthalpies of sublimation, it is not unusual to find values reported in the literature that differ by several kJ/mol. For instance, the ΔH_{sub} values reported for imidazole by Chickos and Acree[210] in the vicinity of room temperature vary from ∼75–85 kJ/mol. Even if one discards the largest outlier (74.5 kJ/mol), the remaining values span a 5 kJ/mol range. Appreciable errors are also likely for the empirical entropies of sublimation.

Finally, note that the some of the carbon dioxide data reported here comes from an earlier publication, while other data is new. Specifically, the volumes in Figure 4.2, sublimation enthalpies in Figure 4.3a, and sublimation entropies in Figure 4.4a have been reported previously,[93] while the CCSD(T)/CBS + AIFF enthalpies in Figure 4.3b and sublimation temperatures in Table 4.3 are new here. The data for the other three crystals is reported here for the first time.

4.5.1 Molar Volumes

The predicted molar volumes of each crystal were computed as a function of temperature using several different electronic structure method/basis set combinations. These results are compared against the corresponding experimental data in Figure 4.2. The "No QHA" data in Figure 4.2 refers to the structures obtained by minimizing the electronic

Table 4.1: Percent molar volume expansion arising from the zero-point vibrational energy and thermal vibrational contributions. Percentages are reported relative to the electronic energy minimized structure (no QHA).

Crystal	Structures	T_{max}	Zero-Point[a]	Thermal[b]	Total
Carbon Dioxide	CCSD(T)/CBS	190 K	2.9%	7.6%	10.5%
Ice	MP2/CBS	265 K	3.3%	2.1%	5.4%
Acetic acid	MP2/aTZ	278 K	2.1%	5.6%	7.7%
Imidazole	MP2/aTZ	283 K	2.1%	4.2%	6.3%

[a] Expansion between the structures optimized at 0 K with and without quasi-harmonic zero-point vibrational energy.
[b] Quasi-harmonic thermal expansion occurring between 0 K and T_{max}.

energy instead of the quasi-harmonic Gibbs free energy.

Experimental temperature-dependent volume data for carbon dioxide and ice were obtained from Refs 57 and 58, respectively. Experimental volumes for acetic acid at 40 K (RefCode ACETAC07),[211] 83 K (ACETAC02),[127] 133 K (ACETAC03),[212] and 278 K (ACETAC01)[127] were found in the Cambridge Structure Database. Experimental volumes for imidazole were similarly obtained at 103 K (IMAZOL06),[129] 123 K (IMAZOL04),[213] 173 K (IMAZOL14),[214] and room temperature.[213,215,216] The Cambridge database contains three different room-temperature structures for imidazole: 55.18 cm^3/mol (IMAZOL05), 54.41 cm^3/mol (IMAZOL10), and 55.23 cm^3/mol (IMAZOL13). All three values are reported here.

Before studying the performance of various models in detail, we examine broad trends associated with the thermal expansion. Table 4.1 summarizes the predicted percent expansions using the highest-level model chemistry reported for each crystal. First, we observe that ice expands the least (5.4%), while carbon dioxide expands the most (10.5%). The 6–8% percent expansion for the other two crystals lies in between these two extremes. These expansion trends can be rationalized based on the crystal packing. Carbon dioxide is held together primarily by relatively weak quadrupolar and van der Waals interactions, which allow for substantial expansion. In contrast, the strong three-dimensional hydrogen bond network in ice inhibits thermal expansion. The other two crystals exhibit one-dimensional hydrogen-bonded chains. Thermal expansion is hindered along the hydrogen bond chain directions, but it occurs more readily in directions orthogonal to the chains.

Second, Table 4.1 decomposes the total expansion into the portion which occurs solely due to zero-point vibrational energy (ZPVE) and the portion which arises from the finite-temperature contributions. The former is the difference between the volumes of the electronic energy minimized structures ("No QHA") and the 0 K structures. The latter reports the expansion between the 0 K and maximum temperature structures considered. The expansion behaviors of the four crystals are also evident in Figure 4.2. In all cases, zero-point vibrational energy drives a substantial fraction of the overall expansion. In ice, 60% of the overall expansion occurs due to zero-point energy. Zero-point expansion represents a smaller fraction of the overall expansion in the other three crystals, but it still amounts to almost a third of the total expansion.

Next, we focus on the detailed performance of the models for individual crystals. We previously[93] examined the thermal expansion of phase I carbon dioxide using basis sets of increasing size up to the MP2/CBS and CCSD(T)/CBS levels. These results are repeated in Figure 4.2a for comparison with the other three crystals. The predicted MP2/aDZ molar volume is significantly too large at low temperatures, and that it expands too quickly compared to the experimental crystal. Increasing the basis set toward the CBS limit decreases the volume for all temperatures, improving the agreement with experiment. In the

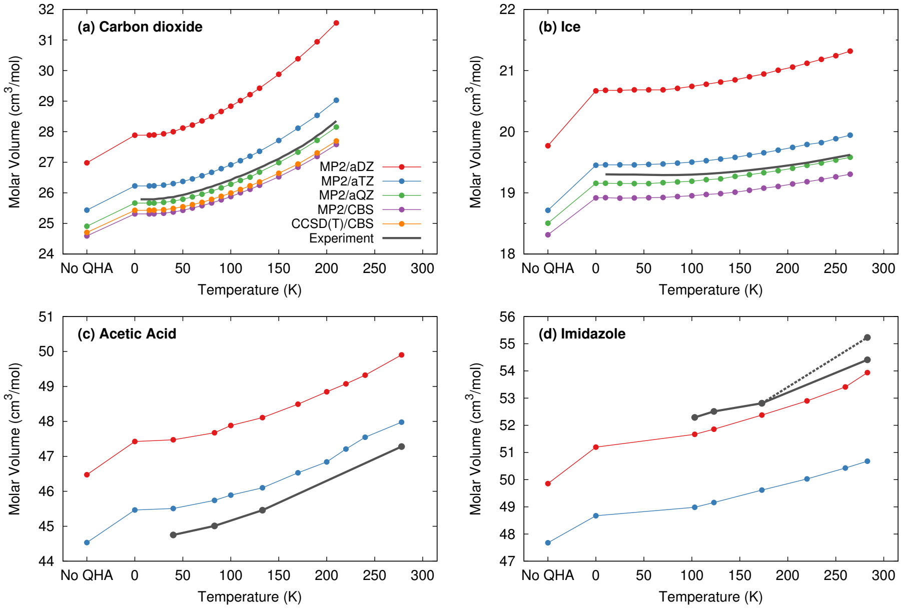

Figure 4.2: Predicted thermal expansion of crystalline carbon dioxide, ice, acetic acid, and imidazole. The "No QHA" volumes were obtained via conventional minimization of the electronic energy.

CBS limit, MP2 underestimates the low-temperature volume, and this is slightly corrected at the CCSD(T) level. This underestimation probably occurs in part due to the neglect of repulsive three-body dispersion, which is significant carbon dioxide.[168] Using the larger basis sets also improves the rate of thermal expansion, producing expansion curves that are nearly parallel to the experimental one.

Similar behavior is observed for ice. MP2/aDZ substantially overestimates the molar volume at all temperatures, but using larger basis sets produces volumes that agree with experiment to within 0.4 cm^3/mol or less. Once again, MP2/aQZ and MP2/CBS underestimate the molar volume. The models also modestly overestimate the rate of thermal expansion in ice, most notably at higher temperatures. Experimentally, the crystal expands by 1.6% between 10 K and 265 K. Over the same range, the quasi-harmonic MP2 calculations predict expansion of 3.1%, 2.5%, 2.2%, and 2.0% as the basis is increased along the series aDZ, aTZ, aQZ, and extrapolated to the CBS limit, respectively. Increasing the basis set size reduces the non-parallelarity error in the temperature dependence of the volume, just as was seen for carbon dioxide.

Ice Ih is unusual in that it exhibits negative thermal expansion (i.e. the volume actually decreases with increasing temperature) at low temperatures. Experimentally, the volume at 70 K (roughly the minimum) is 0.06% smaller than the volume at 10 K. The MP2/aDZ predictions miss this negative thermal expansion entirely. The larger-basis set calculations do capture the correct qualitative behavior, though they are not quantitative. The predicted volume decreases range from less than -0.01% in MP2/aTZ to -0.04% for MP2/CBS). The models also underestimate the temperature range over which the negative expansion occurs. Both MP2/aTZ and MP2/aQZ predict a minimum volume around 40 K and MP2/CBS predicts the minimum around 25 K, compared to 70 K experimentally.

Previous studies indicate that the electronic energy minimum structure and the lattice energy change little between MP2 and CCSD(T) treatments of the one-body and two-body terms,[44,107] so it seems unlikely that performing coupled cluster calculations would alter the MP2 predictions significantly. The residual errors here most likely stem from the force field treatment of the strong many-body polarization effects which arise from cooperative hydrogen bonding and/or from the quasi-harmonic treatment. It would be instructive to examine to what extent a more elaborate many-body water potential[217,218] could correct the residual errors.

Like carbon dioxide and ice, the acetic acid molar volume improves as the basis set is increased from aDZ to aTZ. MP2/aDZ and MP2/aTZ consistently overestimate the volume by $\sim$2.5 cm^3/mol and $\sim$0.7 cm^3/mol, respectively. The acetic acid quasi-harmonic optimizations suffered from minor numerical convergence issues (particularly for MP2/aDZ) which led to the slight roughness in the predicted expansion curves, but both curves are generally parallel to the experimental one. Experimentally, the acetic acid crystal expands by 5.6% between 40 K and 278 K, versus expansion of 5.3% and 5.4% with MP2/aDZ and aTZ respectively. Based on the trends observed for carbon dioxide and ice, one might anticipate smaller errors and improved parallelarity in a larger aQZ basis set. Those calculations were not performed, however, due to their high computational expense.

Finally, imidazole provides an interesting case. In the previous three crystals, MP2/aDZ substantially overestimated the crystal volume, and larger basis sets correct this. In imidazole, MP2/aDZ already slightly underestimates the volume but is (fortuitously) in excellent agreement with the experimental volumes. Increasing the basis set from aDZ to aTZ shrinks the volume further, leading to much larger errors.

The problem stems from the strong π-electron van der Waals interactions in imidazole. MP2 exhibits known problems overestimating the strength of such non-covalent interactions.[62,219] Similarly, MP2 overestimates the lattice energy of crystalline imidazole by 10–15% compared to CCSD(T).[44,45] This overbinding apparently leads to a crystal structure that is too dense. From a practical point of view, this is potentially problematic for optimizing such structures with fragment-based methods. Coupled cluster theory corrects these weaknesses of MP2, but such calculations are generally too expensive for crystal structure optimization beyond the simplest crystals. Applying a dispersion correction to MP2 as in the MP2C model[63,65] helps, but analytic gradients of MP2C are not currently available. At present, dispersion-corrected periodic density functional theory methods probably provide the best alternative for optimizing such crystals.[67,220,221]

As noted earlier, several experimental values have been reported for the molar volume of imidazole at 283 K: two values around 55.2 cm^3/mol, and one at 54.4 cm^3/mol (a difference of about 1.5%). The predicted thermal expansion of MP2/aDZ exhibits an increased slope above 260 K that could support the two larger experimental volumes, while the flatter slope of the MP2/aTZ thermal expansion curve is in better qualitative agreement with the smaller experimental volume. Unfortunately, the predictions here do not appear reliable enough to adjudicate among the different reported experimental volumes.

Overall, the thermal expansion results discussed here have a few important general implications. First, the quasi-harmonic approximation reproduces the thermal expansion fairly well in small, mostly rigid molecule crystals like these. Basis sets of at least triple-zeta quality appear to be necessary to achieve fairly good volume agreement with experiment, though the larger basis sets do slightly underestimate the volumes in carbon dioxide and ice.

Second, unit cell volumes predicted by minimizing the electronic energy clearly differ substantially from those observed experimentally at room temperatures. Based on the crystals examined here and other studies,[222,223] the volume expansion lies in the range of ∼5–10%, especially if one also considers the zero-point expansion.

Third, it is common to benchmark models by comparing predictions against low-temperature crystal structures.[224] The results here highlight crystal structure optimizations which neglect zero-point vibrational energy should actually be a few percent smaller in volume than the low-temperature experimental structures. In practice, small double-zeta basis sets have often been used in fragment-based model crystal structure optimizations. Assuming the pairwise interactions are counterpoise-corrected for basis set superposition error as in the calculations here, double-zeta basis sets consistently underbind the molecules (see Section 4.5.2) and overestimate the unit cell volume. This fortuitously leads to partial error cancellation between the small basis set and the neglect of zero-point contributions.

4.5.2 Enthalpy of Sublimation

The previous section demonstrates that one can predict temperature-dependent volumes for simple crystals consisting of small, rigid molecules fairly well using a quasi-harmonic approximation. The next question is how important is treating thermal expansion for predicting other crystal properties at finite temperatures? Given the general interest in predicting polymorph/phase stability, we focus on the enthalpies (this section), entropies, and Gibbs free energies (following sections) of sublimation here.

Figure 4.3 plots the predicted temperature-dependent sublimation enthalpies for each of the four crystals against the experimental values. The left column (Figures 4.3a–d) reports the enthalpies obtained by optimizing the crystal structure and evaluating the electronic energy with a given level of theory (e.g. MP2/aTZ), and it compares the results with (solid lines) and without (dotted lines) the quasi-harmonic approximation. The right column (Figures 4.3e–h) plots the same enthalpies, except with the electronic energy U_{el} at each data point replaced with a single-point energy on that structure evaluated using ex-

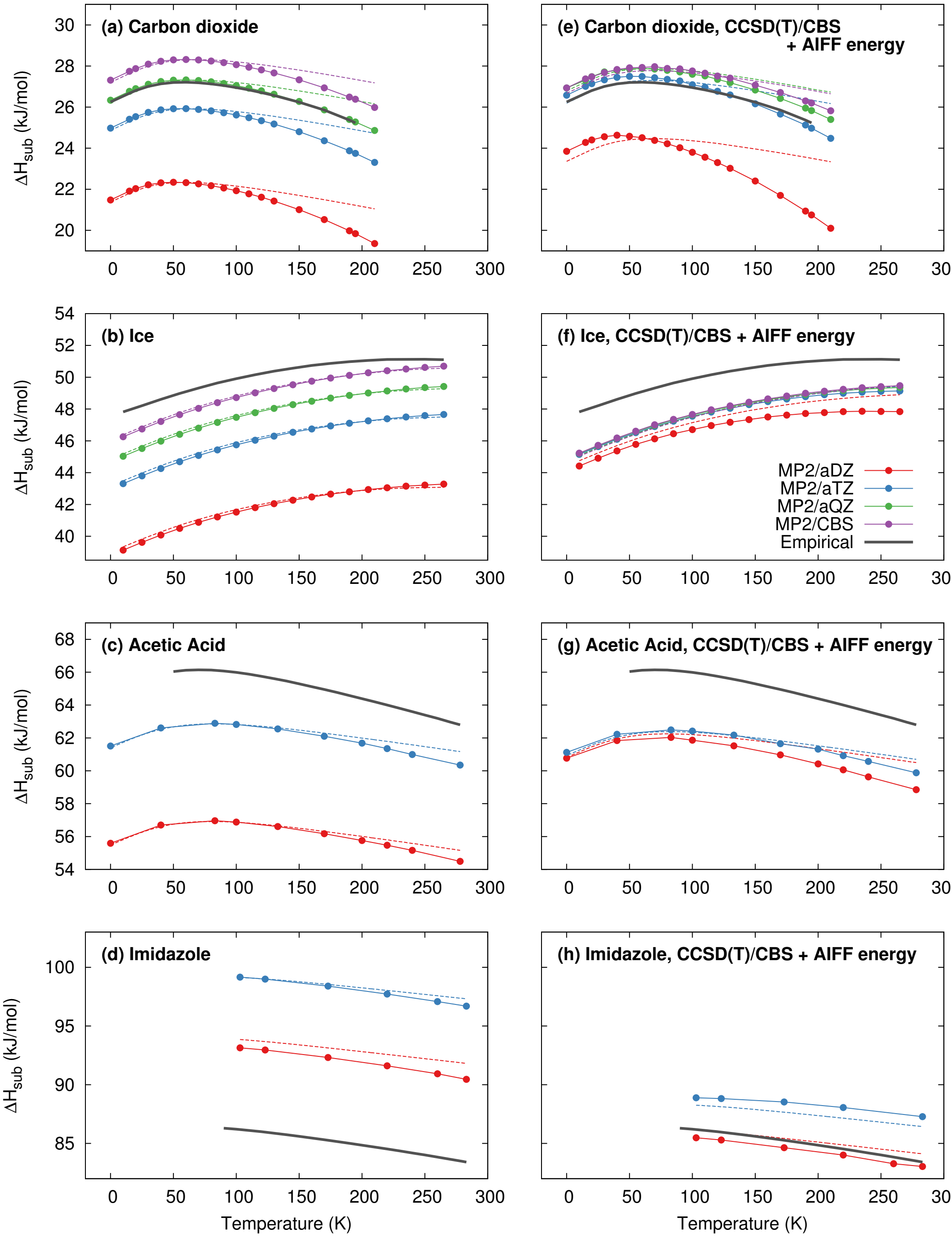

Figure 4.3: Predicted enthalpies of sublimation at 1 atm. Figures on the left use the same level of theory (e.g. MP2/aXZ + Amoeba) to optimize the structure and compute the sublimation enthalpy. Those on the right replace the lattice energy with single-point energies computed using CCSD(T)/CBS + AIFF. Curves drawn with solid lines include quasi-harmonic thermal expansion, while dotted lines neglect it.

trapolated CCSD(T)/CBS QM and AIFF force field MM contributions (instead of Amoeba MM).

Consider first Figures 4.3a–d. As discussed previously,[93] large-basis sets and the quasi-harmonic approximation are important for capturing the correct temperature-dependence of the sublimation enthalpy in carbon dioxide. In the absence of the quasi-harmonic approximation, the enthalpy decreases too slowly at high temperatures. Fortuitously, quasi-harmonic MP2/aQZ reproduces the experimental carbon dioxide enthalpies almost perfectly, while MP2/CBS and CCSD(T)/CBS slightly overestimate them by ∼1 kJ/mol.

