UNIVERSITY OF CALIFORNIA
RIVERSIDE

# The Importance of Thermal Expansion in the Simulation of Molecular Crystals

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

Doctor of Philosophy

in

Chemistry

by

Jessica Lynee McKinley

June 2019

Dissertation Committee:

Dr. Gregory J. O. Beran, Chairperson
Dr. De-en Jiang
Dr. Leonard Mueller

The Dissertation of Jessica Lynee McKinley is approved:

---

---

---

Committee Chairperson

University of California, Riverside

## Acknowledgments

I am grateful to my advisor Dr. Gregory Beran for his guidance and patience. Greg created a stimulating research environment where I was constantly challenged and allowed to continually exceed my limits. I am also grateful to my lab mates Dr. Dominique Nocito, Dr. Ctirad Červinka, Dr. Yuanhang Huang, Dr. Yonaton Heit, Dr. Joshua Hartman, Watit Sontising, Chandler Greenwell, Pablo Unzueta, and Cameron Cook for all the helpful discussions. I would like to thank my family for all their support throughout my career. I am grateful to Nichole Johnston for the many talks we shared. Finally I would like to thank my finacé Dominique Nocito for his love, unwavering support, and encouragement.

To my parents Carolyn and Frank McKinley.

# ABSTRACT OF THE DISSERTATION

The Importance of Thermal Expansion in the Simulation of Molecular Crystals

by

Jessica Lynee McKinley

Doctor of Philosophy, Graduate Program in Chemistry
University of California, Riverside, June 2019
Dr. Gregory J. O. Beran, Chairperson

Molecular crystals are commonly used in pharmaceuticals, organic semi-conductor materials, explosives, and many other areas of Chemistry. Molecular crystal packing interactions are governed by subtle balances between intra- and intermolecular interactions, providing a severe challenge for theoretical crystal structure modeling. Additionally, molecular crystals can expand appreciably upon heating due to both zero-point and thermal vibrational motion, yet this expansion is often neglected in molecular crystal modeling studies. Quasi-harmonic (QHA) approaches provide an economical route to modeling the temperature dependence of molecular crystal structures and properties but can be cost-prohibitive when evaluated at higher levels of theory.

In this thesis, we introduce a hierarchy of models (tiered-QHA) in which the energies, geometries, and phonons are computed either with correlated methods (such as second-order Møller-Plesset perturbation theory (MP2)) or density functional theory (DFT). We examine which combinations produce useful predictions for properties such as the molar volume, enthalpy, and entropy as a function of temperature. Compared to performing the

entire calculation using pure MP2, this leads to a modest loss in chemical accuracy and provides a necessary increase in the speed of computation. Additionally, employing this method increases the system size we can feasibly simulate to over 100 atoms per central unit cell, allowing us to simulate pharmaceutically-relevant molecular crystals. We apply this new tiered-QHA method to examine the phase-transition properties of $\alpha-$ and $\beta-$ resorcinol.

Finally, neglecting thermal expansion can significantly affect simulated spectroscopic properties. In particular, nuclear magnetic resonance (NMR) chemical shift predictions will suffer since a small change in atomic position translates to a large change in the chemical shift spectra. We investigate how accounting for thermal expansion in molecular crystals via the QHA refines isotropic 68 $^{13}$C and 28 $^{15}$N predicted chemical shifts on a number of molecular crystals. We demonstrate that chemical shifts computed using quasi-harmonic room-temperature structures rival those based on the experimental unit cell parameters. We also show increased discrimination between candidate structures amongst five theophylline structures that were generated via crystal structure prediction.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Molecular crystal polymorphism, or the ability for molecules to adopt many different packing motifs in the solid state, is a challenging topic in chemistry today. Molecular crystal polymorphs are quite common due to the varity of intermolecular interactions that can occur, including hydrogen bonding, dispersion, electrostatics, induction, and occasionally $\pi - \pi$ stacking interactions. Due to the delicate balance of interactions that must be maintained, more complex molecules in particular tend to exhibit polymorphism. One particularly difficult system to study is 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophenecarbonitrile, affectionately known as ROY due to the red, orange, and yellow hues the polymorphs exhibit.[13–16] It has been found that these polymorphs can have both different intermolecular packing configurations and different intramolecular configurations (via a rotation of two intra-molecular torsion angles).[17–20] These small changes in crystal packing can lead to changes in physical properties, such as the aforementioned color changes.

However knowledge of crystal structure polymorphism has been surprisingly slow

to propagate through the general public. In as early as 1965 W. C. McCrone can be quoted saying "It is at least this author's opinion that every compound has different polymorphic forms, and that, in general, the number of forms known for that compound is proportional to the time and money spent in research on that compound." [21] Recently, the question has begun to change from which molecular crystals have polymorphs to which do not. [22] This problem of unknown polymorphs and disappearing polymorphs [23,24] can have significant effects especially in patent law. [25]

It is important for pharmaceutical companies to know which polymorphs are stable under various conditions. Polymorphs have been found to undergo phase transitions at elevated temperatures. [26–37] In the production phase of pharmaceutical drugs the crystals often experience high pressures when being ground into tablets. This process is also well known to induce a phase transition, something that can often go undetected. [38–41] For example, one well-known case is the anti-HIV drug Ritonavir which was previously thought to only have one stable polymorph. [27,42] However, upon mass-production a second insoluble polymorph was accidentally produced which forced the drug's removal from market. This cost the company Abbot Laboratories an estimated $250 million and left patients without their medication. [43]

The work presented in this thesis focuses on computational efforts to distinguish between molecular crystal polymorphs and, at times, predict their stability under thermodynamic conditions. The remainder of this chapter will establish what is currently considered to be state of the art for modeling molecular crystals. The next two chapters will explore the uses of crystal structure prediction and how simulated nuclear magnetic

resonance (NMR) chemical shifts can be utilized to enhance discrimination between crystal candidates. Finally, the last two chapters will focus on what level of theory is needed to accurately simulate molecular crystals under thermodynamic conditions.

## 1.1 Crystal Structure Prediction

While there is a clear desire to map out all the stable polymorphs for every molecular crystal, this has proved challenging to do experimentally. Alternatively one could instead pursue computational crystal structure prediction.[44–51] Crystal structure prediction can provide a powerful and increasingly reliable candidate structures, as evidenced by recent blind tests of crystal structure prediction.[52–54] Especially in areas where it can be difficult to perform experimental techniques, computational techniques are being sought as a viable solution to this problem.

Ideally crystal structure prediction algorithms should require only knowledge of the stick diagram of the molecule and the desired level of theory the energies should be evaluated at. From this information the goal would be to generate all stable polymorphs and correctly predict their relative stabilities and thermodynamical observables.[22] While each crystal structure prediction approach has subtleties in their exact differences a general procedure can be derived: 1. Generate all viable crystal candidates and 2. Rank the stability of these candidates correctly.

### 1.1.1 Generating Candidate Structures

While the concept of crystal structure prediction is simple, thoroughly searching the configurational space remains challenging in practice. The search space of the crystal structure grows combinatorially as a function of the number of molecules in the cell and the internal degrees of freedom of the molecule itself. Molecules with many degrees of freedom remain some of the most challenging species to model to this day.[17–20] Due to the large search space that needs to be covered, exhaustive searching approaches are ineffective for all but the simplest of crystals.

Crystal structure generation has largely become somewhat of an art form. Some algorithms take a simple approach to generating crystals by enforcing a basic "crystal logic" on the structure generation. For example the *ab-initio* random structure search (AIRSS)[55] generates crystal structures primarily from space group symmetry operations. Other more sophisticated software attempt to assign a penalty for adopting less optimal configurations, however this requires some knowledge of the energy landscape.[48,56] Genetic and evolutionary algorithms build up the search space by "breeding" successive generations of crystal candidates.[57,58] This is done by defining a fitness criteria (typically lowest-energy) and the crystals of the current generation that best satisfy that criterion are chosen to undergo a series of mutations and distortions to generate a new daughter species. There have also been a number of promising ventures into incorporating machine leaning and other data-driven learning models into the structure search routines.[59,60] This problem of search coverage remains a major source of active research in the field today.

Despite the challenges that exist most structure searching algorithms are capable of providing a well-sampled search space after generating thousands to millions of structures. This itself creates another major problem as not every candidate structure is useful. It is quite often the case that candidate structures will differ by slight rotations of internal degrees of freedom in the molecule or of the relative packing motifs. While these changes are enough to constitute a new polymorph, the differences in lattice energies between these structures are often quite small, making it difficult to predict which of these polymorphs will be observed experimentally.

### 1.1.2 Polymorphic energy rankings

Rapid and efficient sampling of configurational space often necessitates the use of computationally inexpensive methods such as semi-empirical[61] or force-field methods. However, neither force fields nor semi-empirical methods provide the requisite accuracy required for resolving the energetics of crystal polymorphs. In order to have confidence in the discrimination of candidate structures, the model must be capable of resolving errors less that 1 kcal/mol relative to experiment (referred to as chemical accuracy). A common strategy to address this deficiency is to perform an initial pass with a "cheap" level of theory and later refine the candidate structures using a more accurate level of theory.

There is a question as to how many crystal structures should be refined. Structure searches commonly generate hundreds to thousands of candidate structures even after eliminating redundant candidates. A previous study done by Nyman and Day had found that half of experimentally observed molecular crystal polymorphs differ by less than 2 kJ/mol and the majority of all crystal polymorphs lie within 10 kJ/mol of the most stable

form.[62,63] Occasionally polymorphs can be found whose lattice energy differences exceed this window but those cases are rare.[62,64] Hence it is currently common practice to limit the energy refinement pass to only those structures which lie within 10 kJ/mol of the current lowest-energy candidate.[22]

This pruning reduces the number of structures that need to be refined to the range of tens or hundreds of molecular crystals. Because it provides generally accurate energetics at reasonable computational cost periodic Kohn-Sham density functional theory (DFT)[65,66] is the most commonly employed method for refinement. Oftentimes this will be supplemented with another method for describing the van-der Waals dispersion interaction such as Grimmes's dispersion correction (D,[67] D2,[68] D3,[69] and D4[70]), the exchange-hole dipole moment model (XDM[4]), Tkatchenko-Scheffler (TS),[71,72] or many-body dispersion (MBD).[73–75] DFT has previously been used to predict the lattice energies of the X23 benchmark set[76–78] with a mean absolute deviation (MAD) of 3.8 kJ/mol. It has also been shown that DFT is capable of describing relative polymorphic stabilities of oxalic acid,[78] glycine,[79] and coumarin[80] with an accuracy of about 1 kJ/mol with respect to experiment.