The use of large basis sets is similarly important for the sublimation enthalpies in the other crystals. The largest jump occurs between the aDZ and aTZ basis sets, though the effects of larger basis sets still account for several kJ/mol. This matches earlier observations for lattice energies. Like carbon dioxide, increasing basis set size improves the agreement of the sublimation enthalpy with experiment for ice and acetic acid. In imidazole, however, MP2 substantially overestimates the lattice energy,[44,45] which translates to a significant overestimation of the sublimation enthalpy. Agreement with the experimental enthalpy actually deteriorates with increasing basis set size in imidazole.

Though significant basis-set dependence is observed in Figures 4.3a–d, it is unclear to what extent this behavior reflects changes in the optimized crystal structure versus differences in the lattice energy. To decouple these two effects, Figures 4.3e–h re-compute the lattice energies U_{el} for every geometries using CCSD(T)/CBS plus AIFF MM terms. Doing so dramatically reduces the differences in sublimation enthalpy across the structures

Table 4.2: Comparison between the electronic lattice energy and the quasi-harmonic sublimation enthalpy based on CCSD(T)/CBS + AIFF energies, in kJ/mol.

Crystal	Structure	$E_{lattice}{}^a$	$\Delta H_{sub}(T)^b$	Difference
Carbon Dioxide	CCSD(T)/CBS	29.7	26.3 (190 K)	4.1 (16%)
Ice	MP2/CBS	58.9	49.5 (265 K)	9.5 (19%)
Acetic acid	MP2/aTZ	66.0	59.9 (278 K)	6.2 (10%)
Imidazole	MP2/aTZ	91.6	87.3 (283 K)	4.3 (5%)

[a] Purely electronic lattice energy at the electronic energy minimum structure.
[b] Quasi-harmonic sublimation enthalpy at the highest temperature considered for each crystal.

optimized with various basis sets. For ice and carbon dioxide, the structures computed with MP2/aTZ or better produce CCSD(T)/CBS sublimation enthalpies that are nearly indistinguishable. For all four crystals, enthalpies computed on the MP2/aDZ structures are only moderately worse, though MP2/aDZ does not capture the temperature dependence as well at higher temperatures.

For all temperatures, the CCSD(T) accuracy on the larger-basis structures here lies within ~1 kJ/mol of the experimental sublimation enthalpies for carbon dioxide and within ~3–4 kJ/mol for the other three crystals. Most notably, using CCSD(T) instead of MP2 for U_{el} also corrects a substantial portion of the error in the imidazole sublimation enthalpies, as one might expect from earlier studies of the lattice energy.[44,45] As discussed earlier, errors in the experimentally-derived sublimation enthalpies themselves are also probably up to several kJ/mol.

It is interesting to consider the difference between the lattice energy as computed without consideration of temperature and the finite-temperature sublimation enthalpy. Table 4.2 compares the CCSD(T)/CBS + AIFF lattice energy at the electronic energy minimum structure (using the largest-basis structure optimization) to the sublimation enthalpy at the maximum temperature computed here. In all four crystals, the lattice energy is 4–10 kJ/mol larger than the sublimation enthalpy, which amount to up to 20% error. Zero-point vibrational and finite-temperature effects clearly have a substantial impact on the thermodynamic stability of the crystal.

Finally, how important is thermal expansion? The answer depends on the degree of thermal expansion that occurs in the crystal. From Table 4.1, the amount of thermal expansion increases according to:

$$\text{ice} < \text{imidazole} < \text{acetic acid} < \text{CO}_2$$

As seen in Figure 4.3, the importance of quasi-harmonic thermal expansion behaves similarly. The differences between the enthalpies of ice predicted with and without thermal expansion are barely observable. Including thermal expansion when modeling imidazole leads to only a slight improvement in the slope of the enthalpy curve. For acetic acid and carbon dioxide, however, including thermal expansion clearly improves agreement with the experimental sublimation enthalpies. Without thermal expansion, the sublimation enthalpy in those species decreases too slowly at high temperatures. On the other hand, it is clear that the quantitative effects of thermal expansion on the enthalpies here are relatively small at ∼1–2 kJ/mol.

To summarize, temperature-dependent enthalpies of sublimation for these four crystals can be predicted to within a few kJ/mol. Large-basis sets (and sometimes post-MP2 correlation) are important for the lattice energy component. On the other hand, a triple-zeta basis is probably sufficient for the geometry optimization. The effects of thermal

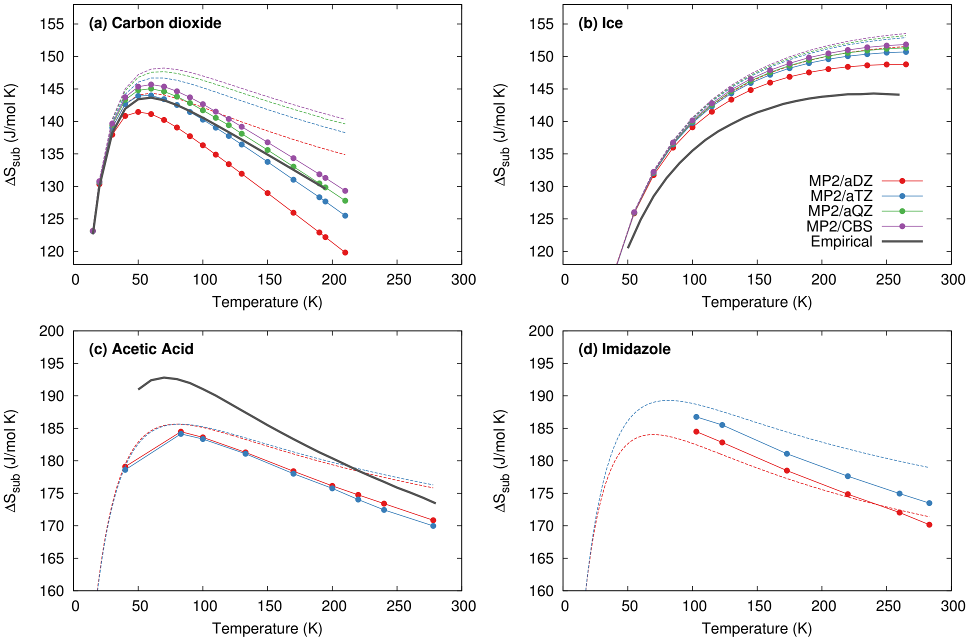

Figure 4.4: Predicted entropy of sublimation at 1 atm. Curves drawn with solid lines include quasi-harmonic thermal expansion, while dotted lines neglect it. Insufficient experimental data was available to derive an empirical sublimation entropy for imidazole.

expansion on the sublimation enthalpies are modest—comparable to or smaller than the underlying errors in the predicted sublimation enthalpies. Unsurprisingly, thermal expansion becomes more important at high temperatures, especially in crystals that exhibit high thermal expansivity.

4.5.3 Entropy of Sublimation

Next we examine the predicted entropies of sublimation, which are plotted in Figure 4.4. As before, the solid lines correspond to the entropies predicted when including quasi-harmonic thermal expansion, while the dotted lines neglect thermal expansion. Note that we were unable to locate sufficient data to derive sublimation entropies for imidazole.

The carbon dioxide entropies are identical to those reported previously.[93]

Two main features are notable in Figure 4.4. First, the sublimation entropies of carbon dioxide and imidazole both exhibit modest basis-set dependence, while the basis-set dependence is much smaller for ice and acetic acid. In the latter two cases, the individual frequencies do exhibit the normal, expected variations with basis set, but these variations largely disappear in the summed vibrational entropy contributions.

Second, including thermal expansion clearly improves the agreement of the predicted sublimation entropies with the empirical values derived from experiment. Most notably, thermal expansion improves the slopes of the entropies at higher temperatures. Unsurprisingly, the impact of thermal expansion on the sublimation entropies is largest in the crystals which expand the most—carbon dioxide, acetic acid and imidazole. In carbon dioxide near the sublimation point (195 K), the MP2/CBS sublimation entropy is overestimated by 7% when thermal expansion is neglected. In acetic acid at 278 K, neglecting thermal expansion increases the sublimation entropy by 4% at the MP2/aTZ level. In imidazole, the MP2/aTZ error is 3% at 283 K. Viewed another way, neglecting thermal expansion in these cases introduces an error of about 1.5–2 kJ/mol in $T\Delta S$ near room temperature (or near the sublimation point in the case of carbon dioxide). That corresponds to about half the magnitude of the error in the CCSD(T)/CBS sublimation enthalpies. For ice, in contrast, the small degree of thermal expansion alters the MP2/CBS sublimation entropy by only 1% at 265 K, or $\sim$0.5 kJ/mol in $T\Delta S$.

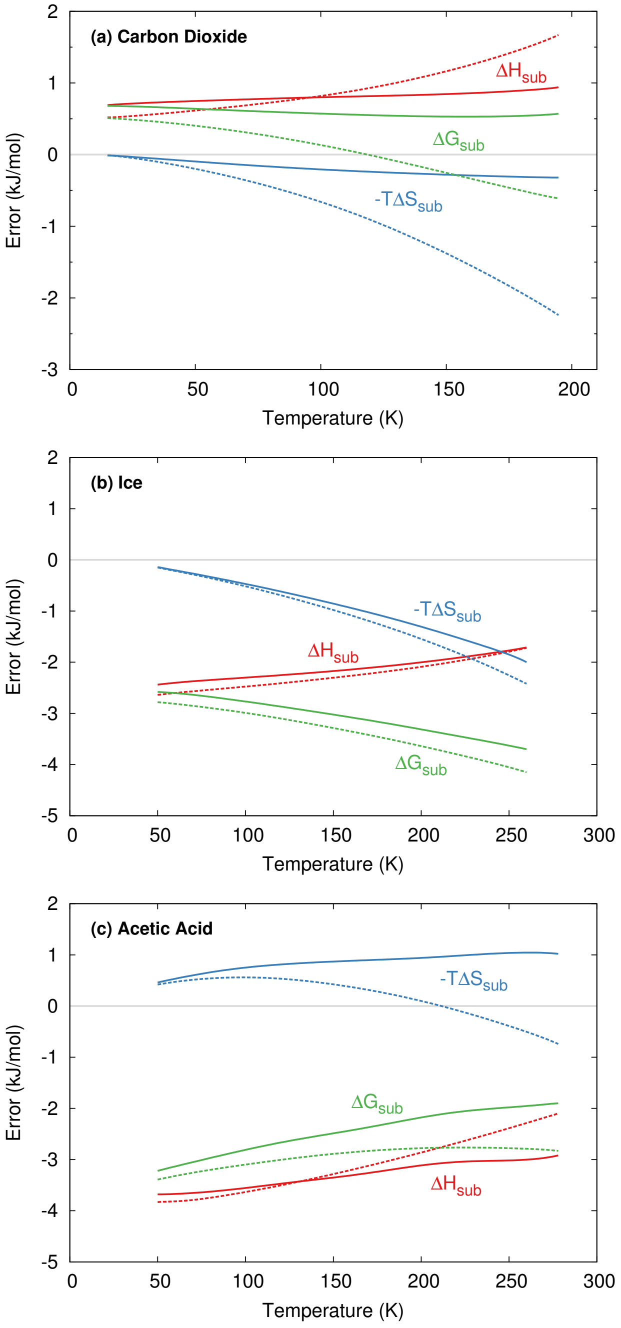

Figure 4.5: Estimated errors in the predicted enthalpies, entropies, and Gibbs free energies of sublimation relative to experiment. Curves were generated by smoothing and splining the available experimental and predicted values. Predictions with (solid lines) and without (dotted lines) quasi-harmonic expansion are shown.

4.5.4 Gibbs Free Energy of Sublimation

Finally, one might combine the enthalpies and entropies of sublimation to predict the Gibbs free energy and determine phase stability. Given the sparsity of temperatures with enthalpy and entropy data for each crystal (especially experimentally), Gibbs free energy curves were generated as a function of temperature by smoothing and splining over the available data. Based on these curves, Figure 4.5 plots the errors between the predicted and empirical sublimation enthalpies, entropies (as $-T\Delta S$), and free energies. The enthalpies here are the CCSD(T)/CBS + AIFF results using the MP2/CBS geometries (or MP2/aTZ for acetic acid). The entropies used were computed at the same level as the geometry optimizations. Imidazole is excluded from this analysis due to the lack of experimental entropy data.

For the three crystals shown here, the sublimation free energy errors range from $\sim$0.5 kJ/mol in carbon dioxide to up to $\sim$4 kJ/mol in ice. For carbon dioxide and acetic acid, the errors in ΔH and $-T\Delta S$ have opposite signs and cancel somewhat when combined into the free energy. For ice, on the other hand, the errors have the same sign, and the error in the free energy is larger.

Focus now on the error introduced by neglecting thermal expansion. Figure 4.5 makes it clear that thermal expansion is more important for the entropy than the enthalpy at high temperatures. However, the errors due to ignoring thermal expansion in carbon dioxide and acetic acid partially cancel between the enthalpy and entropy at higher temperatures. The resulting differences in the free energies with and without thermal expansion are $\sim$1 kJ/mol at the highest temperatures modeled. For ice, the thermal expansion errors in ΔH

and $-T\Delta S$ do not cancel in ΔG for the temperature range considered here, but they remain small (< 0.5 kJ/mol).

To understand the cancellation of errors associated with thermal expansion, consider the temperature dependence of the enthalpy and entropy of sublimation and how crystal expansion affects them. Typically, the sublimation enthalpy initially rises with temperature before reaching a maximum and turning over. This behavior is readily apparent for carbon dioxide, acetic acid, and imidazole in Figure 4.3. For ice, this turnover is subtle and difficult to see in the plots—the experimental enthalpy decreases by less than 0.1 kJ/mol between its maximum around 240 K and 270 K.

In terms of internal energy, the crystal is more stable than the gas due to its favorable intermolecular interactions. Thermal contributions act to destabilize both phases. The temperature dependence of the gas-phase enthalpy is dominated by the translational and rotational contributions ($4RT$), while for the crystal it is dominated by destabilization of the lattice energy and the low-frequency mode vibrational contributions. Thermal expansion weakens the intermolecular interactions (hence the change in lattice energy). It can also soften the intermolecular lattice phonon modes (though the phonon density of states is not strictly correlated with packing density[30]), which slightly stabilizes the crystal due primarily to reduction in zero-point energy. In the crystals exhibiting appreciable thermal expansion here, the changes in enthalpy are dominated by the decrease in lattice energy rather than the vibrational contributions.

At low temperatures, heating destabilizes the gas more quickly than the crystal. At higher temperatures, the thermal destabilization of the crystal becomes more pronounced relative to the gas, creating the observed maximum in the sublimation enthalpy. If thermal expansion is neglected, however, the lattice energy does not decrease as it should. Accordingly, neglecting thermal expansion causes the turnover in the sublimation enthalpy to occur too slowly at high temperatures, and ΔH_{sub} is overestimated.

The entropy of the gas is much higher than that of the crystal, leading to a large, positive ΔS_{sub}. The sublimation entropy also exhibits a temperature dependence similar to that of the enthalpy, with a maximum at intermediate temperatures. Once again, the entropy of the gas increases more rapidly at low temperatures, but at higher temperatures the low-frequency modes of the crystal dominate and cause the crystal entropy to rise faster than that of the gas. This transition produces the turnover observed in the sublimation entropies (Figure 4.4). Without thermal expansion, the lattice phonon frequencies are typically too large, and their entropic contribution to the crystal too small. Thus, ΔS_{sub} turns over too slowly. Note too that whereas thermal expansion generally destabilizes the enthalpy of the solid (unfavorable), it increases the entropy (favorable). In other words, thermal expansion is driven by entropy rather than enthalpy.

In the end, neglecting thermal expansion typically leads to overestimation of both ΔH_{sub} (because the solid is bound too strongly) and ΔS_{sub} (because the entropy of the solid is underestimated) These enthalpy and entropy contributions enter the Gibbs free energy with opposite signs, meaning that these errors arising from the neglect of thermal expansion cancel somewhat. The $\sim$0.5–1 kJ/mol free energy errors associated with neglecting thermal expansion are smaller than the size of the overall errors in the sublimation free energies. On the other hand, about half the polymorph pairs in the Nyman and Day survey were separated by less than 2 kJ/mol (and many of those by <1 kJ/mol).[30] In other words, despite the small magnitude of these errors, thermal expansion might prove important for polymorphs with very similar free energies.

Finally, we use the computed Gibbs free energies to predict the sublimation point of carbon dioxide at 1 atm. We previously demonstrated[93] (1) a strong basis set dependence in the sublimation temperature, ranging from 163.6 K with MP2/aDZ to 199.2 K for MP2/CBS, and (2) that CCSD(T)/CBS + Amoeba calculations with thermal expansion slightly overestimate the sublimation point at 201.0 K, versus 194.7 K experimentally. Despite overestimating both the enthalpy and entropy of sublimation, the CCSD(T)/CBS model without thermal expansion predicts a sublimation point of 194.9 K due to fortuitous error cancellation.

Here, Table 4.3 revises these predictions by performing CCSD(T)/CBS + AIFF single-point energies on structures optimized with each level of theory. This both applies a uniformly high-level electronic structure model to the QM electronic energy terms in all cases, and it replaces the force field with the more accurate AIFF model that includes estimates for Axilrod-Teller-Muto three-body dispersion. As expected from Figure 4.3e, replacing the smaller-basis MP2 electronic energies with single-point CCSD(T)/CBS ones calculation substantially improves the sublimation enthalpies and eliminates much of the apparent basis set dependence in the predicted sublimation temperatures. The quasi-harmonic temperature predictions now vary from 190.6 K (MP2/aDZ structures) to 198.0 K (CCSD(T)/CBS structures).

The refined force field model also slightly decreases the sublimation enthalpies (due

Table 4.3: Predicted sublimation temperatures for phase I carbon dioxide using various electronic structure methods for the structure optimization/phonons and CCSD(T)/CBS + AIFF single point energies.