However, DFT also suffers from its own limitations. Although there are clear cases where DFT has proven useful, it has been unable to consistently achieve chemical accuracy.[81–83] Furthermore the performance is functional-dependent and while there are a hierarchy of methods that can be used to improve the calculation ("Jacob's ladder")[84–87] this can quickly become computationally intractable especially in a planewave basis. Recent studies have shown even the more expensive hybrid density functionals have failed to provide enhanced discrimination of candidate structures due to underestimation of dispersion

interactions between conformational polymorphs (despite employing the aforementioned corrections),[83] and excessive electron delocalization in candidate structures of ROY.[14,16] Problems with delocalization error have also caused spontaneous proton transfer in salt co-crystals.[88]

Ideally one would use large-basis dispersion-corrected second-order Møller-Plesset perturbation theory (MP2C or MP2D) and/or coupled clusters singles and doubles with perturbative triples (CCSD(T)) to rank the relative stability as these methods have been found to give accurate performance.[89–93] However, the computational cost associated with these methods makes it feasible to apply this to only a few tens of atoms which greatly limits the systems that can be studied.

## 1.2 The Hybrid Many-Body Interaction Model

One way to circumvent the high computational cost associated with evaluating the energetics of a periodic crystal with highly-accurate electronic structure methods is to use fragment-based methods. Fragment-based methods decompose the problem into a series of interacting subsystems. When modeling molecular crystals, a natural definition of these subsystems is to assign them to increasing orders of molecular interactions. The hybrid many-body interaction (HMBI)[94–97] model decomposes the total energy of the system into a many-body expansion

$$E_{total} = \sum_{i} E_i + \sum_{ij} \Delta^2 E_{ij} + \sum_{ijk} \Delta^3 E_{ijk} + ... \tag{1.1}$$

where $E_i$ represents the energy of the isolated molecule (monomer), $\Delta^2 E_{ij}$ is the pairwise

interaction energy between molecules i and j (dimer), and $\Delta^3 E_{ijk}$ is the three-body interaction energy between molecules i, j, and k (trimer). Due to the high computational cost associated with calculating trimer interactions, the three-body and higher interactions are approximated with an inexpensive polarizable force field.

HMBI treats individual molecules in the unit cell and their short-range pairwise interactions quantum mechanically (QM), while the longer-range and many-body interactions are typically approximated using a classical molecular mechanics (MM) polarizable force field.

$$U_{el}^{HMBI} = E_{1-body}^{QM} + E_{SR\ 2-body}^{QM} + E_{LR\ 2-body}^{MM} + E_{many\ body}^{MM} \tag{1.2}$$

In some cases, the MM terms will be computed from periodic Hartree-Fock (HF) instead, in which case HMBI is equivalent to the method of increments.[98–100]

There are a number of advantages to this method. For one, the calculations now scale linearly with respect to the number of molecules in the periodic cell rather than $N^5$−$N^7$ with respect to the number of atoms in the crystal. This also encodes a natural way of systematically improving the quality of prediction, either through the use of highly-accurate electronic structure methods or by explicitly calculating higher-order interactions. Since these calculations are broken up, they are highly parallelizable. This allows the use of more expensive calculations that were previously infeasible including MP2 or CCSD(T) and large basis sets.

A number of studies have demonstrated the power of HMBI. It has previously been shown to predict the lattice energy of small molecule clusters within 1-2 kJ/mol of benchmark calculations and crystal lattice energies within 2-4 kJ/mol of experiment.[100–102]

Spectral properties can also be accurately reproduced allowing for the prediction of nuclear magnetic resonance chemical shifts and Raman spectra.[12,26,103,104] HMBI also allows for accurate prediction of thermodynamic properties including stability ordering of small-molecules, accurate reproduction of equation of state, bulk-moduli, and entropies and enthalpies of sublimation.[26,105–107] Even more impressively it has been shown that complicated phase behavior such as the phase boundaries of crystalline methanol can be reliably modeled with this method.[108,109]

## 1.3 The Harmonic Approximation

As previously stated accurate discrimination of crystal candidates requires the use of electronic structure modeling. However, the majority of rankings are done using the lattice energy of a crystal structure which can lead to an overly large configurational space. Free energy calculations are often required to provide accurate thermodynamic modeling at ambient conditions.

From statistical thermodynamics,[110] the Gibbs free energy combines the electronic internal energy $U_{el}$, the Helmholtz vibrational free energy $F_{vib}$, and a pressure-volume ($PV$) contribution,

$$G(T,P) = U_{el} + F_{vib}(T) + PV \tag{1.3}$$

In crystals at ambient pressure, the $PV$ term contributes negligibly and hence is often neglected which reduces the Gibbs free energy to the Helmholtz free energy:

$$F(T) = U_{el} + F_{vib}(T) = U_{el} + H_{vib} - TS_{vib} \tag{1.4}$$

The Helmholtz vibrational free energy is computed from standard harmonic oscillator vibrational partition functions where the vibrational enthalpy ($H_{vib}$) is defined as

$$H_{vib}(T) = N_a \sum_i \left(\frac{\hbar\omega_i}{2} + \frac{\hbar\omega_i}{\exp(\frac{\hbar\omega_i}{k_b T}) - 1}\right) \tag{1.5}$$

and the vibrational entropy ($S_{vib}$) as

$$S_{vib}(T) = N_a \sum_i \left(\frac{\hbar\omega_i}{T(\exp(\frac{\hbar\omega_i}{k_b T}) - 1)} - k_b \ln\left[1 - \exp\left(-\frac{\hbar\omega_i}{k_b T}\right)\right]\right) \tag{1.6}$$

where $N_a$ is Avogadro's number, $\hbar$ is Plank's constant, $k_b$ is the Boltzmann constant, and $\omega_i$ is the vibrational frequency of mode $i$. Summing these vibrational contributions together yields the Helmoltz vibrational energy.

$$F_{vib}(T) = N_a \sum_i \left(\frac{\hbar\omega_i}{2} + k_b T \ln\left[1 - \exp\left(-\frac{\hbar\omega_i}{k_b T}\right)\right]\right) \tag{1.7}$$

The first term corresponds to the zero-point vibrational contribution, while the second gives the thermal vibrational contribution.

The harmonic approximation is a fixed-volume correction which does not account for thermal-vibrationally driven expansion of the cell. However the inclusion of harmonic phonons can be important for accurate modelling of relative stabilities.[62,63] Harmonic vibrational contributions impact the polymorph stability ordering in glycine,[111,112] oxalyl dihydrazide,[102] aspirin,[113] and modern drug targets,[114] for example.

However, the inclusion of these vibrational effects cannot overcome the choice of less accurate energy landscapes. Take for example a study done on two different molecular crystal polymorphs, acetamidobenzamide and oxalyl dihydrazide (See Figure 1.1). When these systems are modeled using periodic/planewave density functional theory the harmonic

approximation only narrows the energy gap between the polymorphs. Upon the inclusion of correlated wavefunctions (MP2) and correcting the inherent dispersion correction with one that can correctly model both intra- and inter-molecular dispersion (the D-correction[115]) the stability ordering of both crystals re-rank.

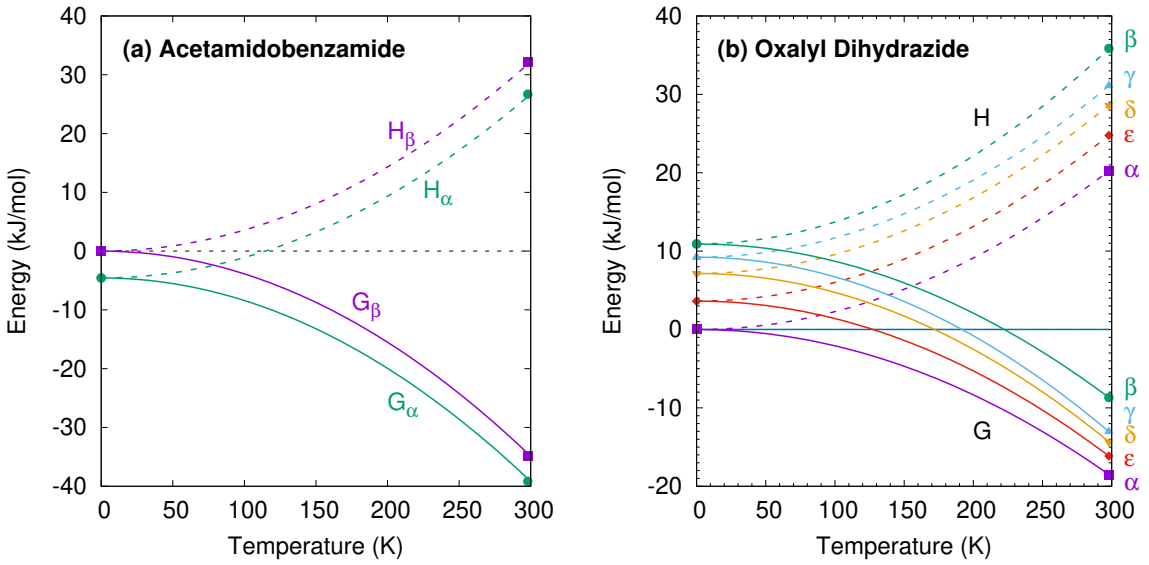



Figure 1.1: Relative stability ordering of a) acetamidobenzamide and b) oxalyl dihydrazide. Energies were calculated using Quantum Espresso[1] with the B86bPBE functional[2,3] and the XDM dispersion correction.[4]

## 1.4 The Quasi-Harmonic Approximation

While the harmonic approximation does capture a significant fraction of the vibrational contribution, it neglects the thermal expansion of the crystal. Molecular crystals typically expand its molar volume by ∼3–4% between the electronic minimum energy structure and the experimental room temperature structure, systems dominated by weaker intermolecular interactions such as van der Waals or $\pi - \pi$ stacking can exhibit volume

expansion of nearly ∼9%. Surveys by Nyman and Day found that accounting for these free energy effects can re-rank the relative energetic stabilities of molecular crystal polymorphs in up to 21% of cases.[62,63]

Molecular dynamics techniques provide a natural means of treating these effects.[116–125] However, obtaining force fields capable of describing the subtle energetic balances that occur in molecular crystal polymorphs is not always possible. The quasi-harmonic approximation (QHA) provides a computationally practical strategy for improving on the purely harmonic model and capturing this thermal expansion and its impact on crystal properties. The QHA models how the Gibbs free energy depends on temperature and pressure by incorporating a dependence of the harmonic phonons with volume. A number of studies have shown use of the QHA with DFT and other correlated-wavefunction approaches allows the prediction of small-molecule structures and their associated properties in excellent agreement with experiment.[61,105–107,126–130]

There are a few ways to build in this volume-dependence of the phonons, but all methods start by optimizing the crystal structure. The crystal structure which minimizes the electronic energy $U_{el}$ (i.e. those obtained from relaxing both atomic positions and unit cell parameters) is referred to as the reference structure. Harmonic phonons for the reference crystal ($\omega_i^{ref}$) are typically calculated to ensure the optimized structure lies at a minimum on the potential energy surface.