Structure	T_{sub} (K)	ΔH_{sub}(194.7 K) (kJ/mol)	ΔS_{sub}(194.7 K) (J/mol K)
		No Thermal Expansion	
MP2/aDZ	192.0	26.1	135.7
MP2/aTZ	191.8	26.7	139.1
MP2/aQZ	192.0	27.0	140.3
MP2/CBS	190.3	26.9	141.2
CCSD(T)/CBS	190.4	26.9	141.1
		With Thermal Expansion	
Structure	T_{sub} (K)	ΔH_{sub}(194.7 K) (kJ/mol)	ΔS_{sub}(194.7 K) (J/mol K)
MP2/aDZ	190.6	23.3	122.2
MP2/aTZ	197.2	25.3	127.7
MP2/aQZ	198.3	25.9	129.8
MP2/CBS	198.7	26.3	131.3
CCSD(T)/CBS	198.0	26.2	131.3
Giauque and Egan[178]	194.7	25.2	129.6

to the repulsive three-body dispersion), reducing the sublimation temperature by a few degrees Kelvin relative to the earlier work. Now both models with and without quasi-harmonic expansion lie within 3–4 K of the experimental sublimation temperature of 194.7 K. For comparison, a few degree change in the sublimation temperature corresponds to changing the sublimation enthalpy by ~0.5 kJ/mol, which is smaller than the level of accuracy one can reasonably expect from the models. On the other hand, the quasi-harmonic model predicts the sublimation temperature reliably through accurate predictions of both the enthalpy and entropy, while the calculations without thermal expansion rely on large error cancellations between the overestimated enthalpies and entropies of sublimation.

4.6 Conclusions

In the end, the results here demonstrate that thermal expansion does indeed affect molecular crystal properties, though its significance will depend on the specific crystal and application. Unsurprisingly, the unit cell volume is most notably affected by thermal expansion. For the four crystals considered here, expansion of 5-10% is predicted between the electronic energy structures and structures near room temperature. A sizable fraction of the crystal expansion arises from zero-point vibrational energy. Accordingly, caution should be taken when citing agreement between predicted electronic energy crystal structures which omit zero-point energy and low-temperature experimental structures which include it. Crystal properties that depend strongly on the distances and orientations of molecules, such as charge transport in organic semiconductor materials, will also likely be affected by these changes in unit cell volume with temperature.

For thermochemistry, neglecting thermal expansion leads to errors of up to a few kJ/mol. Errors in the free energy tend to be smaller than those in the enthalpy and entropy due to error cancellation, but they can still be of the same order of magnitude as the stability differences between many crystal polymorphs. Additional error cancellation might occur when comparing free energies between two different polymorphs, though the extent of that cancellation would likely depend on how similarly the crystals expand with temperature. In any case, for crystals in which the thermal expansion is appreciable and the energetic separations between polymorphs is small, neglecting thermal expansion might lead to an incorrect stability ordering.

Finally, the results here suggest that MP2/aug-cc-pVTZ provides a reasonable level of theory for optimizing crystal structures and calculating phonon modes, except in cases like imidazole where MP2 is known to have problems describing the van der Waals dispersion interactions. The smaller aug-cc-pVDZ basis set tends to overestimate the unit cell volume significantly. The improvements in geometries offered by quadruple-zeta basis sets or larger are small relative to the increase in computational costs. On the other hand, using larger basis sets is important for achieving quantitative accuracy in the lattice energy. Finally, the quasi-harmonic approach provides an effective tool for modeling thermal expansion and free energies, at least for small rigid molecules like the ones studied here. The performance of the quasi-harmonic model for conformationally flexible molecules requires future investigation.

Chapter 5

Toward predicting the carbon dioxide phase diagram

5.1 Outline

The polymorph phase diagram for carbon dioxide remains controversial. This chapter describes preliminary efforts to model the phase diagram and potentially resolve some of the open questions. We particularly focus on the phase transitions between phases I–III. At the HMBI MP2/aug-cc-pVTZ level, the phase I/III transition is predicted to occur 6.8–10.1 GPa, depending on the temperature. Phase II is predicted to be less stable than phase I and III until pressures above 25 GPa. However, it is widely believed that phase III is only metastable relative to phase II, in contrast to the predictions here. Further study is needed.

5.2 Introduction

In the previous chapters, the significance of using thermal expansion when determining thermodynamic properties, especially for phase I carbon dioxide, was demonstrated. It was established that HMBI, with the quasi-harmonic approximation, has the ability to determine sublimation enthalpy and entropy as functions of temperature. The next question is to what extent can the model be used to determine the relative stability of a compound's polymorphs at different temperatures and pressure?

Carbon dioxide exhibits multiple solid phases/polymorphs (as shown by a by Figure 5.1) with vastly different structures. Phase I carbon dioxide, or dry ice, is the longest known and most studied of any of the polymorphs.[21,31,35,36,50,57,146,174,177,178,183,225–228] It belongs to the cubic Pa$\bar{3}$ space group[228] and is a molecular crystal. A second polymorph of carbon dioxide, phase II, was first discovered by Liu in 1983.[229] This polymorph was assigned a space group of P4_2/mnm by Yoo et al.[187] There is a controversy surrounding the nature of phase II carbon dioxide. Yoo et al claims that phase II carbon dioxide has elongated C=O bond distance and a large bulk modulus. From this, they proposed phase II as an intermediate phase between the molecular phase (such as phase I and III) and the covalent phases (such as phase V and the amorphous phases).[187] More recent findings of Datchi et al did not match those of Yoo et al. They found bond lengths similar to that of gas phase carbon dioxide and much smaller bulk modulus, which would indicate that phase II is molecular rather than intermediate in nature.[230]

Phase I carbon dioxide distorts into phase III when compressed under pressure at room temperature. This phase was first discovered by Harson in 1985[231] but detected by Liu a year earlier.[183] While it is traditionally thought that this phase transition occurs around 11.8 GPa at room temperature[183], it may occur anywhere between 7-15 GPa.[226] Phase III carbon dioxide transition to phase II by heating but phase II does not transition back to phase III upon quenching. This has lead many to believe that phase III is a metastatic state relative to phase II.[46] In general, the kinetic nature and considerable hysteresis observed with many of the phase transitions makes precise determination of the phase boundaries experimentally challenging.[47,48]

Phase III was assigned a Cmca space group by Aoki et al.[190] Interestingly, computational studies of phase III have reported the lattice lengths where a > b with experimental results finding the opposite.[36,50,51,227] Giordano and Datchi discovered another Cmca phase which they labeled phase VII.[232] The lattice parameters for this Cmca phase closely match those predicted by theory and exist in the pressure-temperature range predicted by Bonev et al.[36] This brings into question whether the Cmca structure appearing in theoretical papers is phase VII instead of phase III. Here we will refer to the Cmca structure as phase III.

While there are other polymorphs of carbon dioxide, this chapter will focus on phases I, II, and III of carbon dioxide. Most of the other polymorphs are non-molecular and HMBI is a poor model for non-molecular crystals. The structure of phase IV, which is either molecular[233] or molecular/covalent intermediate[234], is in question[233,235] so it was left out of this study. The Gibbs free energy of each crystal was predicted at various pressure and temperature to determine which polymorph is the thermodynamically favored form. The unit cell of all crystals included are on Figure 5.2.

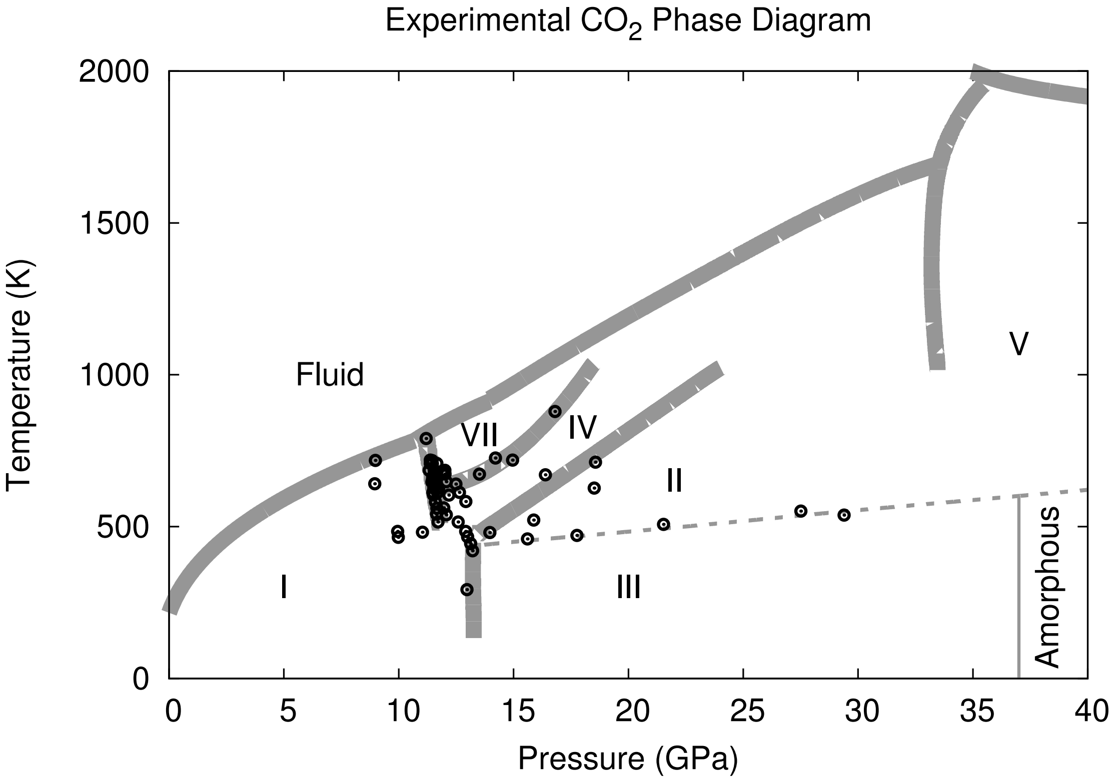

Figure 5.1: Phase diagram for carbon dioxide. Open dots are reported at Ref[46,225,232] Dotted lines between phase II and III is generally thought to be a kinetic barrier rather than a thermodynamic one . Fluid/solid phase transition report in Ref 225. Rough phase III/amorphous transition reported at Ref 226.

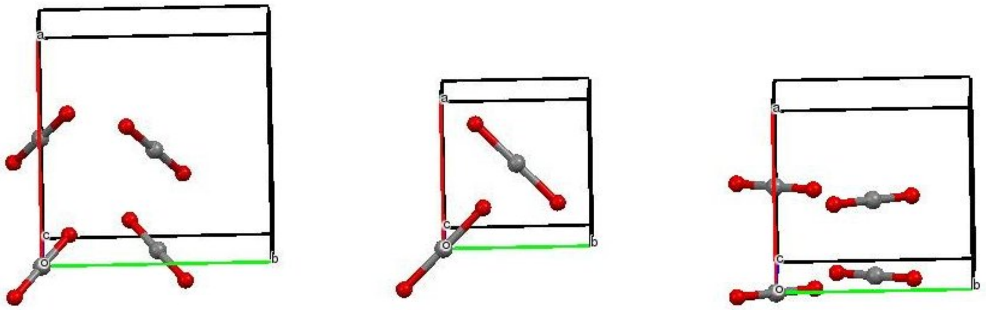

Figure 5.2: The structure for phase I–Pa$\bar{3}$ (left), phase II–P4$_2$/mnm (middle) and phase III/VII–Cmca (right) of carbon dioxide

5.3 Theory and Computational Method

The structure of the crystal unit cell were determined by minimizing the Gibbs free energy.

$$G(T,P) = U_{el} + PV + F_{vib}(T) \tag{5.1}$$

The quasi-harmonic approximation as detailed by Chapters 3 and 4 was utilized. The relative thermodynamical favorability of a polymorph was determined by normalizing the crystal by the number of monomers (i.e. dividing it by the number of monomers in the cell).

In Chapter 4, structures optimized at the MP2/aTZ + Amoeba level proved particularly effective, especially when the single-point electronic energy was replaced with CCSD(T)/CBS + AIFF. Therefore the same approach was used here. The frequencies for the quasi-harmonic approximation were evaluated using a $3 \times 3 \times 3$ Monkhorst-Pack grid on a $3 \times 3 \times 3$ supercell. Both Phase I and Phase III used 10 Å^3 finite difference steps to determine the Grüneisen parameters. Phase II used 5 Å^3 finite difference steps since its unit cell is approximately half the volume and has half the number of monomer of the other cells.

Because the phase II unit cell has half the number of monomers in the unit cells, it has significantly fewer vibrational frequencies in the $3 \times 3 \times 3$ Monkhorst-Pack grid. While the Gibb's free energy was normalized by number of monomers, there was concern that the difference between the number vibrational frequencies of the unit cell may create a bias in the Holmholtz vibrational energy. A 1 × 1 × 2 cell for phase II was created. The Holmholtz vibrational energy was determined for a 3 × 3 × 3 supercell of this cell using a 3 × 3 × 3 Monkhorst-Pack grid. Once the Gibbs free energy was normalized for the number of monomers, there was little energy difference (0.006 kJ/mol) between the 2 × 1 × 1 cell and the unit cell. This indicates that at the current Monkhorst-Pack scheme, Holmholtz vibrational energy is converged.

5.4 Results and Discussion

All normalized Gibbs free energy on Figure 5.3 were set relative to the value for phase I at the same temperature and pressure. By definition, phase I has a zero Gibbs free energy at each temperature and pressure. At each temperature, the Gibbs free energy vs pressure values are virtually linear. A thermodynamic phase transition occurs at the intersection of two curves. At 0 K, the phase I/III intersection occurs at 10.2 GPa, and the transition pressure decreases with increasing temperature. At 298 K, the transition is about 8.5 GPa which is lower than the generally accepted 11.8 GPa.[183] Shieh et al found that the phase I→III transition occurs slowly with pressure, observing a mixture of phase I/III between 7-15 GPa and a 50/50 percent composition between 8-9 GPa.[226] Our results fit well with their findings.

Our calculations predict phase III to be more stable than phase II until 29-34 GPa, depending on the temperature. At 0 K, the phase II/III transition was extrapolated to 34.0 GPa, which decreases to 28.6 GPa at 700 K. In other words, phase III is predicted to be thermodynamically stable over a broad region of the phase diagram, contrary to the

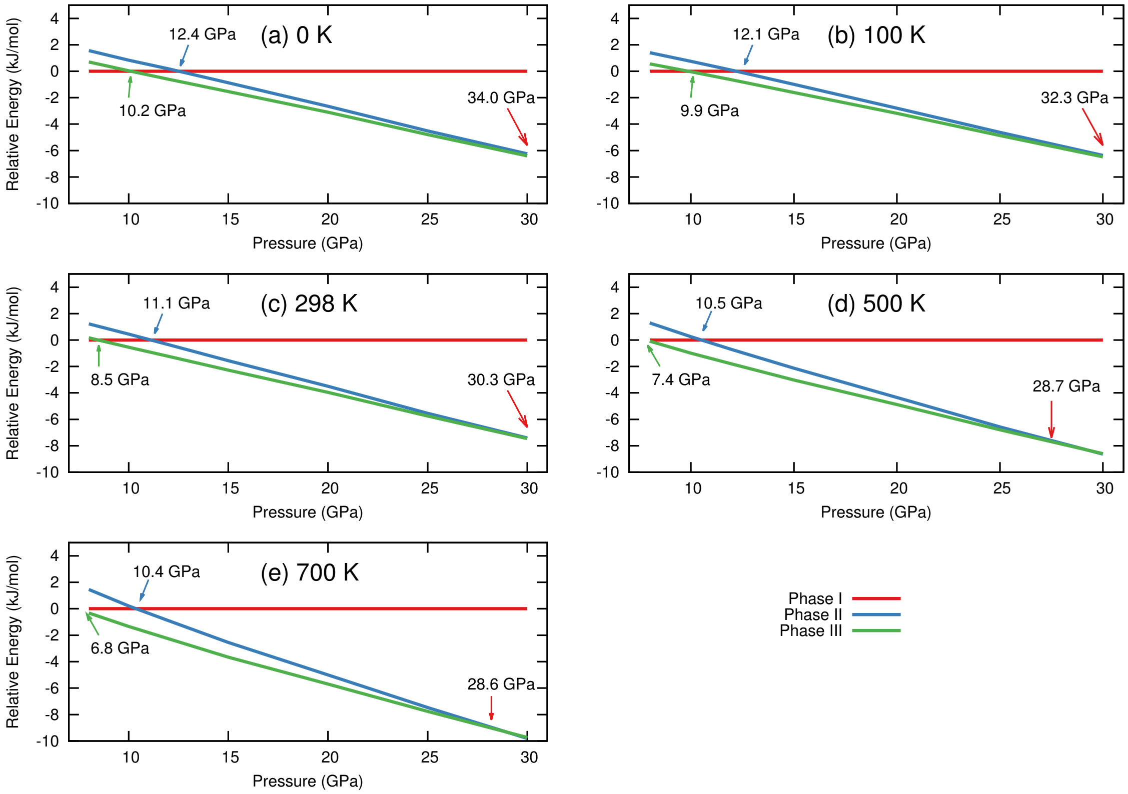

Figure 5.3: The Gibbs free energy phases I, II, III relative to phase I at the same temperature and pressure. The arrows marks the pressure in which the free energy lines intersect and phase transition occurs. Some of the intersections occurred outside the predicted pressure range. In these cases, the intersection was extrapolated.

conventional notion that phase III is experimentally metastable relative to phase II in this region. The figure diagram predicted from these results is on Figure 5.4.