How the remainder of the phonons are evaluated is where the different flavors of the quasi-harmonic approximation come into play. In principle, one ought to re-evaluate the phonons at every volume along the potential energy surface, but that is computationally

impractical given the cost of typical electronic structure theory calculations. Instead one could employ mode-specific Grüneisen parameters $\gamma_i$ to estimate how individual phonon modes vary with unit cell volume,

$$\gamma_i = -\left(\frac{\partial \ln \omega_i}{\partial \ln V}\right) \tag{1.8}$$

Integrating Eq 1.8 yields,

$$\omega_i = \omega_i^{ref}\left(\frac{V}{V^{ref}}\right)^{-\gamma_i} \tag{1.9}$$

which allows the reference phonons $\omega_i^{ref}$ to be scaled to any given new volume $V$.

Heit et al approximates that the differential in eq 1.8 can be accurately represented as a finite difference of the phonons at two different structures. To compute these Grüneisen parameters, the reference structure is isotropically expanded and contracted by a known change in volume typically chosen to be about $\pm 10$ Å$^3$ away from the reference volume.[105,106] Harmonic phonons are then computed at each of these volumes and the mode-specific Grüneisen parameter is then obtained via finite difference.

While this method is significantly cheaper than evaluating the mass-weighted hessian at every volume, there are two major drawbacks. First, an accurate mode-matching scheme is required otherwise the reference phonons could be incorrectly scaled across the modes. Secondly, the molecular crystal must undergo a full optimization at every desired temperature and pressure along the free energy surface. Especially when using high-level electronic structure theory calculations (MP2 or higher) this can quickly become computationally intractable.

One could attempt to estimate the change in the electronic energy surface ahead of time by mapping out electronic energy versus volume curves $E(V)$. While it is relatively

straightforward to do this with isotropic expansion and contraction of the lattice volume, crystals containing larger molecules or more layered systems typically expand in a more anisotropic manner. Instead, the E(V) curves can be mapped out by minimizing the electronic energy of the reference cell under positive (cell compression) or negative pressure (cell expansion). This approach allows the cell to deform anisotropically along the lowest-energy path, which can be important for reliable modeling of the free energies.[108,109,131]

Rather than addressing the mode-matching problem one might instead side-step the problem altogether by evaluating additional harmonic phonons at a number of volumes spaced around the reference structure. The Helmholtz vibrational energy ($F_{vib}$) at a given temperature can then be fitted with either a linear[108,109] or a polynomial fit.[132] Summing the Helmholtz vibrational energy curves together with the electronic energy curve again produces the free energy. Since the particular volume which minimizes the free energy for a given temperature is unlikely to correspond to one of the sampled volumes, each free energy curve is fitted to a weighted double-Murnaghan equation of state. The Murnaghan equation of state is given by,

$$F(V) = F_0 + \frac{B_0 V}{B_0'}\left[\frac{(V_0/V)^{B_0'}}{B_0' - 1} + 1\right] - \frac{B_0 V_0}{B_0' - 1} \tag{1.10}$$

where $F_0$, $V_0$, $B_0$, and $B_0'$ are the fit parameters. $F_0$ gives the free energy at the minimum, $V_0$ is the molar volume at the minimum energy, $B_0$ is the bulk modulus, and $B_0'$ is the first derivative of the bulk modulus with respect to pressure. This method identifies the free-energy minimum effectively while avoiding artifacts that can be caused by simpler equation of state fits or splines.[108,109]

While this alternative approach allows one to readily back out the volume which minimizes the free energy surface at any temperature and pressure, it does so at significant computational cost. It requires five to seven phonon evaluations, compared to only three for the Grüneisen approach. As the evaluation of harmonic phonons are a rate-limiting step in all modeling, this is particularly undesirable. Additionally, ten to fifteen optimizations are required to produce a smooth electronic energy surface versus the Grüneisen approach required three optimizations plus additional optimizations for each desired pressure and temperature. Chapter 4 introduces a new tiered quasi-harmonic approach that attempts to address some of these deficiencies that are present in both these methods.

## 1.5 Solid State Nuclear Magnetic Resonance

Experimentally a number of techniques exist for molecular crystal structure elucidation. Single crystal neutron diffraction methods remains the gold standard. However, it can be difficult to obtain large single crystals which are free of impurities. Diffraction methods characterize the long-range order effectively, but they can sometimes have difficulty resolving local features such as protonation states. Alternative strategies are needed for cases where suitable single crystals are not obtainable.

Nuclear magnetic resonance (NMR) crystallography represents one such alternative approach. It combines solid-state NMR, powder x-ray diffraction, and *ab initio* chemical shielding predictions to solve crystal structures.[133–137] Solid-state NMR complements powder x-ray diffraction by providing detailed information about the local chemical environments. However, mapping from the chemical shifts in an NMR spectrum to a three-

dimensional crystal structure can be difficult. Computational chemical shift predictions can facilitate NMR-driven structure determination. However, doing so requires one or more candidate crystal structures on which to perform the chemical shift predictions. Crystal structure prediction can provide a powerful and increasingly reliable candidate structures, as evidenced by recent blind tests of crystal structure prediction.[52–54] Given a set of candidate structures, the central computational challenge in NMR crystallography becomes the discrimination between correct and incorrect structures. The more accurately the chemical shifts can be predicted, the greater the discrimination between structures.

### 1.5.1 *Ab-Initio* Fragment-based Shielding Calculations

Periodic density functional theory (DFT) has come to play a central role in NMR spectral assignment and structure elucidation of crystalline systems. In particular the planewave DFT-based gauge-inducing projector augmented wave (GIPAW)[138,139] is one of the more commonly used models and has shown great success in NMR crystallography.[140,141]

Despite their widespread success, planewave DFT methods commonly suffer from two main limitations. First, they are limited in practice to generalized gradient approximation (GGA) functionals. Hybrid density functionals improve the accuracy of the chemical shifts but they typically require at least an order of magnitude more computational effort to evaluate in a planewave basis. Secondly, the mapping of absolute chemical shifts obtained from calculations to experimentally determined chemical shifts are generally done using a linear regression model. These regression parameters are specific to the given functional/basis set combination.[104] Regression models obtained from GIPAW are not transfer-

able to chemical shielding calculations computed from all-electron models in non-periodic systems.

Modern fragment-based methods provide a competitive alternative to planewave methods for chemical shift prediction.[142] Fragment-based methods have been shown to improve the accuracy of predicted chemical shifts by ∼20-30% relative to GIPAW. Additionally, due to the many-body breakdown, each calculation has a lower computational cost and are highly parallelizable which enable the chemical shifts of even complicated crystals to be computed within a few hours.[12,143,144]

The chemical shielding tensor is defined as the second derivative of the electronic energy with respect to the $\alpha$-th component of the external magnetic field ($B_\alpha$) and the $\beta$-th component nuclear magnetic moment experienced on atom A ($\mu_\beta^A$).

$$\sigma_{\alpha\beta}^A = \frac{\delta^2 E}{\delta B_\alpha \delta \mu_\beta^A} \tag{1.11}$$

Applying eq 1.11 to the many-body decomposition of the electronic energy in eq 1.1 yields

$$\tilde{\sigma}^A = \sum_i \sigma_i^A + \sum_{ij} \Delta^2 \sigma_{ij}^A + \sum_{ijk} \Delta^3 \sigma_{ijk}^A + ... \tag{1.12}$$

This expression corresponds to expressing the chemical shielding of atom $A$ in the crystal, $\tilde{\sigma}^A$ in terms of the chemical shielding $\sigma^A$ computed for the isolated monomer plus a series of corrections to that shielding due the monomers interactions with other molecules in the lattice. Given the high computational cost of computing three-body (trimer) interactions, Eq 1.12 is truncated after the two-body terms. To account for the polarization/electrostatic effects that are neglected by this truncation, the monomer and dimer calculations are electrostat-

ically embedded in a set of self-consistent point charges designed to mimic the crystalline lattice,

$$\tilde{\sigma_i}^A = \sum_i \sigma_i^{A,emb} + \sum_{ij} \Delta^2 \sigma_{ij}^{A,emb} \tag{1.13}$$

### 1.5.2 The SCRMP model

Truncation to two-body fragments can introduce errors in the form of missing long-range and many-body polarization effects arising from the crystalline environment. One way to address this is to embed the system with electrostatic point charges that are fitted to mimic the crystalline environment. For example the embedded-ion method (EIM)[145,146] and the surface charge representation of the electrostatic potential (SCREEP)[147] mimic the Madelung potential experienced by a given molecule inside an infinite crystal lattice by embedding the molecule of interest in an optimized set of point charges. The self-consistent reproduction of the Madelung potential (SCRMP) model adapts ideas from the EIM and SCREEP in order to construct a point charge array which more accurately mimics the many-body electrostatic environment in the infinite crystal.

Atom-centered charges are first computed for each isolated monomer in the unit cell using charges from electrostatic potentials using a grid-based method (CHELPG).[148] CHELPG avoids errors associated with truncated multipolar expansions that are more common with charges computed with gaussian distributed multiple analysis (GDMA).[149–151] A number of probe points (N) are then placed on replicated atom centers within a radius $R_{QM}$ around the region of interest. For the two-body fragment calculations the region of interest consists of the monomers located within the unit cell. The CHELPG charges are

then used to compute the Madelung potential ($V_{MP}(r_i)$) at each of these N probe point according to Ewald[152] formalism under tinfoil boundary conditions

$$V_{MP}(r_i) = \sum_{n}\sum_{j}^{N} q_j \frac{erfc(\sqrt{\kappa}|\mathbf{r_i} - \mathbf{r_j} + \mathbf{n}|)}{|\mathbf{r_i} - \mathbf{r_j} + \mathbf{n}|} + \sum_{\mathbf{k}\neq\mathbf{0}}\sum_{j}^{N} q_j (\frac{4\pi}{Ck^2}) e^{i\mathbf{k}\cdot(\mathbf{r_i}-\mathbf{r_j})} e^{-\mathbf{k^2}/\mathbf{4}\kappa} \tag{1.14}$$

where $q_j$ is the partial charge of the $j$th atom located at $\mathbf{r_j}$, $C$ is the volume of the unit cell, $\mathbf{n}$ is the real space vector, and $\mathbf{k}$ represents the reciprocal space vector.

Beyond the $R_{QM}$ radius a two-layer point charge array is then constructed. In the first layer the CHELPG point charges are placed on all atoms within a radius $R_{fix}$ of the asymmetric unit ($R_{fix}$ is typically chosen to be 30 Å due to empirical testing).[12] Next an additional shell of N point charges ($\Delta R_{opt}$) are placed around the first layer of fixed CHELPG charges ($R_{fix}$). These outer N charges are optimized such that the sum of the potentials from the fixed charges ($V_{fix}$) and the optimized charges ($V_{opt}$) approximates the Madelung potential ($V_{MP}$) at every probe point $r_i$:

$$V_{MP}(r_i) \approx V_{fix}(r_i) + V_{opt}(r_i) \tag{1.15}$$

The charge fitting of the outer shell is done using a least-squares minimization of the objective function

$$||V_{MP}(r_i) - V_{fix}(r_i) - \frac{1}{|r_i + r_j|} q_j||^2 = ||b - Ax||^2 \tag{1.16}$$

where $r_j$ is the position of the charge $q_j$ in the outer shell.

New CHELPG calculations are then carried out on each monomer in the unit cell with the optimized charge array embedded. The new charges are then compared to the

previously computed charges to see if the root-mean-square (RMS) deviations are less than $10^{-3}$ atomic charge units. If not, the Madelung potential is re-evaluated using the new CHELPG charges, the CHELPG charges are translated onto the all atom centers within $R_{fix}$, and charges in $R_{opt}$ are refitted to mimic the new Madelung potential. This process will continue until self-consistency is achieved. See Ref 12 for further details.

While the SCRMP method can be used to address some of the missing long-range many-body and polarization effects, the fragment method can fail to capture more local many-body effects. Typically these are addressed by doing a more expensive hybrid fragment/cluster approach. For $^{13}$C and $^{15}$N chemical shielding calculations this has been shown to have negligible effect on the chemical shift predictions.[143,144] The SCRMP model has been found to be particularly important for treating $^{17}$O chemical shifts and for crystals with polar unit cells. Previous studies have shown that combining SCRMP-embedded fragment or cluster/fragment calculations with the hybrid PBE0 functional allows one to predict chemical shifts with accuracy that matches ($^{17}$O) or exceeds ($^{1}$H, $^{13}$C, and $^{15}$N) the accuracy of the widely used GIPAW PBE model.[12,153]

## 1.6 Outline of the Dissertation

Controversy has long surrounded the high-pressure phase diagram of carbon dioxide. Obtaining high-quality diffraction data for phase III has proved challenging, and the currently accepted structure was extracted from powder X-ray diffraction on a sample believed to contain a mixture of phases I and III. However, high-level *ab initio* calculations suggest that phases III and VII are actually the same. In Chapter 2, I will show an applica-

tion of the crystal structure prediction program USPEX to investigate whether an unknown structure can plausibly explain the phase III powder X-ray diffraction pattern.

Accurate NMR chemical shift calculations often requires knowledge of experimental crystal structures. However it is not always guaranteed that the structure for a given spectrum will be known. While it is possible to get candidate structures from a crystal structure prediction search, these structures will not fully reproduce the experimental structure. Specifically, whereas typical solid state NMR experiments are performed at room temperature, most crystal structure prediction studies are performed with classical force fields or electronic structure models that predict the electronic energy, rather than the free energy. These calculations neglect the vibrationally-induced thermal expansion of the crystal which would introduce inherent error in an NMR chemical shift simulation. Chapter 3 gives an evaluation of the errors in $^{13}$C and $^{15}$N chemical shifts obtained by neglecting this thermal expansion. We demonstrate how one can improve the accuracy of the chemical shifts by computing them on quasi-harmonic room-temperature structures, obtaining results that rival those computed with the experimental unit cell parameters.

As previously mentioned in Section 1.4, the quasi-harmonic approximation is well known to give good agreement with experiment for predicted equation of states, bulk moduli, entropy and enthalpies of sublimation, and other related thermodynamic properties for small molecules. However, oftentimes computing the energies, structures, and phonons dispersion-corrected MP2 or CCSD(T) is required to reach this level of agreement. At this level it quickly becomes infeasible to model anything beyond 20 atoms per molecule, which greatly limits the systems that can be studied. In Chapter 4, I will discuss how we can use periodic

density functional theory to expedite the evaluation of the quasi-harmonic approximation with an acceptable loss in accuracy. In Appendix D, I outline how we can use this new procedure to obtain the thermally-expanded structures.

Finally, a previous study has shown that the quasi-harmonic approximation can be used to model the phase-transition of crystalline methanol accurately. While that phase diagram prediction was highly successful, it required some 300,000 CPU hours. Performing similar predictions in a larger organic molecule would be cost prohibitive. Chapter 5 presents an effort to utilize the new, lower-cost QHA approach developed in Chapter 4 to predict the phase transition properties of $\alpha-$ and $\beta-$ resorcinol.

# Chapter 2

# Theoretical Predictions Suggest Carbon Dioxide Phases III and VII are Identical

In the previous chapter we discussed how one might go about performing a crystal structure prediction search. In this chapter we demonstrate one example of a crystal structure prediction search which attempts to generate candidate structures to confirm whether phase III of carbon dioxide can be produced using electronic structure theory methods. The work presented in this chapter was originally included in Ref 26 as part of a study trying to understand the strong structural and spectroscopic similarities between carbon dioxide phases III and VII. This chapter focuses primarily on the crystal structure prediction component of the full study.

## 2.1 Introduction

Carbon dioxide is one of the most fundamental chemical species on earth, yet its solid-phase behavior at high pressures continues to confound. Starting with the 1994 powder X-ray diffraction (PXRD) structure of phase III,[10] researchers have mapped out a rich phase diagram with 8-10 crystalline phases that range from molecular crystals at lower pressures to extended covalent and ionic phases at high pressures.[154] Experimental characterization of these crystal structures and their solid-solid phase boundaries has often proved challenging, with considerable kinetic path-dependence and hysteresis in the phase transitions, difficulty in obtaining high-quality diffraction data, and sharp pressure gradients within samples that complicate spectroscopic measurements. As a result, the literature on high-pressure carbon dioxide contains numerous contradictory experimental interpretations. Fortunately, substantial advances in computational modeling of molecular materials[100] mean that *ab initio* calculations can help resolve such controversies and play an integral role in molecular crystallography. Using high-level electronic structure calculations, the present study investigates several molecular crystal phases of carbon dioxide and demonstrates quantitative agreement between predicted and experimentally observed structural, mechanical, and spectroscopic properties for several of them. However, the same theoretical evidence indicates that the long-accepted structure of phase III carbon dioxide is inconsistent with spectroscopic data, and that phases III and VII are likely identical.

Controversy has long surrounded the high-pressure phase diagram of carbon dioxide (Figure 2.1). Substantial research developed a framework in which lower pressure molecular crystal phases I, III, and VII transition to "intermediate bonding" phases II and IV

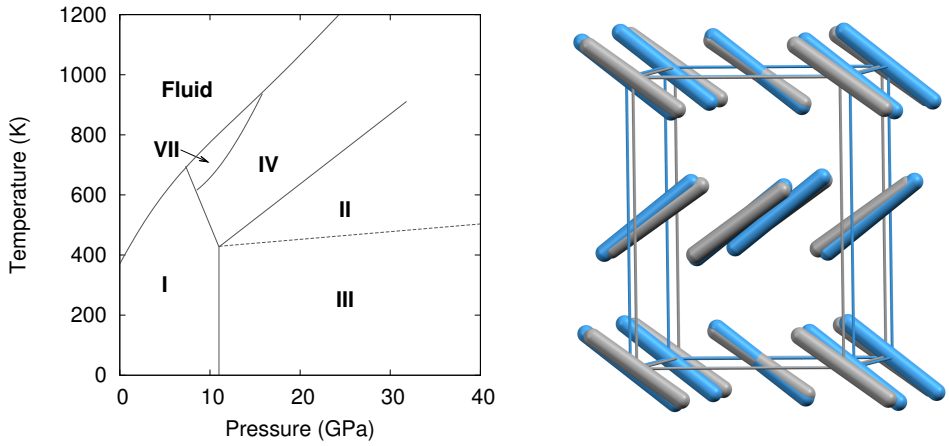



Figure 2.1: Phase diagram of carbon dioxide up to 40 GPa, and structure overlay of the experimental crystal structures for phases III (blue) and VII (gray). Root-mean-square deviation[5] = 0.24 Å.

at moderate pressures (∼20–40 GPa), before eventually forming extended covalent or ionic phases at higher pressures.[154] The intermediate bonding phases purportedly exhibit elongated and/or bent carbon dioxide molecules and abnormally large bulk moduli.[155–157] However, subsequent experiments[7,8] and density functional theory (DFT) calculations[158,159] have challenged this interpretation, suggesting instead more traditional molecular crystal structures at intermediate pressures, with linear carbon dioxide molecules and typical bulk moduli.