The free energy difference between phases II and III in this region are only 1 kJ/mol, which is on par with the 1–2 kJ/mol errors in the phase I carbon dioxide sublimation energies predicted in Chapter 3. This might suggest that it will be difficult to resolve these free energy differences theoretically. On the other hand, one might hope for partial error cancellation in the free energy difference between two similar phases like these. The

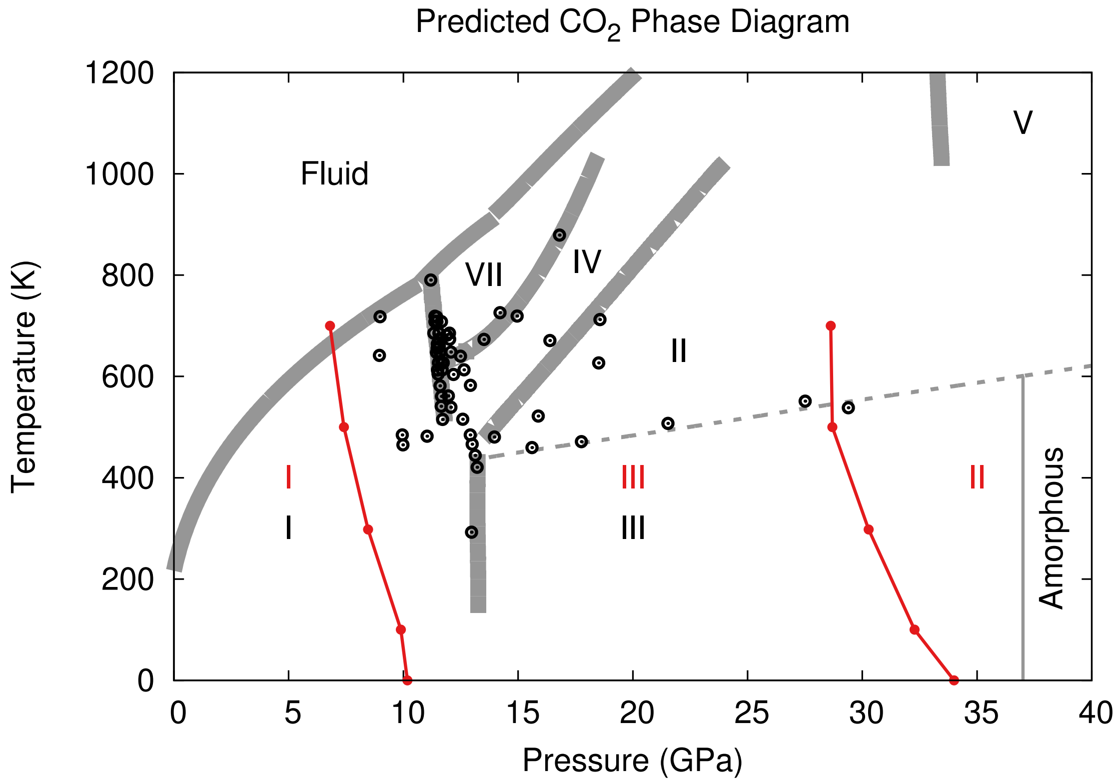

Figure 5.4: Predicted HMBI CCSD(T)/CBS + AIFF phase boundaries overlaid on the experimental phase diagram for carbon dioxide.

predicted unit cell volume could also be an important factor. At high pressures like these, small differences in unit cell volume can lead to sizable contributions in the PV term that appears in the enthalpy. Agreement between the predicted and experimental structures is quite good here, at least for phase I and II. The structure of phase III presents a more interesting set of questions, as discussed below.

Phase III and phase VII have an identical space group (Cmca). Like most other theoretical studies found in the literature, we predicted phase III lattice parameters closer to phase VII than III. Notably, phase III and VII differ in the lengths of the a and b lattice parameters. Our predicted lattice parameters are in much better agreement with those determined experimentally for phase VII. If the Cmca modeled is phase VII rather than phase III, this may help explain why the Cmca phase is not metastable relative to phase II as is thought. This does not explain why this Cmca phase is predicted to be the preferred polymorph in the pressure-temperature region that phase II is observed. One possible solution to this controversy surrounding the nature of phase II. If phase II carbon dioxide is a molecular/covalent intermediate phase as suggested by Yoo, it may be difficult for HMBI to model. HMBI builds fragments based on molecules. Therefore HMBI is bias toward molecular systems.

Ignoring phase III, the phase I/II transition is about 10-12 GPa depending on the temperature. Assuring the phase III is a metastable state, the phase I/II transition should occur approximately around phase I/III transition. As shown by Figure 5.5, the predict value fits perfectly within the experiment transition.

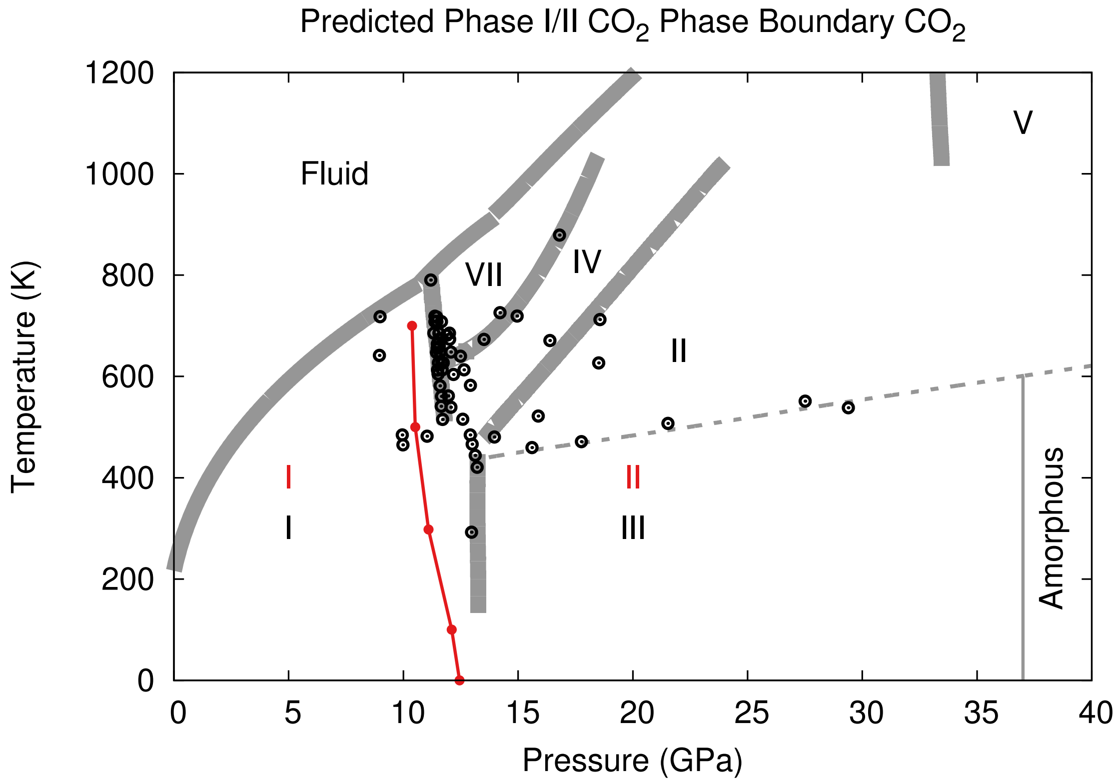

Figure 5.5: The phase I-II phase boundaries overlaid on the experimental phase diagram. Phase III was ignored.

5.4.1 Conclusion

HMBI predicts reasonable phase boundaries between phases I/III and phases I/II but not between phases II/III. There is some question about whether the structure used to model phase III is appropriate and how it differs from phase VII. However, even if the structure modeled actually corresponds to phase VII, it is not obvious that it should be stable in the pressure/temperature region where phase II is observed. One possible explanation for the difficulties could be that phase II may actually be intermediate in nature as suggested by Yoo et al, which could make the HMBI a poor choice for modeling this system. However, Yoo's result is experimentally controversial, and DFT calculations also predict a molecular phase II.[35,36] More research is needed to clarify the structures of these phases and their thermodynamic stabilities.

Chapter 6

Conclusions

Accurate and computationally affordable methods for modeling molecular crystals will play an important role in fields such as pharmaceutical chemistry, energetic materials, and organic semi-conductors. Fragment-based methods like HMBI provide a powerful tool for achieving these goals. This dissertation has detailed two theoretical advances that substantially improve the range of problems that can be tackled with HMBI:

1. An algorithm for accelerating HMBI calculations by exploiting space group symmetry to eliminate redundant monomer and dimer fragment calculations.
2. The treatment of finite temperature effects in crystals by coupling HMBI and the quasi-harmonic approximation.

With these tools, the effects of thermal expansion on molecular crystal properties were examined, first for carbon dioxide and then for several other crystals. The computational savings reaped by exploiting space group symmetry were instrumental in these finite-temperature studies.

For phase I carbon dioxide, neglecting thermal expansion leads to underestimation of the molar volume, overestimation of the enthalpy and entropy of sublimation, and overestimation of the bulk modulus. The importance of thermal expansion helped explain why previous studies had had difficulty predicting the bulk modulus. Of course, carbon dioxide is somewhat atypical for molecular crystals. It is small and very rigid. It is bound relatively weakly, and it undergoes an unusually large degree of thermal expansion 8% between 15 K and 190 K.[57] Therefore, similar studies were repeated on ice, acetic acid, and imidazole crystals. Thermal expansion has little impact in the enthalpy and entropy of sublimation for Ih ice but this is not surprising considering that only expands by about 2% as the temperature raises from 10 K to 265 K.[58] On the other hand, both acetic acid and imidazole show appreciable thermal expansion that affects the predictions of their properties.

Accurate thermochemistry predictions are particularly relevant to molecular crystal polymorphism problems. Based on these thermal expansion studies, we conclude that quasi-harmonic HMBI with high-level electronic structure methods can predict sublimation enthalpies to within several kJ/mol of experiment. Neglecting thermal expansion can introduce additional errors of a couple kJ/mol. The errors in the room-temperature entropies are similar in magnitude. However, a sizable fraction of the errors in the enthalpy and entropy cancel in the Gibbs free energy. In the end, it appears that neglecting thermal expansion leads to errors of 1–2 kJ/mol in the Gibbs free energy at room temperature. Free energy gaps of $\sim$1 kJ/mol or less occur moderately often between crystal polymorphs, so these errors due to neglecting thermal expansion may be important in determining polymorph stabilities.

The next phase of this research is to extend these finite-temperature and pressure modeling capabilities toward the prediction of molecular crystal phase diagrams. Preliminary results for the phase I, II, and III region of carbon dioxide were presented. Plausible boundaries for phase I/III and phase II/III were found, but the phase II/III boundary does not appear in good agreement with experiment. Phase III is believed to be metastable relative to phase II, but the predictions here suggest the opposite. A number of outstanding questions regarding the phase diagram of carbon dioxide complicate the interpretation of these results. First, there remains debate in the literature about whether phase II is a molecular crystal or an molecular/covalent intermediate phase. If it is the latter, a fragment-based method based on the many body expansion may not be an appropriate model. Ambiguity also surrounds the structure of phase III versus phase VII, and whether the structure used to model phase III here is correct.

In the longer term, these tools will hopefully enable more advanced molecular crystal modeling in a variety of problems. Beyond the polymorphism and phase diagram problems discussed here, these models may prove useful in predicting other crystal properties which are sensitive to unit cell volume. Charge transport properties, for example, can vary by large amounts upon small changes in packing density. Nuclear magnetic resonance (NMR) chemical shifts are also sensitive to the fine details of crystal packing. Crystal structure prediction, NMR chemical shift prediction, and solid-state NMR experiments are often combined to solve crystal structures. The ability to predict room-temperature structures instead of 0 K ones might increase the discrimination with which one can identify the correct structure out of a set of candidate structures.

Right now, the main bottleneck to employing quasi-harmonic HMBI is the evaluation of the Hessian elements for the vibrational frequencies. Future research should explore possibilities for utilizing more approximate models to compute the phonons. Such models might include the use of small basis sets, DFT, or even MM potentials. In addition, all of the systems considered here consist of largely rigid molecules. Future work is needed to explore the performance of the quasi-harmonic model in crystals of conformationally flexible molecules.

Appendix A

Symmetry for crystals with high symmetry point group molecules

In Chapter 2, a simple algorithm for implementing space group symmetry was described. While the details given there will work for most crystals, that algorithm becomes problematic for crystals that contain molecules that belong to high symmetry point groups. The algorithm assumes that each atom on a symmetrically unique monomer is also symmetrically unique and, therefore, in the asymmetric unit. If these atoms are treated as independent degrees of freedom during a crystal structure optimization, space group symmetry may be broken. For such crystals, the symmetrical equivalence among atoms in symmetrically unique monomers needs to be exploited using point group symmetry to preserve the space group symmetry. Here we are defining a high symmetry molecule as any molecule that contains atoms that are symmetrically equivalent. Examples of high-symmetry crystals include phase I, II, and III carbon dioxide, phase I benzene, phase I urea, and phase II and IV ammonium nitrate. Note the details in this chapter assume that the lattice angles are 90° and that there is a 1 to 1 mapping between x, y, and z Cartesian and fractional coordinates. This will have to be modified for crystal unit cells whose lattice angles are not 90°.

A.1 Atom Rotation and Translation Vectors

The molecular point group is determined using known point group operators on a symmetrical unique monomer in its standard nuclear orientation (for information on standard nuclear orientation see Section 2.3.1). Once the point group is identified, the operators are applied to each atom individually. If the coordinates match those of another atom, two atoms are symmetrically equivalent. The list of symmetrically unique atoms are stored similarly to how it is for the monomers and dimers. In principle, the only atoms left should be the ones in the asymmetric unit. The rotational operator mapping atom p to atom q on monomer 1 in monomer 1's center of mass coordinates (not standard nuclear orientation) is obtained from matrix $\mathbf{R}^{q \leftarrow p}$,

$$\mathbf{R}^{q \leftarrow p} = \mathbf{R}_1^T O^{q,p} \mathbf{R}_1 \tag{A.1}$$

where operator $O^{q,p}$ is the point group operation that maps atom p to q in monomer 1's standard nuclear orientation and $\mathbf{R}_1$ is the rotation matrix rotating monomer 1's to its standard nuclear orientation. Atoms p and q should have a space group operation demonstrating their symmetrical equivalence along with a point group operator.

$$\mathbf{q}_2 = \mathbf{R}\mathbf{q}_1 + \mathbf{t} \tag{A.2}$$

The rotational matrix matches that of a point group operator. The translational vector $\mathbf{t}^{q \leftarrow p}$ is given by the vector difference between the atoms after the rotation of p is performed with respect to the origin of the global coordinate system.

The rotational matrix found in Section 2.3.1 mapping monomer 2 to monomer 1 may map atom r on monomer 2 to atom p on monomer 1 instead of atom q in the asymmetric unit. In order to map r to q, one must take the product of that rotational matrix mapping monomer 2 to 1 and the matrix mapping atom p to q.

$$\mathbf{R}^{q \leftarrow r} = \mathbf{R}^{q \leftarrow p}\mathbf{R}^{2 \leftarrow 1} \tag{A.3}$$

A translational vector mapping r to q on different monomers can be similarly found as the translational vector mapping p to q on the same monomer.

A.2 Preserving Fractional Symmetry

In many of these high symmetry point groups, relative coordinates of certain atoms in the molecule cannot be altered. For example, carbon dioxide is a linear molecule and carbon is equidistant from both oxygens. The molecule belongs to the point group $\mathrm{D}_{\infty h}$. If the carbon moves towards either of the oxygen the point group changes to $\mathrm{C}_{\infty v}$. In order to maintain point group symmetry, the carbon must be fixed. The same principle is true for space group symmetry. Phase 1 carbon dioxide has a carbon at fractional coordinate (0, 0, 0). The space group is Pa$\bar{3}$ and one of the operators in this space group is inversion. When this operator is applied, the carbon stays in the same location. If the carbon shifts in any direction the space group symmetry is broken. Therefore the location of a carbon has to be fixed during optimization. In order to preserve the space group symmetry, atoms may have to be fixed at a particular fractional coordinate. Note that often an atom is not entirely fixed; perhaps only one or two of the fractional coordinates are fixed. For an atom's fractional coordinate to be fixed it must satisfy one of the following criteria:

1. The atom's corresponding Cartesian coordinate is zero when using the monomer's center of mass coordinates
2. An atom has a fractional coordinate of exactly 0, 1/6, 1/5, 1/4, 1/3, 1/2, or any of these number multiplied by a positive or negative integer

In addition to fixing fractional coordinates of atoms, the linkage between the x, y, and z coordinates of an atom must be considered. Consider the oxygen in phase I carbon dioxide bonded to the carbon at (0, 0, 0). It is at 45° angles relative to the 3 lattice vectors. In the fractional and Cartesian coordinates, the x, y, and z are the same. The space group has a few operators where the fractional x, y, and z coordinates are switched. When these operators are applied, the oxygen atoms is left in the same location. If the oxygen shifts so that any of the angles are no longer 45° these coordinates no longer match and space group symmetry is broken. Therefore the x, y and, z coordinates are linked. These coordinates are turned into a single degree of freedom. The criteria for x, y, and/or z coordinates of an atom to be turned into a single degree of freedom are:

1. If any of an atom's Cartesian coordinates have the same absolute value using the monomers center of mass coordinate.
2. If the sum or difference of any atom's fractional coordinates is exactly 0, 1/6, 1/5, 1/4, 1/3, 1/2, or any of these number multiplied by a positive or negative integer.

In most cases, either both or neither of these criteria is true. Note that all three coordinates of an atom can be transformed to a single degree of freedom. In this case, the sum or difference of all the fractional coordinates do not have to be a certain number, only the two individually. For example, if an atom's fractional coordinates are (0.567, 0.433, 1.567). The difference of the x and z fractional coordinates coordinate is 1 and the sum of y and z is also 2. Both x and y are linked to z and they are treated to a single degree of freedom, but their sum or difference is no number of interest.

A.3 Optimizing Structure While Preserving Symmetry

Even though a fractional coordinate are may be fixed by symmetry, the corresponding Cartesian coordinates are not. When the lattice parameters are relaxed, the Cartesian coordinates can be shifted to maintain space group symmetry. In order to fix fractional coordinates without fixing Cartesian positions, we must introduce the shift vector $\mathbf{s}_q$. The shift vector is also used when the fraction x, y, and/or z coordinates of an atom are linked but must maintain a certain sum or difference in order to preserve symmetry. Once the linked x, y, and/or z coordinate single degree of freedom is updated by the optimizer, adding the shift vector multiplied by the change in lattice vectors as a column matrix will maintain symmetry.

$$(\Delta \mathbf{v}_1 \Delta \mathbf{v}_2 \Delta \mathbf{v}_3)\mathbf{s}_q \tag{A.4}$$

The entries of the shift vector are the determined by the follow criteria:

1. If an atom's x, y, or z fractional coordinate is exactly 0, 1/6, 1/5, 1/4, 1/3, 1/2, or any of these number multiplied by a positive or negative integer, that number (times the multiplied integer) is placed into the corresponding shift vector entry.

2. If the sum of any atom's fractional coordinate is exactly 0, 1/6, 1/5, 1/4, 1/3, 1/2, or any of these number multiplied by a positive or negative integer, the sum is placed into an entry of the shift vector. If the x and y sum is this number, that sum is placed into the shift vector's x entry. If the x and z sum is this number, that sum is placed into the shift vector's x entry. If the y and z sum is this number, that sum is placed into the shift vector's y entry.

3. If the difference of any atom's fractional coordinates is exactly 0, 1/6, 1/5, 1/4, 1/3, 1/2, or any of these number multiplied by a positive or negative integer, that difference is placed into shift vector. If the x and y difference is this number, that difference is placed into the shift vector's x entry. If the x and z difference is this number, that difference is placed into the shift vector's x entry. If the y and z difference is this number, that difference is placed into the shift vector's y entry.

Note that only an atom coordinate that is not a degree of freedom or that is linked to another coordinate on the same atom can have a non-zero value in its entry in the shift vector. The shift vector for an atom not in the asymmetric cell is determined by rotating the shift vector using the rotation matrix mapping its symmetrically equivalent atom in the asymmetric cell to it. During the optimization, the following steps are taken for high symmetry point group molecules.

1. Determine the number of symmetrically unique monomers as described by Chapter 2.
2. Determine atoms in the asymmetric unit from the atoms on the symmetrically unique monomers.
3. Determine which atoms in the asymmetric unit have to be fixed or have their x, y, and/or z fractional coordinates linked. No fixed atom is a degree of freedom for optimization. Linked x, y, and/or z coordinates are treated as a single degree of freedom.
4. Determine shift vector for atoms in asymmetric cell.
5. Determine shift vector for atoms not in asymmetric cell using rotation matrices.
6. Determine number of symmetrical unique dimers described by Chapter 2.
7. Get gradient and coordinates from optimizer for every every degree of freedom for the next optimization step.
8. Use linkage to update x, y and/or z for atoms in asymmetric unit based on the coordinate for single degree of freedom given by the optimizer for the next optimization step.
9. Use shift vector to update non-degrees of freedom or linked coordinates based on the lattice parameters given for the next optimization step in order to maintain space

group symmetry.

10. Apply space groups operators to the remaining coordinates for the atoms not in the asymmetric unit.

11. Repeat steps 6-10 until crystal is relaxed.

Note that this method has not been tested for crystals with high symmetry point group molecule and have lattice angles other than 90°. Therefore it is assumed that there is a one-to-one relationship between x, y, and z fractional and Cartesian coordinates.

A.4 Gradient for High Symmetry Point Groups

The nuclear gradient is mostly unchanged from the nuclear gradient in Section 2.3.4.

$$\begin{aligned}
\frac{\partial E^{HMBI}_{crystal}}{\partial q^z_l} &= \frac{E^{MM}_{crystal}}{\partial q^z_l} + \sigma_z \left(\frac{\partial E^{QM}_z}{\partial q^z_l} - \frac{\partial E^{MM}_z}{\partial q^z_l} \right) \\
&\quad + {\sum_{i<j}}' \sigma_{ij} d_{ij} \left(\frac{\partial \Delta^2 E^{QM}_{ij}}{\partial q^z_l} - \frac{\partial \Delta^2 E^{MM}_{ij}}{\partial q^z_l} \right) \\
&\quad + {\sum_{i<j}}' \sigma_{ij} \frac{\partial d_{ij}}{\partial q^z_l} \left(\Delta^2 E^{QM}_{ij} - \Delta^2 E^{MM}_{ij} \right) \\
&\quad + \frac{1}{2} {\sum_{ik}}' \sigma_{ik} d_{ik} \left(\frac{\partial \Delta^2 E^{QM}_{ik}}{\partial q^z_l} - \frac{\partial \Delta^2 E^{MM}_{ik}}{\partial q^z_l} \right) \\
&\quad + \frac{1}{2} {\sum_{ik}}' \sigma_{ik} \frac{\partial d_{ik}}{\partial q^z_l} \left(\Delta^2 E^{QM}_{ik} - \Delta^2 E^{MM}_{ik} \right) \qquad \text{(A.5)}
\end{aligned}$$

Once the gradient for all the atoms in the symmetrically unique monomers are determined, the gradient of atoms not in the asymmetric unit are mapped to atoms in the asymmetric unit using the atom rotation matrices.

$$\frac{\partial E}{\partial q_l^z} = \sum_p \sum_{l'} R_{ll'}^{q\leftarrow p} \left(\frac{\partial E}{\partial p_{l'}^z}\right) \tag{A.6}$$

The gradient for fixed atoms coordinates are zeroed. Any atoms with linked x, y, or z coordinates combine their gradient into a single entry. The sign of the linked gradient may have to be changed before combining if two of the linked coordinates in center of mass coordinates have difference signs or if they are linked because the difference, rather than the sum, between the coordinates in a particular number.

The lattice vector gradient from Eq 2.15 now includes an additional one-body term.

$$\begin{aligned}
\frac{\partial E'}{\partial \mathrm{v}_{\epsilon l}} &= \sum_i \sum_q \frac{\partial q_l^i}{\partial \mathrm{v}_{\epsilon l}} \left(\frac{\partial E_i^{QM}}{\partial q_l^i} - \frac{\Delta^2 E_i^{QM}}{\partial q_l^i}\right) \\
&\quad + \sum_{i<j} d_{ij} \sum_q \frac{\partial q_l^{ij}}{\partial \mathrm{v}_{\epsilon l}} \left(\frac{\partial \Delta^2 E_{ij}^{QM}}{\partial q_l^{ij}} - \frac{\partial \Delta^2 E_{ij}^{QM}}{\partial q_l^{ij}}\right) \\
&\quad + \sum_{i<j} \left(\Delta^2 E_{ij}^{QM} - \Delta^2 E_{ij}^{MM}\right) \sum_q \frac{\partial q_l^{ij}}{\partial \mathrm{v}_{\epsilon l}} \frac{\partial d_{ij}}{\partial q_l^{ij}} \\
&\quad + \sum_{ik} d_{ik} \sum_q \frac{\partial q_l^{ik}}{\partial \mathrm{v}_{\epsilon l}} \left(\frac{\partial \Delta^2 E_{ik}^{QM}}{\partial q_l^{ik}} - \frac{\partial \Delta^2 E_{ik}^{QM}}{\partial q_l^{ik}}\right) \\
&\quad + \sum_{ik} \left(\Delta^2 E_{ik}^{QM} - \Delta^2 E_{ik}^{MM}\right) \sum_q \frac{\partial q_l^{ik}}{\partial \mathrm{v}_{\epsilon l}} \frac{\partial d_{ik}}{\partial q_l^{ik}}
\end{aligned} \tag{A.7}$$

The $\frac{\partial q_l^i}{\partial \mathrm{v}_{\epsilon l}}$ term is the derivative of the change in the Cartesian coordinates of atom q relative to monomer i's center of mass. This term is equal to the difference in the monomer and atom translational vector derivative with respect to the change in the lattice vector plus the entry in the shift vector.

$$\frac{\partial q_l^i}{\partial \mathrm{v}_{\epsilon l}} = \frac{\mathrm{t}_l^q}{\partial \mathrm{v}_{\epsilon l}} - \frac{\mathrm{t}_l^i}{\partial \mathrm{v}_{\epsilon l}} + \hat{}s_l^q = \hat{l}_t^q - \hat{l}_t^i + \hat{s}_l^q \tag{A.8}$$

where $\hat{l}_t^i$ and $\hat{l}_t^q$ are the elements of the fractional coordinate representation of the translational operator of monomer i and atom q respectively. The dimer $\frac{\partial q_l^{ij}}{\partial \mathrm{v}_{\epsilon l}}$ term uses the atom's fractional coordinate translational operator instead of the monomer's and includes the atom's shift vector.

$$\frac{\partial q_l^{ij}}{\partial \mathrm{v}_{\epsilon l}} = \frac{\mathrm{t}_l^q}{\partial \mathrm{v}_{\epsilon l}} + \hat{s}_l^q = \hat{l}_t^q + \hat{s}_l^q \tag{A.9}$$

Appendix B

Experimentally Derived Thermochemistry Data

The empirical enthalpies and entropies of sublimation used to validate the predictions in Chapters 3 and 4 were obtained as a function of temperature for each crystal as follows.

B.1 Carbon Dioxide

For carbon dioxide, temperature-dependent heats of sublimation at 1 atm were taken from Azreg-Aïnou.[177] The entropies of sublimation were derived according to:

$$\begin{aligned}\Delta S_{sub}^{empr}(T) &= \Delta S_{sub}^{expt}(194.7K) - \int_{194.7K}^{T} \frac{C_{p,solid}(T)}{T} dT \\ &\quad + (S_{gas}(T) - S_{gas}(194.7K))\end{aligned} \tag{B.1}$$

This expression was evaluated using the sublimation entropy at 194.7 K[178] and heat capacities for the crystal[178] and gas-phase rotational constants[179] and vibrational frequencies.[180] See Ref 93 for details.

B.2 Ice Ih

Temperature-dependent experimental enthalpies of sublimation for ice Ih were taken from the equation of state of Feistel and Wagner.[236] Entropies of sublimation were derived from experimental data according to:

$$\begin{aligned}\Delta S_{sub}^{empr}(T) &= S_{gas}(T) - S_{gas}(373.15K) + \Delta S_{vap}(373.15K) \\ &\quad + S_{liquid}(373.15K) - S_{liquid}(273.15K) + \Delta S_{fus}(273.15K) \\ &\quad + \int_{T}^{273.15K} \frac{C_{p,solid}(T)}{T} dT \end{aligned} \tag{B.2}$$

The entropies for the gas and liquid water were found on the NIST/TRC Web Thermo Tables.[237] The entropy of vaporization and fusion were computed from the enthalpy of vaporization and fusion at their transition points ($\Delta S_{fus}(289.15K) = \frac{\Delta H_{fus}(289.15K)}{289.15K}$ and $\Delta S_{vap}(391.7K) = \frac{\Delta H_{vap}(391.7K)}{391.7K}$). The enthalpies of vaporization and fusion at their transition point were taken from the Handbook of Chemistry and Physics.[238] Isobaric heat capacities from Ref 239 were converted to a smooth function using a cubic spline and integrated using Mathematica.

B.3 Acetic Acid

For acetic acid, empirical enthalpies and entropies of sublimation were determined by combining a series of reported values:

$$\begin{aligned} \Delta H_{sub}^{empr}(T) &= H_{gas}(T) - H_{gas}(298.15K) + \Delta H_{vap}(298.15K) \\ &\quad + H_{liquid}(298.15K) - H_{liquid}(289.6K) + \Delta H_{fus}(289.6K) \\ &\quad + H_{solid}(289.6K) - H_{solid}(T) \end{aligned} \tag{B.3}$$

$$\begin{aligned} \Delta S_{sub}^{empr}(T) &= S_{gas}(T) - S_{gas}(391.7K) + \Delta S_{vap}(391.7K) \\ &\quad + S_{liquid}(391.7K) - S_{liquid}(289.6K) + \Delta S_{fus}(289.6K) \\ &\quad + S_{solid}(289.6K) - S_{solid}(T) \end{aligned} \tag{B.4}$$

The gas, liquid, and solid enthalpies ($H(T)$) and entropies ($S(T)$) at various temperatures were taken from the NIST/TRC Web Thermo Tables.[237] Values for the enthalpies of vaporization[240] and fusion[241] were taken from the literature. The entropies of vaporization and fusion were determined by dividing the corresponding enthalpy at the appropriate phase transition temperature ($\Delta S_{fus}(289.6K) = \frac{\Delta H_{fus}(289.6K)}{289.6K}$, $\Delta S_{vap}(391.7K) = \frac{\Delta H_{vap}(391.7K)}{391.7K}$). Since the enthalpy of vaporization was reported at 298.15 K instead of at the boiling point of 391.7 K, it was extrapolated to 391.7 K using data from the NIST/TRC Web Thermo

Tables:

$$\begin{aligned} \Delta H_{vap}(391.7K) &= \Delta H_{vap}(298.15K) + H_{gas}(391.7) - H_{gas}(298.15K) \\ &\quad - H_{liquid}(391.7) + H_{liquid}(298.15K) \end{aligned} \tag{B.5}$$

B.4 Imidazole

Empirical sublimation enthalpies for imidazole were derived similarly to the previous crystals:

$$\begin{aligned} \Delta H_{sub}^{empr}(T) &= \Delta H_{sub}(298.15K) + H_{gas}(T) - H_{gas}(298.15K) \\ &\quad + \int_{T}^{298.15K} C_{p,solid}(T)dT \end{aligned} \tag{B.6}$$

The enthalpy of sublimation at 298.15 K was reported by Jiménez et al.[242] Isobaric heat capacities for solid imidazole[243] were converted to a smooth function via cubic spline and integrated using Mathematica. The enthalpy of gaseous imidazole at various temperatures was computed using Eqs 4.14 and 4.15. Experimental vibrational frequencies were taken from Billes et al.[244] The gas-phase electronic energy contribution was omitted because the value temperature T cancels with the one at 298.15 K. Insufficient data was found to derive the empirical entropies of sublimation for imidazole.

Bibliography

[1] Chemburkar, S.R., Bauer, J., Deming, K., Spiwek, H., Patel, K., Morris, J., Henry, R., Spanton, S., Dziki, W., Porter, W., Quick, J., Bauer, P., Donaubauer, J., Narayanan, B.A., Soldani, M., Riley, D., and Mcfarland, K. "Dealing with the Impact of Ritonavir Polymorphs on the Late Stages of Bulk Drug Process Development." *Org. Proc. Res. Dev.*, **4**, 413–417 (2000).

[2] Bauer, J., Spanton, S., Quick, R., Quick, J., Dzikim, W., Porter, W., and Morris, J. "Ritonavir: An extraordinary example of conformational polymorphism." *Pharm. Res.*, **18**, 859–866 (2001).

[3] Raw, A.S., Furness, M.S., Gill, D.S., Adams, R.C., Holcombe, F.O., and Yu, L.X. "Regulatory considerations of pharmaceutical solid polymorphism in Abbreviated New Drug Applications (ANDAs)." *Adv. Drug Deliv. Rev.*, **56**, 397–414 (2004). doi: 10.1016/j.addr.2003.10.011.

[4] Haas, S., Stassen, A.F., Schuck, G., Pernstich, K.P., Gundlach, D.J., Batlogg, B., Berens, U., and Kirner, H.J. "High charge-carrier mobility and low trap density in a rubrene derivative." *Phys. Rev. B*, **76**, 115203 (2007). doi: 10.1103/PhysRevB.76.115203.

[5] Harikrishnan, G., Ajith, K.M., Chandra, S., and Valsakumar, M.C. "Evolutionary algorithm based structure search for hard ruthenium carbides." *Modelling Simul. Mater. Sci. Eng.*, **25**, 085006 (2015). doi:10.1088/0965-0393/23/8/085006.

[6] Motherwell, W.D.S., Ammon, H.L., Dunitz, J.D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D.W.M., Leusen, F.J.J., Lommerse, J.P.M., Mooij, W.T.M., Price, S.L., Scheraga, H.A., Schweizer, B., Schmidt, M.U., van Eijck, B.P., Verwer, P., and Williams, D.E. "Crystal structure prediction of small organic molecules: a second blind test." *Acta Cryst. B*, **58**, 647–661 (2002). doi:10.1107/S0108768102005669.

[7] Day, G.M., Motherwell, W.D.S., Ammon, H.L., Boerrigter, S.X.M., Della Valle, R.G., Venuti, E., Dzyabchenko, A., Dunitz, J.D., Schweizer, B., van Eijck, B.P., Erk, P., Facelli, J.C., Bazterra, V.E., Ferraro, M.B., Hofmann, D.W.M., Leusen, F.J.J., Liang, C., Pantelides, C.C., Karamertzanis, P.G., Price, S.L., Lewis, T.C., Nowell, H., Torrisi, A., Scheraga, H.A., Arnautova, Y.A., Schmidt, M.U., and Verwer, P. "A third blind test of crystal structure prediction." *Acta. Cryst. B*, **61**, 511–527 (2005). doi: 10.1107/S0108768105016563.

[8] Day, G.M., Cooper, T.G., Cruz-Cabeza, A.J., Hejczyk, K.E., Ammon, H.L., Boerrigter, S.X.M., Tan, J.S., Della Valle, R.G., Venuti, E., Jose, J., Gadre, S.R., Desiraju, G.R., Thakur, T.S., van Eijck, B.P., Facelli, J.C., Bazterra, V.E., Ferraro, M.B., Hofmann, D.W.M., Neumann, M.A., Leusen, F.J.J., Kendrick, J., Price, S.L., Misquitta, A.J., Karamertzanis, P.G., Welch, G.W.A., Scheraga, H.A., Arnautova, Y.A., Schmidt, M.U., van de Streek, J., Wolf, A.K., and Schweizer, B. "Significant progress in predicting the crystal structures of small organic molecules–a report on the fourth blind test." *Acta Cryst. B*, **65**, 107–125 (2009). doi:10.1107/S0108768109004066.

[9] Neumann, M.A., Leusen, F.J.J., and Kendrick, J. "A major advance in crystal structure prediction." *Angew. Chem. Int. Ed.*, **47**, 2427–2430 (2008). doi: 10.1002/anie.200704247.

[10] Kazantsev, A.V., Karamertzanis, P.G., Adjiman, C.S., Pantelides, C.C., Price, S.L., Galek, P.T.A., Day, G.M., and Cruz-Cabeza, A.J. "Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction." *Int. J. Pharm.*, **418**, 168–178 (2011). doi:10.1016/j.ijpharm.2011.03.058.