Phases III and VII represent another conundrum. X-ray diffraction studies suggest that both phases adopt similar *Cmca* space group structures (Figure 2.1).[9,10] The primary differences lie in effectively swapping the $a$ and $b$ lattice constants and slightly altering the angle the molecule forms relative to the $c$ crystallographic axis. Phase III can be formed at room temperature by compressing phase I (dry ice) to pressures above ∼12 GPa,

though the precise phase boundary remains unclear due to the sluggish martensitic phase change.[154] Obtaining high-quality diffraction data for phase III has proved challenging, and the currently accepted structure was extracted from powder X-ray diffraction on a sample believed to contain a mixture of phases I and III.[10] Despite routinely being included in the low-temperature region of the phase diagram, phase III is actually believed to be metastable and monotropic relative to phase II. It converts to phase II upon annealing to ∼500 K at 12 GPa or above.[7,155]

Phase VII occurs in a narrow pressure and temperature region around 15–17 GPa and 750 K,[154] and producing it experimentally can also be challenging.[9,160] Nevertheless, its structure was determined via X-ray diffraction on single crystals grown from the melt. Given the difficulty of obtaining quality diffraction data for phase III and the correspondingly poor constraints on its structure,[6,9,161] the possibility that phases III and VII were actually the same phase was raised immediately.[9] However, the non-contiguous existence domains for III and VII in the phase diagram and subtle differences in the Raman spectra were cited in favor of there being two distinct phases.[9]

Resolving these issues experimentally has proved challenging. *Ab initio* crystallography plays an increasingly important role in molecular crystals, materials, and even biological systems. Computational refinement of experimental crystal structures has long been integral in many studies, and advances in crystal structure prediction[162,163] have made *ab initio* structure determination even more viable. Unfortunately, energy alone is often an insufficient descriptor—one commonly predicts multiple potential structures whose energies are sufficiently close so as to prevent clear discrimination. By predicting and com-

paring additional spectroscopic observables such as infrared, Raman, or nuclear magnetic resonance spectra, however, one can markedly increase confidence in the structural assignments.[100,133,164–166]

Several molecular crystalline phases of carbon dioxide were revisited with large basis, quasi-harmonic second-order Møller-Plesset perturbation theory (MP2) electronic structure theory calculations.[26,105,106] These high-level calculations quantitatively reproduce structures, mechanical properties, and Raman spectra across most of the phases considered. However, these models do not predict a distinct phase III structure whatsoever. Moreover, even if the experimental structure were correct, the predictions indicate that its Raman spectrum would differ from the experimentally observed spectra. In contrast, the predictions for phase VII are highly consistent with those observed experimentally for phase III.

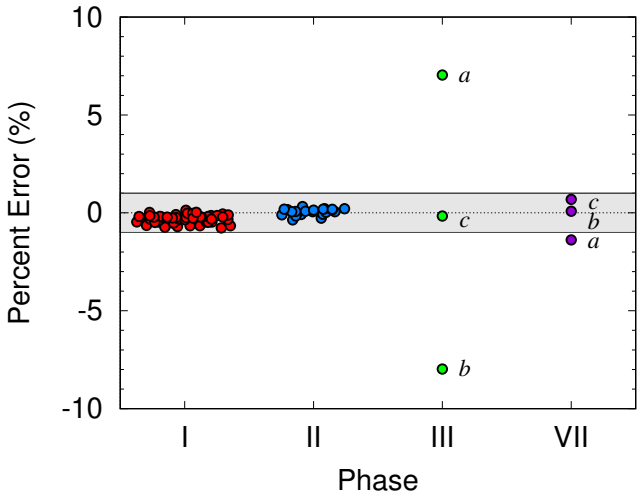



Figure 2.2: Errors in the predicted lattice constants versus experiment. The shaded band indicates ±1% error.

Both conventional electronic and quasi-harmonic MP2 free energy relaxation of the

experimental phase III structure always converge to the phase VII structure, regardless of temperature or pressure. As shown in Figure 2.2, the errors in the $a$ and $b$ lattice parameters relative to the experimental crystal structure are 5–10 fold larger than those for the other phases. In fact, no reported electronic structure calculation on phase III predicts a distinct phase III structure.[158,159,167–169] Even if phase III is only metastable relative to phase II as inferred experimentally, it should exist as a local minimum on the free energy surface that is distinct from phase VII.

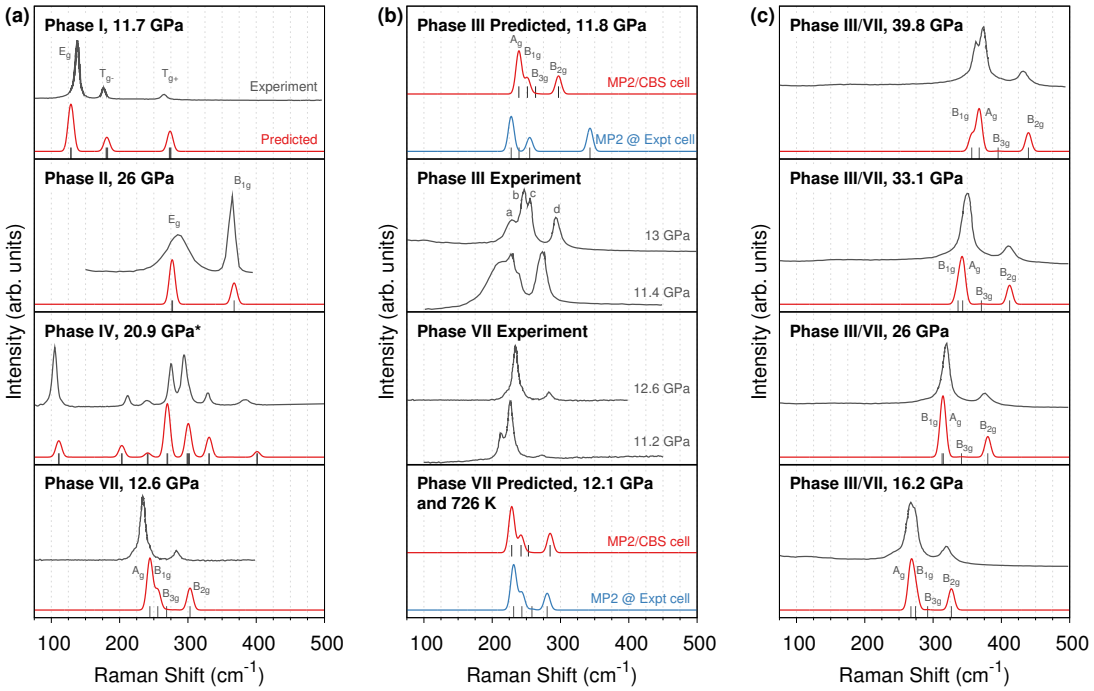



Figure 2.3: (a) Comparison of MP2-predicted and experimental lattice phonon Raman spectra for phases I,[6] II,[7], IV,[8] and VII[9] carbon dioxide. (b) Comparison of Raman spectra for phases III and VII,[6,9] including predicted spectra using either MP2/CBS or experimentally determined unit cell parameters. (c) Pressure dependence of the experimental phase III[6] and predicted phase VII Raman spectra. All spectra are at room temperature unless otherwise indicated. *The Phase IV spectrum employs an MP2/aug-cc-pVDZ optimized cell instead of a quasi-harmonic MP2/CBS one.

Raman spectroscopy provides further evidence for problems with the phase III structure[26]. Figure 2.3 compares room-temperature experimental Raman spectra against those computed at the MP2/aug-cc-pVDZ level using unit cell parameters from the quasi-harmonic MP2/CBS-limit calculations. Predicted Raman peak positions and intensities generally agree very well with experiment for several different molecular phases of carbon dioxide. For phases I and II, the predicted peaks lie within $\sim$10 cm$^{-1}$ of experiment. For phase VII, agreement between the predicted and experimental structures is also reasonably good, with the predicted frequencies $\sim$10–15 cm$^{-1}$ too high. The positions and intensities of these peaks in the MP2/CBS phase VII structure are consistent with both the experimental phase III and VII spectra. Of course, that consistency alone is insufficient to rule out the phase III structure.

Although the phase III structure is not a stationary point on the free energy surface, one can start with the purported experimental structure of phase III, freeze the lattice constants, relax the atomic positions, and predict the Raman frequencies and intensities. The phase III structure optimized with fixed experimental lattice constants reproduces the claimed experimental structure very well but shifts the highest frequency band $\sim$50 cm$^{-1}$ higher than experiment which is several-fold larger than the disagreements observed between theory and experiment for any of the other phases examined here. For comparison, performing the same procedure on the 726 K experimental phase VII structure[9] results in a predicted spectrum that is in excellent agreement with both the MP2/CBS cell Raman spectrum at the same temperature and the experimental room-temperature Raman at $\sim$11-13 GPa. In other words, while the experimentally reported phase VII structure is consistent

with the predicted one in terms of both structure and Raman activity, the putative phase III structure is neither a stationary point on the free energy surface, nor is its predicted Raman spectrum compatible with the experimentally observed one.

One obvious alternative is that phase III and VII are in fact the same, as was first raised (and subsequently discounted) by Giordano and Datchi.[9] Raman spectroscopic evidence supports this hypothesis. The phase VII predictions quantitatively reproduce the pressure dependence of the phase III Raman spectra over tens of GPa (Figure 2.3c). However the possibility that another unknown structure exists which can account for the experimental data can be explored using crystal structure prediction.

## 2.2 Computational details

To investigate the possibility that an unknown structure can account for the phase III experimental data, evolutionary algorithm-driven crystal structure prediction was performed using USPEX[170]. Each of six runs was seeded with ten random structures from randomly chosen space groups, containing either two or four molecules in the unit cell (the unit cell sizes for phases I, II, and III/VII). Structures were relaxed at ambient pressure and energies computed using Tinker[171] and the OPLS-AA force field.[172] New structures were constructed for 15–20 generations via heredity, coordinate/rotational mutations, or lattice mutation.[170] This process generated 660 structures with $Z = 2$ and 1083 structures with $Z = 4$ were generated, though many of these were redundant or clearly energetically unfavorable.