[11] Kendrick, J., Leusen, F.J.J., Neumann, M.A., and van de Streek, J. "Progress in Crystal Structure Prediction." *Chem. Eur. J.*, **17**, 10736–10744 (2011). doi: 10.1002/chem.201100689.

[12] Bardwell, D.A., Adjiman, C.S., Arnautova, Y.A., Bartashevich, E., Boerrigter, S.X.M., Braun, D.E., Cruz-Cabeza, A.J., Day, G.M., Della Valle, R.G., Desiraju, G.R., van Eijck, B.P., Facelli, J.C., Ferraro, M.B., Grillo, D., Habgood, M., Hofmann, D.W.M., Hofmann, F., Jose, K.V.J., Karamertzanis, P.G., Kazantsev, A.V., Kendrick, J., Kuleshova, L.N., Leusen, F.J.J., Maleev, A.V., Misquitta, A.J., Mohamed, S., Needs, R.J., Neumann, M.A., Nikylov, D., Orendt, A.M., Pal, R., Pantelides, C.C., Pickard, C.J., Price, L.S., Price, S.L., Scheraga, H.A., van de Streek, J., Thakur, T.S., Tiwari, S., Venuti, E., and Zhitkov, I.K. "Towards crystal structure prediction of complex organic compounds–a report on the fifth blind test." *Acta Cryst. B*, **67**, 535–551 (2011). doi:10.1107/S0108768111042868.

[13] Price, S.L. "Predicting crystal structures of organic compounds." *Chem. Soc. Rev.*, **43**, 2098–2111 (2014). doi:10.1039/c3cs60279f.

[14] Reilly, A.M. and Tkatchenko, A. "Role of dispersion interactions in the polymorphism and entropic stabilization of the aspirin crystal." *Phys. Rev. Lett.*, **113**, 1–5 (2014). doi:10.1103/PhysRevLett.113.055701.

[15] Wen, S. and Beran, G.J.O. "Accidental degeneracy in crystalline aspirin: New insights from high-level ab initio calculations." *Cryst. Growth Des.*, **12**, 2169–2172 (2012). doi: 10.121/cg300358n.

[16] Wen, S. and Beran, G.J.O. "Crystal Polymorphism in Oxalyl Dihydrazide: Is Empirical DFT-D Accurate Enough?" *J. Chem. Theory Comput.*, **8**, 2698–2705 (2012). doi:10.1021/ct300484h.

[17] Hirata, S. "Fast electron-correlation methods for molecular crystals: an application to the α, β_1, and β_2 modifications of solid formic acid." *J. Chem. Phys.*, **129**, 204104 (2008). doi:10.1063/1.3021077.

[18] Hirata, S., Gilliard, K., He, X., Li, J., and Sode, O. "Ab initio molecular crystal structures, spectra, and phase diagrams." *Acc. Chem. Res.*, **47**, 2721–2030 (2014). doi:10.1021/ar500041m.

[19] Karamertzanis, P.G., Day, G.M., Kendrick, G.W.W.J., Leusen, F.J.J., Neumann, M.A., and Price, S.L. "Modeling the interplay of inter- and intramolecular hydrogen bonding in conformational polymorphs." *J. Chem. Phys.*, **128**, 244708 (2008). doi: 10.1063/1.2937446.

[20] Reilly, A.M. and Tkatchenko, A. "Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals." *J. Chem. Phys.*, **139**, 024705 (2013). doi:10.1063/1.4812819.

[21] Li, J., Sode, O., and Hirata, S. "Second-Order Many-Body Perturbation Study on Thermal Expansion of Solid Carbon Dioxide." *J. Chem. Theory Comput.*, **11**, 224–229 (2015). doi:10.1021/ct500983k.

[22] He, X., Sode, O., Xantheas, S.S., and Hirata, S. "Second-order many-body perturbation study of ice Ih." *J. Chem. Phys.*, **137**, 204505 (2012). doi:10.1063/1.4767898.

[23] Harris, R.K., Joyce, S.A., Pickard, C.J., and Cadars, S. "Assigning carbon-13 NMR spectra to crystal structures by the INADEQUATE pulse sequence and first principles computation: a case study of two forms of testosterone." *Phys. Chem. Chem. Phys.*, **8**, 137–143 (2006). doi:10.1039/b513392k.

[24] Tripon, C., Kacso, I., Miclaus, M., Filip, X., Bratu, I., and Filip, C. "Molecular Structure Elucidation of a New Anhydrous Polymorph of Acyclovir: Experimental and Computational Approach." *Rev. Chim.-Bucharest*, **65**, 657–663 (2014).

[25] Abraham, A., Apperley, D.C., Gelbrich, T., Harris, R.K., and Griesser, U.J. "NMR crystallography - Three polymorphs of phenobarbital." *Can. J. Chem.*, **89**, 770–778 (2011). doi:10.1139/V11-011.

[26] Marom, N., DiStasio, R.A., Atalla, V., Levchenko, S., Reilly, A.M., Chelikowsky, J.R., Leiserowitz, L., and Tkatchenko, A. "Many-body Dispersion Interactions in Molecular Crystal Polymorphism." *Angew. Chem. Int. Ed.*, **52**, 6629–6632 (2013). doi:10.1002/anie.201301938.

[27] Li, T. and Feng, S. "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).

[28] Neumann, M.A. and Perrin, M. "Can crystal structure prediction guide experimentalists to a new polymorph of paracetamol?" *CrystEngComm*, **11**, 2475–2479 (2009). doi:10.1039/b909819d.

[29] DiStasio, R.A., von Lilienfeld, O.A., and Tkatchenko, A. “Collective many-body van der Waals interactions in molecular systems.” *Proc. Nat. Acad. Sci.*, **109**, 14791–14795 (2012). doi:10.1073/pnas.1208121109.

[30] Nyman, J. and Day, G.M. “Static and lattice vibrational energy differences between polymorphs.” *CrystEngComm.*, **17**, 5154–5165 (2015). doi:10.1039/C5CE00045A.

[31] de-la Roza, A.O. and Johnson, E.R. “A benchmark for non-covalent interactions in solids.” *J. Chem. Phys.*, **137**, 054103 (2012). doi:10.1063/1.4738961.

[32] Otero-De-La-Roza, A., Cao, B.H., Price, I.K., Hein, J.E., and Johnson, E.R. “Predicting the relative solubilities of racemic and enantiopure crystals by density-functional theory.” *Angew. Chem. Int. Ed.*, **53**, 7879–7882 (2014). doi:10.1002/anie.201403541.

[33] Taylor, C.R., Bygrave, P.J., Hart, J.N., Allan, N.L., and Manby, F.R. “Improving density functional theory for crystal polymorph energetics.” *Phys. Chem. Chem. Phys.*, **14**, 7739–7743 (2012). doi:10.1039/c2cp24090d.

[34] Gracia, L., Marqués, M., Beltrán, A., Pendás, A.M., and Recio, J.M. “Bonding and compressibility in molecular and polymeric phases of solid CO2.” *J. Phys. Condens. Mat.*, **16**, S1263–S1270 (2004). doi:10.1088/0953-8984/16/14/038.

[35] Gohr, S., Grimme, S., Söhnel, T., Paulus, B., and Schwerdtfeger, P. “Pressure dependent stability and structure of carbon dioxide–A density functional study including long-range corrections.” *J. Chem. Phys.*, **139**, 174501 (2013). doi:10.1063/1.4826929.

[36] Bonev, S.A., Gygi, F., Ogitsu, T., and Galli, G. “High-pressure molecular phases of solid carbon dioxide.” *Phys. Rev. Lett.*, **91**, 065501 (2003). doi: 10.1103/PhysRevLett.91.065501.

[37] Brandenburg, J.G. and Grimme, S. “Dispersion Corrected Hartree-Fock and Density Functional Theory for Organic Crystal Structure Prediction.” *Top. Curr. Chem.*, **345**, 1–23 (2014). doi:10.1007/128_2013_488.

[38] Moellmann, J. and Grimme, S. “DFT-D3 Study of Some Molecular Crystals.” *J. Phys. Chem. C*, **118**, 7615–7621 (2014). doi:10.1021/jp501237c.

[39] Braun, D.E., McMahon, J.A., Koztecki, L.H., Price, S.L., and Reutzel-Edens, S.M. “Contrasting Polymorphism of Related Small Molecule Drugs Correlated and Guided by the Computed Crystal Energy Landscape.” *Cryst. Growth Des.*, **14**, 2056–2072 (2014). doi:10.1021/cg500185h.

[40] Tkatchenko, A. and Scheffler, M. “Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data.” *Phys. Rev. Lett.*, **102**, 073005 (2009). doi:10.1103/PhysRevLett.102.073005.

[41] Tkatchenko, A., DiStasio, R.A., Car, R., and Scheffler, M. “Accurate and Efficient Method for Many-Body van der Waals Interactions.” *Phys. Rev. Lett.*, **108**, 236402 (2012). doi:10.1103/PhysRevLett.108.236402.

[42] van de Streek, J. and Neumann, M.A. "Validation of experimental molecular crystal structures with dispersion-corrected density functional theory calculations." *Acta Cryst. B*, **66**, 544–558 (2010). doi:10.1107/S0108768110031873.

[43] Reilly, A.M. and Tkatchenko, A. "Seamless and accurate modeling of organic molecular materials." *J. Phys. Chem. Lett.*, **4**, 1028–1033 (2013).

[44] Beran, G.J.O. and Nanda, K. "Predicting Organic Crystal Lattice Energies with Chemical Accuracy." *J. Phys. Chem. Lett.*, **1**, 3480–3487 (2010). doi: 10.1021/jz101383z.

[45] Wen, S. and Beran, G.J.O. "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). doi:10.1021/ct200541h.

[46] Yoo, C. and Iota, V. "Phase Diagram of Carbon Dioxide: Evidence for a New Associated Phase." *Phys. Rev. Lett.*, **86**, 5922–5925 (2001).

[47] Santoro, M. and Gorelli, F.A. "High pressure solid state chemistry of carbon dioxide." *Chem. Soc. Rev.*, **35**, 918–931 (2006). doi:10.1039/b604306m.

[48] Yoo, C., Sengupta, A., and Kim, M. "Phase diagram of carbon dioxide: update and challenges." *High Press. Res.*, **31**, 68–74 (2011). doi:10.1080/08957959.2010.523000.

[49] Schmerler, S. and Kortus, J. "Ab initio study of AlN: Anisotropic thermal expansion, phase diagram, and high-temperature rocksalt to wurtzite phase transition." *Phys. Rev. B*, **89**, 064109 (2014).

[50] Li, J., Sode, O., Voth, G.A., and Hirata, S. "A solid-solid phase transition in carbon dioxide at high pressures and intermediate temperatures." *Nature Commun.*, **4**, 2647 (2013). doi:10.1038/ncomms3647.

[51] Li, J., Sode, O., Voth, G.A., and Hirata, S. "A solid-solid phase transition in carbon dioxide at high pressures and intermediate temperatures." *Nature Commun.*, **6**, 8907 (2015). doi:10.1038/ncomms9907.

[52] Allan, N.L., Barrera, G.D., Barron, T.H.K., and Taylor, M.B. "Evaluation of Thermodynamic Properties of Solids by Quasiharmonic Lattice Dynamics." *Int. J. Thermophys.*, **22**, 535–546 (2001). doi:10.1023/A:1010783000570.

[53] Ramírez, R., Neuerburg, N., Fernández-Serra, M.V., and Herrero, C.P. "Quasiharmonic approximation of thermodynamic properties of ice Ih, II, and III." *J. Chem. Phys.*, **137**, 044502 (2012). doi:10.1063/1.4737862.

[54] Cramer, C.J. *Essentials of Computational Chemistry: Theories and Models*. John Wiley & Sons, second edition (2004).

[55] Jensen, F. *Introduction to Computational Chemistry*. John Wiley & Sons, second edition (2007).

[56] Monkhorst, H.J. and Pack, J.D. "Special points for Brillonin-zone integrations." *Phys. Rev. B*, **13**, 5188 (1976).

[57] Krupskii, I.N., Prokhvatilov, A.I., Erenburg, A.I., and Barylnik, A.S. "Thermal expansion X-ray studies of solid CO_2." *Fiz. Nizk. Temp.*, **8**, 533–541 (1982).

[58] Petrenko, V.F. and Whitworth, R.W. *Physics of ice.* Oxford University Press (1999).

[59] Dove, M.T. *Introduction to lattice dynamics.* Cambridge University Press (1993).

[60] Szabo, A. and Ostlund, N.S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory.* Dover, Mineola, NY (1996).

[61] Sinnokrot, M.O. and Sherrill, C.D. "Highly accurate coupled cluster potential energy curves for the benzene dimer: Sandwich, T-shaped, and parallel-displaced configurations." *J. Phys. Chem. A*, **108**, 10200–10207 (2004).

[62] Riley, K.E., Pitonak, M., Jurecka, P., and Hobza, P. "Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories." *Chem. Rev.*, **110**, 5023–5063 (2010). doi: 10.1021/cr1000173.

[63] Hesselmann, A. "Improved supermolecular second order Møller-Plesset intermolecular interaction energies using time-dependent density functional response theory." *J. Chem. Phys.*, **128**, 144112 (2008).

[64] Huang, Y., Shao, Y., and Beran, G.J.O. "Accelerating MP2C dispersion corrections for dimers and molecular crystals." *J. Chem. Phys.*, **138**, 224112 (2013). doi: 10.1063/1.4809981.

[65] Pitonak, M. and Hesselmann, A. "Accurate Intermolecular Interaction Energies from a Combination of MP2 and TDDFT Response Theory." *J. Chem. Theory Comput.*, **6**, 168–178 (2010). doi:10.1021/ct9005882.

[66] Kristyan, S. and Pulay, P. "Can (semi)local density functional theory account for the London dispersion forces?" *Chem. Phys. Lett.*, **229**, 175–180 (1994).

[67] Grimme, S. "Density functional theory with London dispersion corrections." *WIRES: Comput. Mol. Sci.*, **1**, 211–228 (2011). doi:10.1002/wcms.30.

[68] Antony, J. and Grimme, S. "Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules." *Phys. Chem. Chem. Phys.*, **8**, 5287–5293 (2006). doi:10.1039/b612585a.

[69] Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu." *J. Chem. Phys.*, **132**, 154104 (2010). doi:10.1063/1.3382344.

[70] Ambrosetti, A., Reilly, A.M., DiStasio, R.a., and Tkatchenko, A. "Long-range correlation energy calculated from coupled atomic response functions." *J. Chem. Phys.*, **140**, 18A508 (2014). doi:10.1063/1.4865104.

[71] R. Salmon-Ferrer, D.A.C. and Walker, R.C. "An overview of Amber biomolecular simulation package." *WIREs Comput. Mol. Sci.*, **3**, 198–210 (2011). doi:10.1002/wcms.1121.

[72] Stone, A.J. *The Theory of Intermolecular Forces.* Clarendon Press, Oxford (2002).

[73] Thole, B.T. "Molecular Polarizabilities Calculated With A Modified Dipole Interaction." *Chem. Phys.*, **59**, 341–350 (1981). doi:10.1016/0301-0104(81)85176-2.

[74] Halgren, T.A. "Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters." *J. Am. Chem. Soc.*, **114**, 7827–7843 (1991).

[75] Ponder, J.W., Wu, C., Ren, P., Pande, V.S., Chodera, J.D., Schnieders, M.J., Haque, I., Mobley, D.L., Lambrecht, D.S., DiStasio, R.A., Head-Gordon, M., Clark, G.N.I., Johnson, M.E., and Head-Gordon, T. "Current Status of the AMOEBA polarizable force field." *J. Phys. Chem. B*, **114**, 2549–2564 (2010). doi:10.1021/jp910674d.

[76] Kennedy, M.R., McDonald, A.R., DePrince, A.E., Marshall, M.S., Podeszwa, R., and Sherrill, C.D. "Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory." *J. Chem. Phys.*, **140**, 121104 (2014). doi:10.1063/1.4869686.

[77] Yang, J., Hu, W., Usvyat, D., Matthews, D., Schutz, M., and Chan, G.K.L. "Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy." *Science*, **345**, 640–643 (2014). doi:10.1002/14356007.

[78] Ladd, M.F.C. *Symmetry in molecules and crystals.* Ellis Horwood Halsted Press, Chichester, West Sussex, England New York (1989). ISBN 9780853122555.

[79] Gordon, M.S., Fedorov, D.G., Pruitt, S.R., and Slipchenko, L. "Fragmentation Methods: A Route to Accurate Calculations on Large Systems." *Chem. Rev.*, **112**, 632–672 (2012). doi:10.1021/cr200093j.

[80] Wen, S., Nanda, K., Y.Huang, and Beran, G.J.O. "Practical quantum mechanics-based fragment methods for predicting molecular crystal properties." *Physical chemistry chemical physics : PCCP*, **14**, 7578–7590 (2012). doi:10.1039/c2cp23949c.

[81] Raghavachari, K. and Saha, A. "Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules." *Chem. Rev.*, **115**, 5643–5677 (2015). doi:10.1021/cr500606e.

[82] Richard, R.M. and Herbert, J.M. "A generalized many-body expansion and a unified view of fragment-based methods in electronic structure theory." *J. Chem. Phys.*, **137**, 064113 (2012). doi:10.1063/1.4742816.

[83] Mayhall, N.J. and Raghavachari, K. "Many-Overlapping-Body (MOB) Expansion: A Generalized Many Body Expansion for Nondisjoint Monomers in Molecular Fragmentation Calculations of Covalent Molecules." *J. Chem. Theory Comput.*, **8**, 2669–2675 (2012). doi:10.1021/ct300366e.

[84] Toukmaji, A.Y. and Board, J.A. "Ewald summation techniques in perspective: a survey." *Comput. Phys. Commun.*, **95**, 73–92 (1996).

[85] Sebetci, A. and Beran, G.J.O. "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). doi:10.1021/ct900545v.

[86] Beran, G.J.O. "Approximating quantum many-body intermolecular interactions in molecular clusters using classical polarizable force fields." *J. Chem. Phys.*, **130**, 164115 (2009). doi:10.1063/1.3121323.

[87] Nanda, K. and Beran, G.J.O. "Improved prediction of organic molecular crystal geometries from MP2-level fragment QM/MM calculations." *J. Chem. Phys.*, **137**, 174106 (2012). doi:10.1063/1.4764063.

[88] Nanda, K. and Beran, G.J.O. "What governs the proton-ordering in ice XV?" *J. Phys. Chem. Lett.*, **4**, 3165–3169 (2013). doi:10.1021/jz401625w.

[89] Ben, M.D., Vandevondele, J., and Slater, B. "Periodic MP2, RPA, and Boundary Condition Assessment of Hydrogen Ordering in Ice XV." *J. Phys. Chem. Lett.*, **5**, 4122–4128 (2014). doi:10.1021/jz501985w.

[90] Hartman, J.D. and Beran, G.J.O. "Fragment-Based Electronic Structure Approach for Computing Nuclear Magnetic Resonance Chemical Shifts in Molecular Crystals." *J. Chem. Theory Comput.*, **10**, 4862–4872 (2014). doi:10.1021/ct500749h.

[91] Hartman, J.D., Monaco, S., Schatschneider, B., and Beran, G.J.O. "Fragment-based 13C nuclear magnetic resonance chemical shift predictions in molecular crystals: An alternative to planewave methods." *J. Chem. Phys.*, **143**, 102809 (2015). doi:10.1063/1.4922649.

[92] Heit, Y. and Beran, G.J.O. "Exploting space group symmetry in fragment-based molecular crystal calculations." *J. Comp. Chem.*, **35**, 2205–2214 (2014). doi:10.1002/jcc.23737.

[93] Heit, Y.N., Nanda, K.D., and Beran, G.J.O. "Predicting finite-temperature properties of crystalline carbon dioxide from first principles with quantitative accuracy." *Chem. Sci.*, **7**, 246–255 (2016). doi:10.1039/C5SC03014E.

[94] Heit, Y. N. and Beran, G. J. O. "How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures?" Submitted (2016).

[95] Singhal, D. and Curatolo, W. "Drug polymorphism and dosage form design: a practical perspective." *Adv. Drug Deliv. Rev.*, **56**, 335–347 (2004). doi: 10.1016/j.addr.2003.10.008.

[96] Datta, S. and Grant, D.J.W. "Crystal structures of drugs: advances in determination, prediction and engineering." *Nature reviews. Drug discovery*, **3**, 42–57 (2004). doi: 10.1038/nrd1280.

[97] Schön, J., Kloc, C., Laudise, R., and Batlogg, B. "Electrical properties of single crystals of rigid rodlike conjugated molecules." *Phys. Rev. B*, **58**, 12952–12957 (1998). doi:10.1103/PhysRevB.58.12952.

[98] Siegrist, T., Kloc, C., Schön, J.H., Batlogg, B., Haddon, R.C., Berg, S., and Thomas, G.A. "Enhanced Physical Properties in a Pentacene Polymorph." *Angew. Chem. Int. Ed.*, **40**, 1732–1736 (2001).

[99] Jurchescu, O., Mourey, D., Subramanian, S., Parkin, S., Vogel, B., Anthony, J., Jackson, T., and Gundlach, D. "Effects of polymorphism on charge transport in organic semiconductors." *Phys. Rev. B*, **80**, 21–23 (2009). doi:10.1103/PhysRevB.80.085201.

[100] Sode, O. and Hirata, S. "Second-order many-body perturbation study of solid hydrogen fluoride under pressure." *Phys. Chem. Chem. Phys.*, **14**, 7765–7779 (2012). doi:10.1039/C2CP40236J.

[101] Manby, F.R., Alfè, D., and Gillan, M.J. "Extension of molecular electronic structure methods to the solid state: computation of the cohesive energy of lithium hydride." *Phys. Chem. Chem. Phys.*, **8**, 5178–5180 (2006). doi:10.1039/b613676a.

[102] Bygrave, P.J., Allan, N.L., and Manby, F.R. "The embedded many-body expansion for energetics of molecular crystals." *J. Chem. Phys.*, **137**, 164102 (2012). doi: 10.1063/1.4759079.

[103] Gillan, M.J., Alfe, D., Bartok, A.P., and Csányi, G. "First-principles energetics of water clusters and ice: A many-body analysis." *J. Chem. Phys.*, **139**, 244504 (2013). doi:10.1063/1.4852182.

[104] Gillan, M.J., Alfè, D., Bygrave, P.J., Taylor, C.R., and Manby, F.R. "Energy benchmarks for water clusters and ice structures from an embedded many-body expansion." *J. Chem. Phys.*, **139**, 114101 (2013). doi:10.1063/1.4820906.

[105] Neill, D.P.O., Allan, N.L., and Manby, F.R. "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).

[106] Podeszwa, R., Rice, B.M., and Szalewicz, K. "Predicting Structure of Molecular Crystals from First Principles." *Phys. Rev. Lett.*, **101**, 115503 (2008). doi: 10.1103/PhysRevLett.101.115503.

[107] Hermann, A. and Schwerdtfeger, P. "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**, 183005 (2008). doi: 10.1103/PhysRevLett.101.183005.

[108] Hermann, A. and Schwerdtfeger, P. "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).

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

[110] Tsuzuki, S., Orita, H., Honda, K., and Mikami, M. "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).

[111] Fedorov, D.G. and Kitaura, K. "Extending the power of quantum chemsitry to large systems with the fragment molecular orbital method." *J. Phys. Chem. A*, **111**, 6904–6914 (2007).

[112] Hahn, T., editor. *International Tables for Crystallography, Vol. A: Space-group symmetry.* Springer, Heidelberg (2005).

[113] Brock, C.P. and Dunitz, J.D. "Towards a grammar of crystal packing." *Chem. Mater.*, **6**, 1118–1127 (1994).

[114] Born, M. and Huang, K. *Dynamical Theory of Crystal Lattices.* Clarendon Press (1954).

[115] Shao, Y., Molnar, L.F., Jung, Y., Kussmann, J., Ochsenfeld, C., Brown, S.T., Gilbert, A.T.B., Slipchenko, L.V., Levchenko, S.V., O'Neill, D.P., Distasio, R.A., Lochan, R.C., Wang, T., Beran, G.J.O., Besley, N.A., Herbert, J.M., Lin, C.Y., Van Voorhis, T., Chien, S.H., Sodt, A., Steele, R.P., Rassolov, V.A., Maslen, P.E., Korambath, P.P., Adamson, R.D., Austin, B., Baker, J., Byrd, E.F.C., Dachsel, H., Doerksen, R.J., Dreuw, A., Dunietz, B.D., Dutoi, A.D., Furlani, T.R., Gwaltney, S.R., Heyden, A., Hirata, S., Hsu, C.P., Kedziora, G., Khalliulin, R.Z., Klunzinger, P., Lee, A.M., Lee, M.S., Liang, W.Z., Lotan, I., Nair, N., Peters, B., Proynov, E.I., Pieniazek, P.A., Rhee, Y.M., Ritchie, J., Rosta, E., Sherrill, C.D., Simmonett, A.C., Subotnik, J.E., Woodcock III, H.L., Zhang, W., Bell, A.T., Chakraborty, A.K., Chipman, D.M., Keil, F.J., Warshel, A., Hehre, W.J., Schaefer, H.F., Kong, J., Krylov, A.I., Gill, P.M.W., and Head-Gordon, M. "Advances in methods and algorithms in a modern quantum chemistry program package." *Phys. Chem. Chem. Phys.*, **8**, 3172–3191 (2006). doi: 10.1039/b517914a.

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

[117] Weigend, F., Häser, M., Patzelt, H., and Ahlrichs, R. "RI-MP2: optimized auxiliary basis sets and demonstration of efficiency." *Chem. Phys. Lett.*, **294**, 143–152 (1998).

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

[119] Distasio, R.A., Steele, R.P., and Head-Gordon, M. "The analytical gradient of dual-basis resolution-of-the-identity second-order Møller-Plesset perturbation theory." *Mol. Phys.*, **105**, 2731–2742 (2007).

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

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

[122] Steele, R.P., Distasio, R.A., and Head-Gordon, M. "Non-covalent interactions with dual-basis methods: Pairings for augmented basis sets." *J. Chem. Theory Comput.*, **5**, 1560–1572 (2009). doi:10.1021/ct900058p.

[123] J. W. Ponder, TINKER v6.2 (2014, `http://dasher.wustl.edu/tinker/`). Accessed May 15, 2014.

[124] Liu, D.C. and Nocedal, J. "On the limited memory BFGS method for large scale optimization." *Math. Program.*, **45**, 503–528 (1989). doi:10.1007/BF01589116.

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

[126] Ottersen, T. "On the Structure of the peptide linkage. The structures of formamide and acetamide at -165 degrees C and an ab initio study of formamide, acetamide, and n-methylformamide." *Acta Chem. Scand. A*, **29**, 939–944 (1975). doi:10.3891/acta.chem.scand.29a-0939.

[127] Nahringbauer, I. "Hydrogen Bond Studies. 39. Reinvestigation of the Crystal Structure of Acetic Acid (at +5 degrees C and -190 degrees C)." *Acta Chem. Scand.*, **24**, 453–462 (1970). doi:10.3891/acta.chem.scand.24-0453.

[128] Zobel, D., Luger, P., Dreissig, W., and Koritsanszky, T. "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**, 837–848 (1992). doi:10.1107/S0108768192005433.

[129] McMullan, R.K., Epstein, J., John, R., and Craven, B.M. "The crystal structure of imidazole at 103 K by neutron diffraction." *Acta Cryst. B.*, **35**, 688–691 (1979). doi:10.1107/S0567740879004441.

[130] Chisholm, J.A. and Motherwell, W.D.S. "COMPACK: A program for identifying crystal structure similarity using distances." *J. Appl. Crystall.*, **38**, 228–231 (2005). doi:10.1107/S0021889804027074.

[131] Wu, C.Y., Hancock, B., Mills, A., Bentham, A., Best, S., and Elliott, J. "Numerical and experimental investigation of capping mechanisms during pharmaceutical tablet compaction." *Powder Technology*, **181**, 121–129 (2008). doi: 10.1016/j.powtec.2006.12.017.

[132] Mazel, V., Busignies, V., Diarra, H., and Tchoreloff, P. "On the links between elastic constants and effective elastic behavior of pharmaceutical compacts: Importance of poisson's ratio and use of bulk modulus." *J. Pharm. Sci.*, **102**, 4009–4014 (2013). doi:10.1002/jps.23710.

[133] Egart, M., Ilic, I., Jankovic, B., Lah, N., and Srcic, S. "Compaction properties of crystalline pharmaceutical ingredients according to the Walker model and nanomechanical attributes." *Int. J. Pharm.*, **472**, 347–355 (2014). doi:10.1016/j.ijpharm.2014.06.047.

[134] Revard, B.C., Tipton, W.W., and Hennig, R.G. "Structure and stability prediction of compounds with evolutionary algorithms." *Top. Curr. Chem.*, **345**, 181–222 (2014).

[135] Glass, C.W., Oganov, A.R., and Hansen, N. "USPEX—Evolutionary crystal structure prediction." *Comp. Phys. Commun.*, **175**, 713–720 (2006). doi: 10.1016/j.cpc.2006.07.020.

[136] Zhu, Q., Oganov, A.R., Glass, C.W., and Stokes, H.T. "Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications." *Acta Cryst. B*, **68**, 215–226 (2012). doi:10.1107/S0108768112017466.

[137] Lonie, D.C. and Zurek, E. "XtalOpt: An open-source evolutionary algorithm for crystal structure prediction." *Comp. Phys. Commun.*, **182**, 372–387 (2011). doi: 10.1016/j.cpc.2010.07.048.

[138] Wang, Y. and Ma, Y. "Perspective: Crystal structure prediction at high pressures." *J. Chem. Phys.*, **140**, 040901 (2014). doi:10.1063/1.4861966.

[139] Raiteri, P., Martonák, R., and Parrinello, M. "Exploring polymorphism: the case of benzene." *Angew. Chem. Int. Ed.*, **44**, 3769–73 (2005). doi:10.1002/anie.200462760.

[140] Karamertzanis, P.G., Raiteri, P., Parrinello, M., Leslie, M., and Price, S.L. "The thermal stability of lattice-energy minima of 5-fluorouracil: metadynamics as an aid to polymorph prediction." *J. Phys. Chem. B*, **112**, 4298–308 (2008). doi: 10.1021/jp709764e.

[141] Kazantsev, A.V., Karamertzanis, P.G., Adjiman, C.S., and Pantelides, C.C. "Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Crystals." *J. Chem. Theory Comput.*, **7**, 1998–2016 (2011).

[142] Thompson, H.P.G. and Day, G.M. "Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape." *Chem. Sci.*, **5**, 3173–3182 (2014). doi:10.1039/c4sc01132e.

[143] Nolan, S.J., Bygrave, P.J., Allan, N.L., and Manby, F.R. "Comparison of the incremental and hierarchical methods for crystalline neon." *J. Phys.: Condens. Matter*, **22**, 074201 (2010). doi:10.1088/0953-8984/22/7/074201.

[144] Beran, G.J.O., Wen, S., Nanda, K., Huang, Y., and Heit, Y. "Accurate molecular crystal modeling with fragment-based electronic structure methods." *Top. Curr. Chem.*, **345**, 59–93 (2014). doi:10.1007/128_2013_502.

[145] Gilliard, K., Sode, O., and Hirata, S. "Second-order many-body perturbation and coupled-cluster singles and doubles study of ice VIII." *J. Chem. Phys.*, **140**, 174507 (2014). doi:10.1063/1.4873919.

[146] Sode, O., Keceli, M., Yagi, K., and Hirata, S. "Fermi resonance in solid CO_2 under pressure." *J. Chem. Phys.*, **138**, 074501 (2013). doi:10.1063/1.4790537.

[147] Price, S.L. "Computed crystal energy landscapes for understanding and predicting organic crystal structures and polymorphism." *Acc. Chem. Res.*, **42**, 117–126 (2009). doi:10.1021/ar800147t.

[148] Yu, T.Q. and Tuckerman, M. "Temperature-Accelerated Method for Exploring Polymorphism in Molecular Crystals Based on Free Energy." *Physical Review Letters*, **107**, 3–6 (2011). doi:10.1103/PhysRevLett.107.015701.

[149] Schnieders, M.J., Baltrusaitis, J., Shi, Y., Chattree, G., Zheng, L., Yang, W., and Ren, P. "The Structure, Thermodynamics and Solubility of Organic Crystals from Simulation with a Polarizable Force Field." *J. Chem. Theory Comput.*, **8**, 1721–1736 (2012). doi:10.1021/ct300035u.

[150] Park, J., Nessler, I., Mcclain, B., Macikenas, D., Baltrusaitis, J., and Schnieders, M.J. "Absolute Organic Crystal Thermodynamics: Growth of the Asymmetric Unit into a Crystal via Alchemy." *J. Chem. Theory Comput.*, **10**, 2781–2791 (2014). doi:10.1021/ct500180m.

[151] Duff, N. and Peters, B. "Polymorph specific RMSD local order parameters for molecular crystals and nuclei: α-, β-, and γ-glycine." *J. Chem. Phys.*, **135**, 134101 (2011). doi:10.1063/1.3638268.

[152] Anwar, J. and Zahn, D. "Uncovering molecular processes in crystal nucleation and growth by using molecular simulation." *Angew. Chem. Int. Ed.*, **50**, 1996–2013 (2011). doi:10.1002/anie.201000463.

[153] Salvalaglio, M., Vetter, T., Giberti, F., Mazzotti, M., and Parrinello, M. "Uncovering molecular details of urea crystal growth in the presence of additives." *J. Am. Chem. Soc.*, **134**, 17221–17233 (2012).

[154] Salvalaglio, M., Vetter, T., Mazzotti, M., and Parrinello, M. "Controlling and predicting crystal shapes: The case of urea." *Angew. Chem. Int. Ed.*, **52**, 13369–72 (2013). doi:10.1002/anie.201304562.

[155] Sun, J., Klug, D.D., Martonák, R., Montoya, J.A., Lee, M.S., Scandolo, S., and Tosatti, E. "High-pressure polymeric phases of carbon dioxide." *Proc. Nat. Acad. Sci.*, **106**, 6077–6081 (2009). doi:10.1073/pnas.0812624106.

[156] Nanda, K. D. PhD thesis, University of California, Riverside, 2013. `https://escholarship.org/uc/item/2qv949f5` Accessed August 10, 2015.

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

[158] Weigend, F. "A fully direct RI-HF algorithm: Implementation, optimised auxiliary basis sets, demonstration of accuracy and efficiency." *Phys. Chem. Chem. Phys.*, **4**, 4285–4291 (2002).

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

[160] Watts, J.D., Gauss, J., and Bartlett, R.J. "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**, 8718–8733 (1993).

[161] MOLPRO, version 2012.1, a package of ab initio programs, H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, see http://www.molpro.net.

[162] Werner, H.J., Knowles, P.J., Knizia, G., Manby, F.R., and Schütz, M. "Molpro: a general-purpose quantum chemistry program package." *WIREs Comput Mol Sci*, **2**, 242–253 (2012).

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

[164] Karton, A. and Martin, J.M.L. "Comment on: "Estimating the Hartree-Fock limit from finite basis set calculations"." *Theor. Chem. Acc.*, **115**, 330–333 (2006).

[165] Helgaker, T., Klopper, W., Koch, H., and Noga, J. "Basis-set convergence of correlated calculations on water." *J. Chem. Phys.*, **106**, 9639–9646 (1997).

[166] J. W. Ponder, TINKER v6.3 (2014, `http://dasher.wustl.edu/tinker/`). Accessed August 10, 2015.

[167] Wu, J.C., Chattree, G., and Ren, P. "Automation of AMOEBA polarizable force field parameterization for small molecules." *Theor. Chem. Acc.*, **131**, 1138 (2012). doi:10.1007/s00214-012-1138-6.

[168] Yu, K. and Schmidt, J.R. "Many-body effects are essential in a physically motivated CO_2 force field." *J. Chem. Phys.*, **136**, 034503 (2012). doi:10.1063/1.3672810.