After removal of obvious duplicates, the most stable 91 structures were then refined under 11.8 GPa of pressure (the pressure for the experimental powder X-ray diffraction of phase III) using density functional theory (DFT)[1]. Structures were analyzed in terms of energy, unit cell volume, root-mean-square deviations in atomic positions, and by comparing simulated powder X-ray diffraction patterns (wavelength 0.6888 Å) generated by Mercury.[173] These were then compared against the experimental phase III diffraction pattern and the simulated pattern for the experimentally reported phase III crystal structure.[10]

## 2.3 Results and Discussion

The crystal structure prediction generated Phase I, II, and VII, along with 22 other potential structures within 10 kJ/mol of the most stable one (Figure 2.4). The known experimental structures for phase IV was not found in the search since it has more than four molecules in the unit cell. However, none of the other structures has a simulated powder X-ray diffraction pattern that is plausibly consistent with the experimental phase III one (Figure 2.5). Simulated powder X-ray diffraction patterns were also examined for the stable 91 structures that were generated on the OPLS-AA landscape however no plausible match with the experimental phase III was found here either.

The 1994 PXRD experiment that was originally used to solve the structure of phase III was performed on a sample believed to consist of a mixture of phase I and III.[10] Examination of all the diffraction patterns reveals that none of the structures which are not experimentally known provides a good match to the experimental PXRD. The best matches come from either the claimed phase III structure or the phase VII one (with

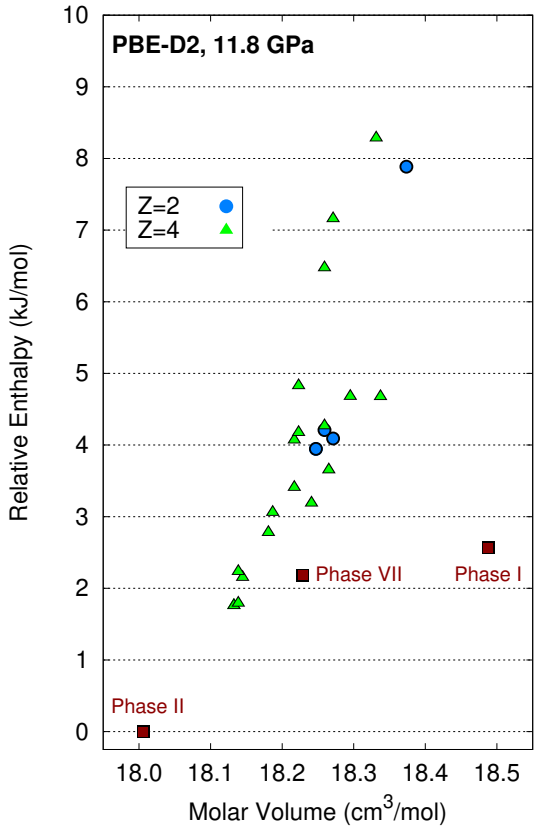



Figure 2.4: PBE-D2 crystal energy landscape for potential carbon dioxide crystal structures at 11.8 GPa with $Z = 2$ or 4 molecules in the unit cell.

some mixture of phase I). Of course, the potential for structures with a different number of molecules in the cell or otherwise missed by this crystal structure prediction cannot be ruled out. Nevertheless, in the absence of other viable phase III structures, the most obvious alternative is that phase III and VII are in fact the same, as was first raised (and subsequently discounted) by Giordano and Datchi.[9]

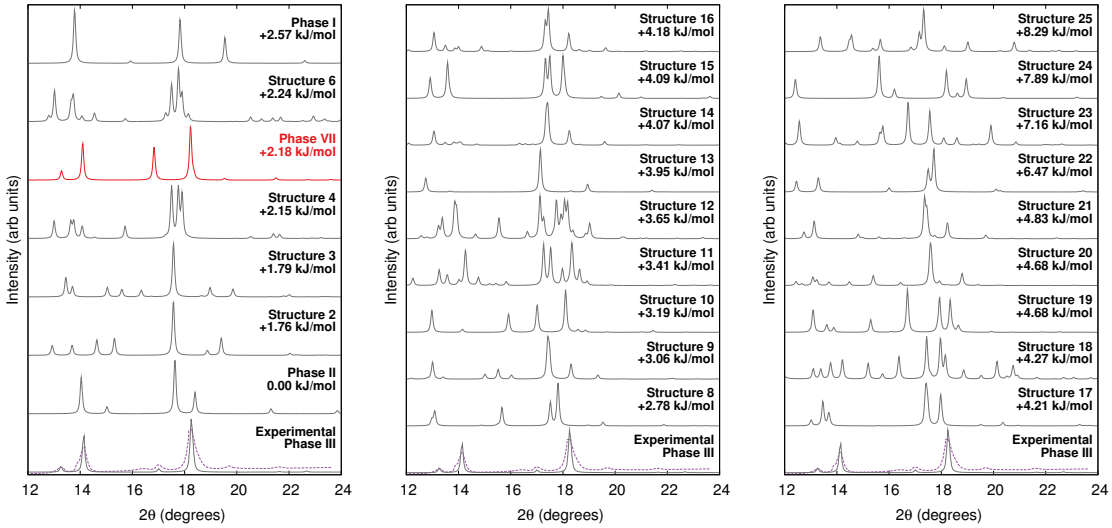



Figure 2.5: Simulated powder X-ray diffraction patterns for the 25 predicted crystal structures which lie within 10 kJ/mol of the stable phase II one. All were optimized at the PBE-D2 level and 11.8 GPa, including those identified as Phase I, II, or VII. These patterns are compared against the actual experimental phase III powder X-ray diffraction pattern (purple dotted line) and the simulated phase III spectrum based on the claimed experimental structure.[10]

## 2.4 Conclusion

To summarize, large-basis, quasi-harmonic MP2 electronic structure calculations accurately reproduce experimentally observed structural, mechanical, and spectroscopic properties for several different molecular crystal phases of carbon dioxide across broad pressure range. However, theoretical optimization of the purported structure for phase III relaxes directly to phase VII. Even if phase III is only metastable, the existence of two distinct phases should translate to two separate free energy basins, but only one is found computationally. Furthermore, the claimed experimental phase III structure would produce a Raman spectrum in the librational region that disagrees with the experimentally observed ones. In contrast, Raman spectra predicted for phase VII agree well with the experimentally

observed phase III and VII ones over a broad pressure range. Crystal structure prediction was employed to investigate whether an unknown structure could plausibly reproduce the experimental phase III powder X-ray diffraction pattern. Despite successfully finding phases I, II, and VII, no structure was found that could plausibly explain phase III other than phase VII.

Based on the above results and the failure to identify a plausible alternative structure, we propose that phases III and VII are in fact the same. Although this hypothesis accounts for the data discussed above, questions remain. While the subtle differences in the experimental phase III and VII Raman spectra might be attributed to the variations arising from microstrain or other experimental complications, the apparent disconnect between the phase III and VII regions in the phase diagram is more difficult to rationalize. On the one hand, there is no contradiction in having phase III/VII be kinetically accessible in the phase III region and thermodynamically stable in the phase VII region. Instead of phase III being monotropically related to phases II and/or IV, phase III/VII would be enantiotropically related to them.

On the other hand, if the two phases are the same, why is phase VII seemingly difficult to form from phase I (it is typically formed from the melt instead),[160] while phase III forms readily? Perhaps given the sluggish nature of the phase I→III/VII transition and the narrow region of phase VII stability, the transformation from I→VII upon isothermal compression near 725 K is incomplete before one enters the region of phase IV stability. Similarly, why can one not form phase III kinetically via isothermal compression of phase I, then heat it to the phase VII region of thermodynamic stability without it transforming to

phase II? If the kinetic barrier to transforming phase III→II is relatively small, maybe heating metastable phase III/VII from ambient temperatures provides sufficient thermal energy to convert to the more stable phase II before one reaches the regime of phase III/VII thermodynamic stability. Interestingly, one can quench phase VII down to room temperature, suggesting that the rate of heating may be significant. New experiments that investigate the crystal structure of phase III and its relationship to phase VII are clearly needed.

# Chapter 3

# Improving Predicted Nuclear Magnetic Resonance Chemical Shifts Using the Quasi-Harmonic Approximation

Crystal structure prediction (CSP) is increasingly being used in conjunction with solid state nuclear magnetic resonance (NMR) spectroscopy for molecular crystal structure determination. A typical CSP run will produce anywhere from tens to hundreds of candidate structures that must be distinguished between. The more accurately the chemical shifts can be predicted, the greater the discrimination between candidate structures. Whereas typical solid state NMR experiments are performed at room temperature, most CSP studies are performed with classical force fields or electronic structure models that

neglect the zero-point vibrational and finite temperature effects which lead to thermal expansion. Since molecular crystals can expand appreciably between their electronic minima and room temperature, this introduces an error in the predicted chemical shifts. In this chapter we will quantify the errors that are accrued by neglecting thermal expansion of the crystal and investigate how accounting for thermal expansion in molecular crystals via the quasi-harmonic approximation (QHA) refines isotropic $^{13}$C and $^{15}$N NMR chemical shift predictions. The work presented in this chapter has been submitted for publication.[174]

## 3.1 Introduction

Structural characterization of molecular crystals is fundamental in chemistry. Single-crystal diffraction techniques represent the gold standard for structure determination, but alternative strategies are needed for cases where suitable single crystals are not obtainable. Nuclear magnetic resonance (NMR) crystallography represents one such alternative approach. It combines solid-state NMR, powder x-ray diffraction, and *ab initio* chemical shielding predictions to solve crystal structures.[133,135,140,141,175–179] Diffraction methods characterize the long-range order effectively, but they can sometimes have difficulty resolving local features such as protonation states. Solid-state NMR complements powder x-ray diffraction by providing detailed information about the local chemical environments. However, mapping from the chemical shifts in an NMR spectrum to a three-dimensional crystal structure can be difficult.