[169] Cygan, R.T., Romanov, V.N., and Myshakin, E.M. "Molecular simulation of carbon dioxide capture by montmorillonite using an accurate and flexible force field." *J. Phys. Chem. C*, **116**, 13079–13091 (2012). doi:10.1021/jp3007574.

[170] Perez-Sanchez, G., Gonzalez-Salgado, D., Pineiro, M.M., and Vega, C. "Fluid-solid equilibrium of carbon dioxide as obtained from computer simulations of several popular potential models: The role of the quadrupole." *J. Chem. Phys.*, **138**, 084506 (2013). doi:10.1063/1.4792443.

[171] Graham, C., Imrie, D., and Raab, R. "Measurement of the electric quadrupole moments of CO_2, CO, N_2, Cl_2, and BF_3." *Mol. Phys.*, **93**, 49–56 (1998). doi:10.1080/00268979809482187.

[172] Potoff, J.J. and Siepman, J.I. "Vapor-Liquid Equilibria of Mixtures Containing Alkanes, Carbon Dioxide, and Nitrogen." *AICHE J*, **47**, 1676–1682 (2001).

[173] McDaniel, J.G. and Schmidt, J.R. "Physically-motivated force fields from symmetry-adapted perturbation theory." *J. Phys. Chem. A*, **117**, 2053–2066 (2013). doi:10.1021/jp3108182.

[174] Manzhelii, V.G., Tolkachev, A.M., Bagatskii, M.I., and Voitovich, E.I. "Thermal Expansion, Heat Capacity, and Compressibility of Solid CO_2." *Phys. Stat. Sol. (b)*, **44**, 39–49 (1971). doi:10.1002/pssb.2220440104.

[175] Keesom, W.H. and Köhler, J.W.L. "The Lattice Constant and Expansion Coefficient of Solid Carbon Dioxide." *Physica*, **1**, 655–658 (1934). doi:10.1016/S0031-8914(34)80253-4.

[176] Curzon, A. "A comment on the lattice parameter of solid carbon dioxide at -190 C." *Physica*, **59**, 733 (1972).

[177] Azreg-Aïnou, M. "Low-temperature data for carbon dioxide." *Monatshefte fur Chemie*, **136**, 2017–2027 (2005). doi:10.1007/s00706-005-0370-3.

[178] Giauque, W.F. and Egan, C. "Carbon Dioxide. The Heat Capacity and Vapor Pressure of the Solid. The Heat of Sublimation. Thermodynamic and Spectroscopic Values of the Entropy." *J. Chem. Phys*, **5**, 45 (1937). doi:10.1063/1.1749929.

[179] Herzberg, G. *Electronic Spectra and Electronic Structure of Polyatomic Molecules.* D. Van Nostrand Company, Inc (1966).

[180] Shimanouchi, T. *Tables of Molecular Vibrational Frequencies*, volume 1 of *NSRDS NBS-39*. United States Department of Commerce (1972).

[181] Birch, F. “Finite strain isotherm and velocities for single-crystal and polycrystalline NaCl at high pressures and 300°K.” *J. Geophys. Res.*, **83**, 1257–1268 (1978). doi: 10.1029/JB083iB03p01257.

[182] Vinet, P., Smith, J.R., Ferrante, J., and Rose, J.H. “Temperature effects on the universal equation of state.” *Phys. Rev. B*, **35**, 1945–1953 (1987).

[183] Liu, L. “Compression and phase behavior of solid CO2 to half a megabar.” *Earth Planet. Sci. Lett.*, **71**, 104–110 (1984). doi:10.1016/0012-821X(84)90056-6.

[184] Giordano, V.M., Datchi, F., Gorelli, F.A., and Bini, R. “Equation of state and anharmonicity of carbon dioxide phase I up to 12 GPa and 800 K.” *J. Chem. Phys.*, **133**, 0–9 (2010). doi:10.1063/1.3495951.

[185] Bridgman, P.W. “Rough Compressibilities of Fourteen Substances to 45,000 Kg / Cm.” *Proc. Am. Acad. Arts Sci.*, **72**, 207–225 (1938).

[186] Olinger, B. “The compression of solid CO2 at 296 K and 10 GPa.” *J. Chem. Phys.*, **77**, 6255–6258 (1982). doi:10.1063/1.443828.

[187] Yoo, C., Kohlmann, H., Cynn, H., Nicol, M., Iota, V., and LeBihan, T. “Crystal structure of pseudo-six-fold carbon dioxide phase II at high pressures and temperatures.” *Physical Review B*, **65**, 1–6 (2002). doi:10.1103/PhysRevB.65.104103.

[188] Martin Trusler, J.P. “Equation of State for Solid Phase I of Carbon Dioxide Valid for Temperatures up to 800 K and Pressures up to 12 GPa.” *J. Phys. Chem. Ref. Data*, **40**, 043105 (2011). doi:10.1063/1.3664915.

[189] Zhang, J.S., Shieh, S.R., Bass, J.D., Dera, P., and Prakapenka, V. “High-pressure single-crystal elasticity study of CO2 across phase I-III transition.” *Appl. Phys. Lett.*, **104**, 141901 (2014). doi:10.1063/1.4870526.

[190] Aoki, K., Yamawaki, H., Sakashita, M., Gotoh, Y., and Takemura, K. “Crystal Structure of the High-Pressure Phase of Solid CO_2.” *Science*, **263**, 356–358 (1994).

[191] Cruz-Cabeza, A.J., Reutzel-Edens, S.M., and Bernstein, J. “Facts and Fictions About Polymorphism.” *Chem. Soc. Rev.*, **44**, 8619–8635 (2015). doi:10.1039/C5CS00227C.

[192] G. Goldbeck, E. Pidcock, and C. Groom, “Solid Form Informatics for pharmaceuticals and agrochemicals: Knowledge-based substance development and risk assessment.” Cambridge Crystallographic Data Center, 2012. `http://www.ccdc.cam.ac.uk/Lists/ResourceFileList/Solid Form Informatics.pdf` Accessed June 28, 2013.

[193] Lommerse, J.P.M., Motherwell, W.D.S., Ammon, H.L., Dunitz, J.D., Gavezzotti, A., Hofmann, D.W.M., Leusen, F.J.J., Mooij, W.T.M., Price, S.L., Schweizer, B.,

Schmidt, M.U., van Eijck, B.P., Verwer, P., and Williams, D.E. “A test of crystal structure prediction of small organic molecules.” *Acta Cryst. B*, **56**, 697–714 (2000). doi:10.1107/S0108768100004584.

[194] Gavezzotti, A. and Filippini, G. “Polymorphic Forms of Organic Crystals at Room Conditions: Thermodynamic and Structural Implications.” *J. Am. Chem. Soc.*, **117**, 12299–12305 (1995). doi:10.1021/ja00154a032.

[195] van Eijck, B.P. “Ab Initio Crystal Structure Predictions for Flexible Hydrogen-Bonded Molecules. Part III. Effect of Lattice Vibrations.” *J. Comp. Chem.*, **22**, 816–826 (2001).

[196] Anghel, A.T., Day, G.M., and Price, S.L. “A study of the known and hypothetical crystal structures of pyridine: why are there four molecules in the asymmetric unit cell?” *CrystEngComm*, **4**, 348–355 (2002). doi:10.1039/b202084j.

[197] Rivera, S.A., Allis, D.G., and Hudson, B.S. “Importance of Vibrational Zero-Point Energy Contribution to the Relative Polymorph Energies of Hydrogen-Bonded Species.” *Cryst. Growth Des.*, **8**, 3905–3907 (2008).

[198] Cruz-Cabeza, A.J., Day, G.M., and Jones, W. “Towards Prediction of Stoichiometry in Crystalline Multicomponent Complexes.” *Chem. Eur. J.*, **14**, 8830–8836 (2008). doi:10.1002/chem.200800668.

[199] Price, S.L. “Why don’t we find more polymorphs?” *Acta Cryst. B*, **69**, 313–28 (2013). doi:10.1107/S2052519213018861.

[200] Zykova-Timan, T., Raiteri, P., and Parrinello, M. “Investigating the polymorphism in PR179: a combined crystal structure prediction and metadynamics study.” *J. Phys. Chem. B*, **112**, 13231–13237 (2008). doi:10.1021/jp802977t.

[201] Coropceanu, V., Cornil, J., da Silva Filho, D.A., Olivier, Y., Silbey, R., and Brédas, J.L. “Charge Transport in Organic Semiconductors.” *Chem. Rev.*, **107**, 926–952 (2007). doi:10.1021/cr050140x.

[202] Li, Y., Coropceanu, V., and Bredas, J.L. “Thermal Narrowing of the Electronic Bandwidths in Organic Molecular Semiconductors: Impact of the Crystal Thermal Expansion.” *J. Phys. Chem. Lett.*, **3**, 3325–3329 (2012). doi:10.1021/jz301575u.

[203] Filippini, G., Gramaccioli, C.M., Simonetta, M., and Suffritti, G.B. “Lattice-dynamical evaluation of cell parameters of some hydrocarbons at various temperatures.” *Chem. Phys. Lett.*, **35**, 17–20 (1975). doi:10.1016/0009-2614(75)85579-5.

[204] Filippini, G. and Gramaccioli, C.M. “Deriving the equilibrium conformation in molecular crystals by the quasi-harmonic procedure: Some critical remarks.” *Acta Cryst. A*, **37**, 335–342 (1981).

[205] Taylor, M.B., Barrera, G.D., Allan, N.L., and Barron, T.H.K. “Free-energy derivatives and structure optimization within quasiharmonic lattice dynamics.” *Physical Review B*, **56**, 14380–14390 (1997). doi:10.1103/PhysRevB.56.14380.

[206] Pauling, L. "The Structure and Entropy of Ice and of Other Crystals with Some Randomness of Atomic Arrangement." *J. Am. Chem. Soc.*, **57**, 2680–2684 (1935). doi:10.1021/ja01315a102.

[207] Morrison, I., Li, J.C., Jenkin, S., Xantheas, S.S., and Payne, M.C. "Ab-Initio Total Energy Studies of the Static and Dynamical Properties of Ice Ih." *J. Phys. Chem. B,*, **101**, 6146–6150 (1997).

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

[209] A. J. Misquitta and A. J. Stone, CamCASP v5.6 (2011), `http://www-stone.ch.cam.ac.uk/programs.html`. Accessed Februrary 23, 2011.

[210] Acree, W. and Chickos, J.S. "Phase Transition Enthalpy Measurements of Organic and Organometallic Compounds. Sublimation, Vaporization and Fusion Enthalpies From 1880 to 2010." *J. Phys. Chem. Ref. Data*, **39**, 043101 (2010). doi: 10.1063/1.3309507.

[211] Boese, R., D, B., Latz, R., and Bäumen, A. "Acetic acid at 40K." *Acta Cryst. C*, **C55**, 9900001 (1999). doi:10.1107/S0108270199099862.

[212] Jönson, P. "Hydrogen Bond Studies. XLIV. Neutron Diffraction Study of Acetic Acid." *Acta Cryst.*, **B27**, 893–898 (1971).

[213] Craven, B.M., McMullan, R.K., Bell, J.D., and Freeman, H.C. "The crystal structure of imidazole by neutron diffraction at 20C and -150C." *Acta Cryst. B.*, **B33**, 2585 (1977). doi:10.1107/S0567740877008954.

[214] Deschamps, J. R., Cook ,J.M., and Teng Y., *Private Communication*, (2008) doi: 10.5517/ccr4jbq.

[215] Epstein, J., Ruble, J.R., and Craven, B.M. "The Change Density in Imidazole by X-ray Diffraction at 103 and 293 K." *Acta Cryst. B.*, **38**, 140–149 (1982). doi: 10.1107/S0567740882002246.

[216] Will, G. *Z.Kristallogr.,Kristallgeom.,Kristallphys.,Kristallchem.*, **129**, 211 (1969).

[217] Babin, V., Medders, G.R., and Paesani, F. "Development of a First Principles Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters." *J. Chem. Theory Comput.*, **10**, 1599–1607 (2014). doi:10.1021/ct500079y.

[218] Medders, G.R., Götz, A.W., Morales, M.A., Bajaj, P., and Paesani, F. "On the representation of many-body interactions in water." *J. Chem. Phys.*, **143**, 104102 (2015). doi:10.1063/1.4930194.

[219] Sinnokrot, M.O. and Sherrill, C.D. "High-Accuracy Quantum Mechanical Studies of π-π Interactions in Benzene Dimers." *J. Phys. Chem. A*, **110**, 10656–10668 (2006).

[220] Kronik, L. and Tkatchenko, A. "Understanding Molecular Crystals with Dispersion-Inclusive Density Functional Theory: Pairwise Corrections and Beyond." *Acc. Chem. Res.*, **47**, 3208–3216 (2014). doi:10.1021/ar500144s.

[221] Berland, K., Cooper, V.R., Lee, K., Schröder, E., Thonhauser, T., Hyldgaard, P., and Lundqvist, B.I. "van der Waals forces in density functional theory: a review of the vdW-DF method." *Rep. Prog. Phys.*, **78**, 066501 (2015). doi:10.1088/0034-4885/78/6/066501.

[222] Beyer, T. and Price, S.L. "The errors in lattice energy minimisation studies: sensitivity to experimental variations in the molecular structure of paracetamol." *CrystEngComm*, **2**, 183 (2000). doi:10.1039/b006604o.

[223] Gavezzotti, A. "A Molecular Dynamics Test of the Different Stability of Crystal Polymorphs under Thermal Strain." *J. Am Chem. Soc.*, **122**, 10724–10725 (2000). doi:10.1021/ja000588.

[224] Beran, G. J. O. "Modeling Polymorphic Molecular Crystals with Electronic Structure Theory", *Chem. Rev.* submitted (2016).

[225] Litasov, K.D., Goncharov, A.F., and Hemley, R.J. "Crossover from melting to dissociation of CO_2 under pressure: Implications for the lower mantle." *Earth Planet. Sci. Lett.*, **309**, 318 (2011). doi:10.1016/j.epsl.2011.07.006.

[226] Shieh, S.R., Jarriage, I., Wu, M., Hiraoka, N., Tse, J.S., Mi, Z., Kaci, L., Jiang, J., and Cai, Y.Q. "Electronic structure of carbon dioxide under pressure and insights into the molecular-to-nonmolecular transition." *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 1840218406 (2013). doi:10.1073/pnas.1305116110.

[227] Gygi, F. "First-principles simulations of organic compounds: Solid CO_2 under pressure." *Comput. Mater. Sci.*, **10**, 63–66 (1998).

[228] Keesom, W.H. and Köhler, J.W.L. "New Determination of Lattice constant of Carbon Dioxide." *Physica*, **1**, 167–174 (1934). doi:10.1016/S0031-8914(34)90021-5.

[229] Liu, L. "Dry ice II, a new polymorph of CO_2." *Nature*, **303**, 508–509 (1983).

[230] Datchi, F., Mallick, B., Salamat, A., Rousse, G., Ninet, S., Garbarino, G., Bouvier, P., and Mezouar, M. "Structure and compressibility of the high-pressure molecular phase II of carbon dioxide." *Phys. Rev. B*, page 144101 (2014).

[231] Hanson, R.C. "A New High-pressure Phase of Solid CO_2." *J. Phys. Chem*, **89**, 4499–4501 (1985).

[232] Giorano, V.M. and Datchi, F. "Molecular carbon dioxide at high pressure and high temperature." *Europhys. Lett.*, **77**, 46002 (2007). doi:1209/0295-5075/77/46002.

[233] Datchi, F., Giorano, V.M., Munsch, P., and Saitta, A.M. "Structure of Carbon Dioxide Phase IV: Breakdown of the Intermediate Bonding State Scenario." *Phys. Rev. Lett.*, **103**, 185701 (2009).

[234] Yoo, C., Iota, V., and Cynn, H. "Nonlinear Carbon Dioxide at High Pressures and Temperatures." *Phys. Rev. Lett.*, **86**, 444–447 (2001). doi: 10.1103/PhysRevLett.86.444.

[235] Park, J.H., Yoo, C.S., Iota, V., Cynn, H., Nicol, M.F., and Bihan, T.L. "Crystal structure of bent carbon dioxide phase IV." *Phys. Rev. B*, **68**, 014107 (2003).

[236] Feistel, R. and Wagner, W. "Sublimation pressure and sublimation enthalpy of H_2O ice Ih between 0 and 273.16 K." *Geochim. Cosmochim. Acta*, **71**, 36–45 (2007). doi: 10.1016/j.gca.2006.08.034.

[237] NIST/TRC Web Thermo Tables (WTT), NIST Standard Reference Subscription Database 2—Lite Edition Version 2-2012-1-Lite.

[238] Haynes, W.M., editor. *Handbook of Chemistry and Physics.* CRC Press, 96 edition (2004).

[239] Feistel, R. and Wagner, W. "A New Equation of State for H_2O Ice Ih." *J. Phys. Chem. Ref. Data*, **35**, 1021–1047 (2006).

[240] Verevkin, S. "Measurement and Prediction of the Monocarboxylic Acids Thermochemical Properties." *J. Chem. Eng. Data*, **45**, 953–960 (2000).

[241] Martin, J. and Andon, R.L. "Thermodynamic properties of organic oxygen compounds Part LII. Molar heat capacity of ethanoic, propanoic and butanoic acids." *J. Chem. Thermo.*, **14**, 679–688 (1982). doi:10.1016/0021-9614(82)90083-0.

[242] Jiménez, P. and M.V. Roux, C.T. "Enthalpies of combustion,vapour pressures and enthalpies of subimation, and enthalpies of formation of pyrazole, imidazole, indazole, and benzimidazole." *J. Chem. Thermo.*, **19**, 985–992 (1987).

[243] De Wit, H.G.M., De Kruif, C.G., and Van Miltenburg, J.C. "Thermodynamic properties of molecular organic crystals containing nitrogen, oxygen, and sulfur II. Molar heat capacities of eight compounds by adiabatic calorimetry." *J. Chem. Thermo.*, **15**, 891–902 (1983).

[244] Billes, F., Endrédi, H., and Jalsovszky, G. "Vibrational spectroscopy of diazoles." *J. Mol. Struct. (Theochem)*, **465**, 157–172 (1999).