Computational chemical shift predictions can facilitate NMR-driven structure determination. However, doing so requires one or more candidate crystal structures on which

to perform the chemical shift predictions. One might obtain such structures from partial solutions from ambiguous powder x-ray diffraction data.[180] Alternatively, crystal structure prediction can also provide a powerful and increasingly reliable candidate structures, as evidenced by recent blind tests of crystal structure prediction.[52–54] Given a set of candidate structures, the central computational challenge in NMR crystallography becomes the discrimination between correct and incorrect structures. The more accurately the chemical shifts can be predicted, the greater the discrimination between candidate structures. For example, switching from a generalized gradient approximation (GGA) density functional like PBE to a hybrid functional like B3LYP or PBE0 reduces the chemical shift errors by 20–30%[12,104] and increases the structural discrimination.[144]

The electronic structure model used to predict the chemical shifts is not the only source of error, however. The quality of the predicted geometry also matters. If the experimental structure is known, DFT refinement of the atomic positions (both hydrogen and heavy atoms) while constraining the experimental lattice parameters frequently leads to better agreement with neutron diffraction structures[181] and smaller differences between the measured and predicted chemical shifts.[104,143] The situation becomes more challenging when the experimental structure is unknown and has been predicted via crystal structure prediction or other similar lattice energy minimization modeling techniques. Whereas typical solid state NMR experiments are performed at room temperature, most crystal structure prediction studies are performed with classical force fields or electronic structure models that predict the electronic energy, rather than the free energy. By doing so, they neglect the zero-point vibrational and finite temperature effects that lead to thermal expansion.

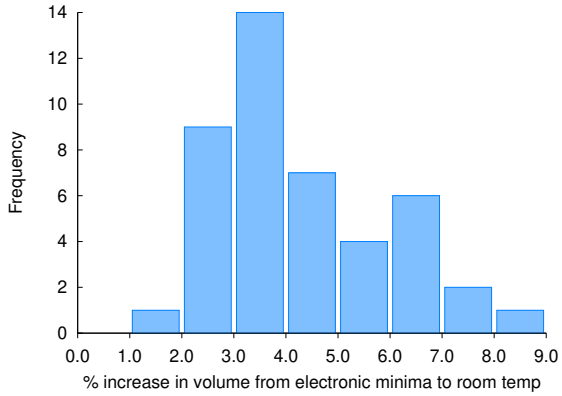



Figure 3.1: Total volume expansion of 44 molecular crystals. On average the unit cell will expand by 4.3% of its total volume. See Appendix A section 8 for the structures used in this graph.

While a typical molecular crystal might expand its molar volume by ∼3–4% between the electronic minimum energy structure and the experimental room temperature structure, systems dominated by weaker intermolecular interactions such as van der-waals or $\pi - \pi$ stacking can exhibit volume expansion of nearly ∼9% (Figure 3.1). Surveys by Nyman and Day found that accounting for these free energy effects can reorder the relative energetic stabilities of molecular crystal polymorphs in up to 21% of cases.[62,63]

Various strategies exist for incorporating the free energy effects into the simulation. Molecular dynamics techniques naturally accounts for finite-temperature effects (though standard classical dynamics omits zero-point vibrational contributions). However, force fields do not always provide the requisite accuracy required for resolving the energetics of crystal polymorphs. Free energy estimates based on harmonic phonon calculations account for a significant fraction of the vibrational contribution, but they neglect thermal expansion of the crystal. The quasi-harmonic approximation (QHA) provides a computationally practical strategy for improving on the purely harmonic model and capturing this thermal

expansion and its impact on crystal properties. The QHA models how the Gibbs free energy depends on temperature and pressure by incorporating a dependence of the harmonic phonons with volume. A number of studies have shown use of the QHA with DFT and other correlated wavefunction approaches allows the prediction of small-molecule structures and their associated properties in excellent agreement with experiment.[61,105–107,126–130,132,182] Of course, the QHA model has its own limitations. Its static harmonic approach can be too simplistic to describe more flexible molecules and dynamical motions in a crystal, and it is expected to break down at higher temperatures.

Here we investigate how accounting for thermal expansion in molecular crystals via the quasi-harmonic approximation refines isotropic $^{13}$C and $^{15}$N NMR chemical shift predictions. First, using a set of twenty small-molecule molecular crystals, we demonstrate that accounting for thermal expansion effects increases the accuracy of the chemical shift predictions moderately, and that the chemical shifts computed using quasi-harmonic room-temperature structures rival those based on the experimental unit cell parameters. While the statistical reductions in error are modest on the whole, selected cases are found where the improvements to individual chemical shifts are significant. In other words, properly accounting for thermal expansion may be helpful in NMR crystallography applications where discriminating between different candidate structures proves challenging. Second, the capabilities of this QHA technique are demonstrated by applying the approach to several pharmaceutical crystals. Finally, we conclude by examining how QHA-based chemical shift refinement helps in the context of discriminating among different candidate structures of theophylline, which represents a challenging case for NMR crystallography.

## 3.2 Theory

### 3.2.1 Quasi-Harmonic Approximation

To model molecular crystals at finite temperatures and pressures the Gibbs free energy must be computed. The Gibbs free energy is composed of the electronic internal energy $U_{el}$, the Helmholtz vibrational free energy $F_{vib}$, and a pressure-volume ($PV$) contribution,

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

At ambient pressure, the $PV$ term contributes negligibly to the total energy and hence will be neglected in this work. In that situation, the Gibbs free energy reduces to the Helmholtz free energy $F$.

The internal electronic energy $U_{el}$ is computed here via dispersion-corrected planewave DFT. The Helmholtz vibrational free energy is computed from standard harmonic oscillator vibrational partition functions as,

$$F_{vib}(T) = N_a \sum_i \left( \frac{\hbar\omega_i}{2} + k_b T \ln\left[1 - \exp\left(-\frac{\hbar\omega_i}{k_b T}\right)\right]\right) \tag{3.2}$$

where $N_a$ is Avogadro's number, $\hbar$ is Plank's constant, $k_b$ is the Boltzmann constant, and $\omega_i$ is the vibrational frequency of mode $i$. The first term corresponds to the zero-point vibrational contribution, while the second term describes the temperature dependence. For computational expedience, the phonons are computed only at the $\Gamma$ point in the present work.

The crystal structures which minimize the electronic energy $U_{el}$ (i.e. those obtained from relaxing both atomic positions and unit cell parameters) are referred to as the reference structure. We compute harmonic phonons for the reference crystal ($\omega_i^{ref}$) to ensure the

optimized structure lies at a minimum on the potential energy surface. Electronic energy versus volume curves $U_{el}(V)$ are then mapped out by minimizing the DFT energy of the reference cell under positive (cell compression) or negative pressure (cell expansion). This approach allows the cell to deform anisotropically along the lowest-energy path, which can be important for reliable modeling of the free energies.[108,109]

In principle, one ought to re-evaluate the phonons at every volume along this potential energy surface slice, but that is computationally impractical given the cost of typical electronic structure theory calculations. Instead we employ mode-specific Grüneisen parameters $\gamma_i$ to estimate how individual phonon modes vary with unit cell volume,

$$\gamma_i = -\left(\frac{\partial \ln \omega_i}{\partial \ln V}\right) \tag{3.3}$$

Integrating Eq 3.3 yields,

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

which allows the reference phonons $\omega_i^{ref}$ to be scaled to any given new volume $V$.

To compute these Grüneisen parameters, harmonic phonons are computed at one expanded and one compressed volume selected from those geometries obtained along the $E(V)$ curve. The selected structures were typically chosen to be about $\pm 10$ Å$^3$ away from the reference volume here.[105,106] The Grüneisen parameter for each mode is then obtained via finite difference of these phonons computed on the expanded and compressed structures.

Using the calculated reference phonons and the mode-specific Grüneisen parameters, the Helmholtz vibrational free energy is evaluated at each volume in the electronic energy curve over a range of temperatures. Summing these two curves together gives the free energy. Since the particular volume which minimizes the free energy for a given tem-

perature is unlikely to correspond to one of the sampled volumes, each free energy curve is fitted to a weighted double-Murnaghan equation of state. The Murnaghan equation of state is given by,

$$F(V) = F_0 + \frac{B_0 V}{B_0'}\left[\frac{(V_0/V)^{B_0'}}{B_0' - 1} + 1\right] - \frac{B_0 V_0}{B_0' - 1} \tag{3.5}$$

where $F_0$, $V_0$, $B_0$, and $B_0'$ are the fit parameters. $F_0$ gives the free energy at the minimum, $V_0$ is the molar volume at the minimum energy, $B_0$ is the bulk modulus, and $B_0'$ is the first derivative of the bulk modulus with respect to pressure. This method identifies the free-energy minimum effectively while avoiding artifacts that can be caused by simpler equation of state fits or splines.[108]

Once the free energy equation of state $F(V)$ has been obtained at a chosen temperature, the free energy can be minimized to find the optimal molar volume. The lattice constants and atomic positions at this optimal volume are interpolated based on the explicitly optimized structures that were obtaining in generating $U_{el}(V)$. The atomic positions are then relaxed subject to those lattice constants being fixed to ensure the structure is at a minimum on the QHA free energy surface.

The computational cost of these QHA structure determinations is considerably higher than that of the conventional approach, which would typically involve only a single relaxation of the crystal structure followed by a single chemical shift prediction. In the QHA approach used here, the crystal was relaxed at ∼15-20 different pressures to map out $U_{el}(V)$. Furthermore, the harmonic phonon frequencies were computed three volumes. Each Hessian was evaluated via finite difference of the forces, and it effectively has a computational cost of ∼2-5 times that of a geometry optimization (depending both on the size and internal

symmetry of the molecule). Lastly, one final optimization is employed on the optimal structure at each given temperature. Overall, the computational cost of this QHA approach is ∼20–40 times higher than a conventional structure optimization. It might be possible to reduce the number of structure optimizations performed to construct $U_{el}$, but the QHA costs will remain at least an order of magnitude higher than that of a conventional calculation regardless.

### 3.2.2 *Ab-Initio* Fragment-based Shielding Calculations

Periodic density functional theory (DFT) has long been used for modeling chemical shift calculations as they are well suited to describe extended crystal systems. In particular the planewave DFT-based gauge-inducing projector augmented wave (GIPAW)[138,139] is one of the more commonly used models and has shown great success in NMR crystallography.[140,141] Modern fragment- and cluster-based methods[142,183,184] provide a competitive alternative to planewave methods for chemical shift prediction. Due primarily to their ability to use hybrid density functionals with much lower computational cost, fragment methods can frequently out-perform the more commonly used GIPAW method.[12,143,144] Switching from the GGA functional PBE to the hybrid PBE0 improves the accuracy of the chemical shifts by ∼20-30%.

Fragment-based methods decompose the total energy of the system into a many-body expansion,

$$E_{total} = \sum_{i} E_i + \sum_{ij} \Delta^2 E_{ij} + \sum_{ijk} \Delta^3 E_{ijk} + ... \tag{3.6}$$

where $E_i$ represents the energy of the isolated molecule, $\Delta^2 E_{ij}$ is the pairwise interaction

energy between molecules i and j, and $\Delta^3 E_{ijk}$ is the three-body interaction energy between molecules $i$, $j$, and $k$. Differentiating this expression with respect to the $\alpha$-th component of the external magnetic field $B$ and the nuclear magnetic moment $\mu$ produces a many-body expansion for the chemical shielding tensor $\sigma$ on atom $A$,

$$\tilde{\sigma}^A = \sum_i \sigma_i^A + \sum_{ij} \Delta^2 \sigma_{ij}^A + \sum_{ijk} \Delta^3 \sigma_{ijk}^A + ... \tag{3.7}$$

This expression corresponds to expressing the chemical shielding of atom $A$ in the crystal, $\tilde{\sigma}^A$, in terms of the chemical shielding $\sigma^A$ computed for the isolated monomer plus a series of corrections to that shielding due that monomer's interactions with other molecules in the lattice. Given the high computational cost of computing three-body (trimer) interactions, Eq 3.7 is truncated after the two-body terms. To account for the polarization/electrostatic effects that are neglected by this truncation, the monomer and dimer calculations are electrostatically embedded in a set of self-consistent point charges designed to mimic the crystalline lattice,

$$\tilde{\sigma_i}^A = \sum_i \sigma_i^{A,emb} + \sum_{ij} \Delta^2 \sigma_{ij}^{A,emb} \tag{3.8}$$

See Ref 12 for details. The fragment approach generally captures the impact of the crystalline environment on chemical shieldings well.[12,104] Furthermore, it reduces the overall computational cost of the chemical shift calculation and can take advantage of highly parallel software implementations which enable the chemical shifts of even complicated crystals to be computed within a few hours.[104]

## 3.3 Computational details

### 3.3.1 Systems studied

To evaluate the effect thermal expansion has on predicted chemical shifts, 20 molecular crystals with 68 $^{13}$C and 28 $^{15}$N experimentally measured chemical shifts were chosen (Figure 3.2). These represent a subset of crystals from our earlier benchmark sets.[104,143] The CSD reference codes and experimental isotropic chemical shifts for these two test sets are:

- Carbon (12 structures, 68 shifts): HXACAN13,[185] GLYCIN03,[186] LALNIN12,[187] LSERIN01,[188] LTYROS11,[189] SUCROS04,[190] NAPHTA36[191] ADENOS12,[189] LTHREO01,[189] GLUTAM01,[189] ASPARM03,[192] and LCYSTN21.[189]

- Nitrogen (14 structures, 28 shifts): HXACAN13,[193] GLYCIN03,[194] ALUCAL04,[104] LSERIN01,[104] GLUTAM01,[104] ASPARM03,[192] LCYSTN21,[104] FUSVAQ01,[195] URACIL,[195] BAPLOT01,[134] LHISTD02,[196] BITZAF,[197] CYTSIN,[195] and THYMIN01.[195]

Several updates/modifications were made to these structures compared to the earlier works. First, the previous acetaminophen (paracetamol) benchmarks used the room-temperature HXACAN26 structure instead of the 20 K HXACAN13 one. However, it turns out that even after geometry optimization, HXACAN26 retains an imaginary vibrational frequency that is associated with the hydrogen bonding. The hydrogen bonding network in HXACAN26 differs from that in all other form I structures in the CSD. To rectify this, we instead choose HXACAN13 as the starting structure for the fully optimized crystal and the subsequent QHA expansions. The resulting vibrational frequencies for HXACAN13 are

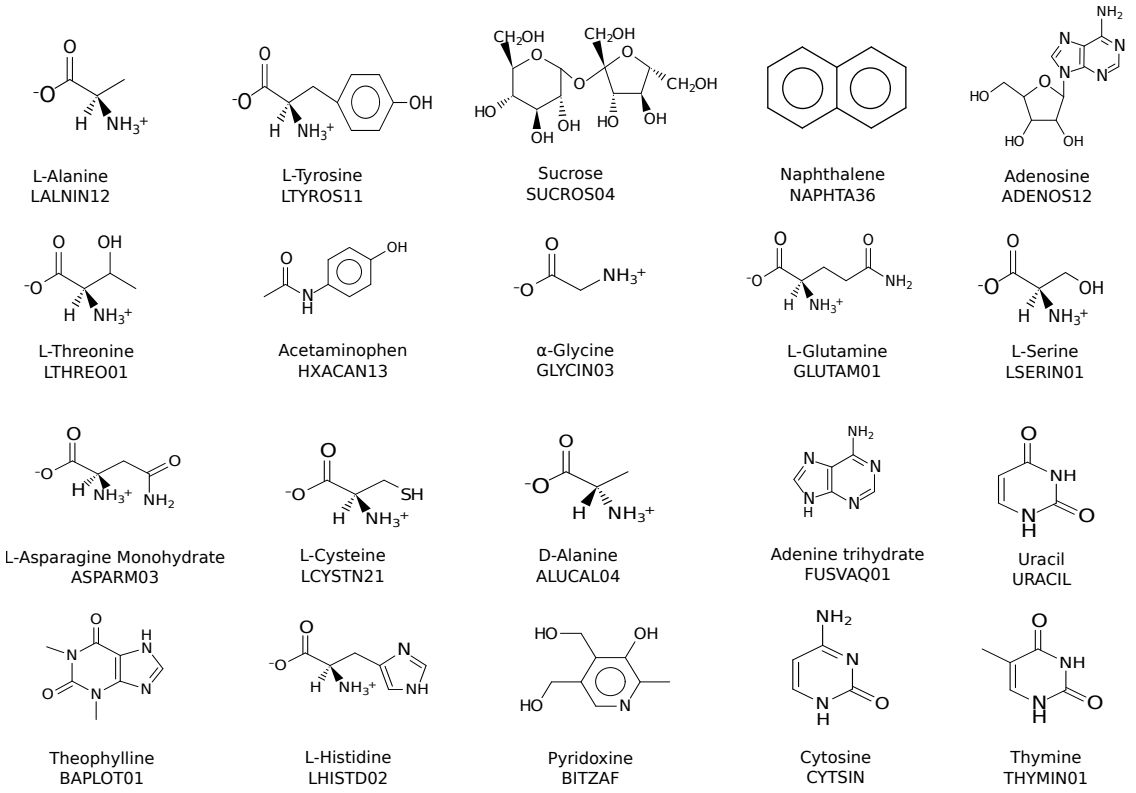



Figure 3.2: The twenty molecular crystals studied here, indicated by their species name and the CSD RefCode.

all real. For the fixed cell optimization on the room-temperature structure, we created a new structure by combining the unit cell parameters from HXACAN26 with fractional coordinates from HXACAN13 and then relaxed the atomic positions. The resulting structure exhibits small root-mean-square deviations relative to other experimental structures in the CSD, indicating that the correct overall packing was preserved.

Second, adenine tri-hydrate currently has two crystal structures reported in the CSD: FUSVAQ at room temperature and FUSVAQ01 at 105 K. These structures differ in their water molecule orientations, with the orientations found in the low-temperature

FUSVAQ01 structure leading to better agreement in the predicted $^{15}$N chemical shifts. To update the room-temperature structure, a new structure was created by combining the unit cell parameters from FUSVAQ with fractional coordinates from FUSVAQ01 and then relaxing the atomic positions. The resulting structure yielded good root-mean-square deviations relative to the original FUSVAQ01 structure and appreciably smaller errors in the $^{15}$N chemical shifts relative to experiment.

Third, experimental $^{13}$C isotropic chemical shifts for ADENOS12, GLUTAM01, LCYSTN21, LTHREO01 and LTYROS11 were recently re-measured by Dračínský et al,[189] correcting errors found in previously published literature that included incorrectly using the racemic instead of enantiopure crystal, chemical shift assignment errors, and referencing issues. Using the revised experimental values improved agreement with the theoretical calculations in all cases, reiterating the importance of obtaining high-quality experimental data when benchmarking theoretical models.

Finally, we investigate five pharmaceutical crystals more closely: acetaminophen (HXACAN13)[185], ibuprofen (IBPRAC16)[198], theophylline (BAPLOT01)[134], carbamazepine (CBMZPN23)[199], and naproxen (COYRUD12)[200]. Fixed-cell geometry optimizations and chemical shift calculations were performed for benchmark comparison on room temperature structures IBPRAC06, BAPLOT01, CBMZPN14, and COYRUD. Again, the acetaminophen room temperature structure was generated as described above. Theophylline candidate structures generated via crystal structure prediction for Section 3.4.5 were obtained from the authors of Ref 134.

### 3.3.2 DFT structure optimizations

Three different optimizations were carried out for each of the structures mentioned here. The first is a full electronic energy optimization of both the atomic positions and the unit cell (labeled "No QHA"). The second uses the QHA to thermally expand the fully optimized structure to its free energy minimum at 0 K (labeled "0 K") and 300 K (labeled "300 K"). For the 0 K structure, the thermal expansion comes solely from zero-point energy (ZPE) contributions which is known to contribute up to 30% of the total expansion of the crystal at ambient temperature and pressure.[106] Finally, to benchmark against the known experimental structure, results were taken from a previous study[12] which performed fixed-cell optimizations on these structures (labeled "Expt. Cell"). In the pharmaceutical crystals for which earlier fixed-cell optimizations were not available, they were performed here.

For the calculations performed here, the crystals were first optimized with periodic boundary planewave DFT in Quantum Espresso v6.1[201], using the B86bPBE density functional[2,3] and the exchange-hole dipole method (XDM) dispersion correction.[4] Core electrons were treated according to the projector augmented wave (PAW) approach, and PAW potentials for H, C, N, O, S, Cl, Br, and F were produced using A. Dal Corso's Atomic code v6.1.[202] Optimizations were carried out using a 50 Ry planewave energy cutoff. Reciprocal space k-points were placed with a Monkhorst-Pack grid[203] with a typical k-point grid spacing of 0.04 $Å^{-1}$ between nearest k-points and maximum spacings of no more than 0.09 $Å^{-1}$. See Appendix A Section 3.3 for the specific k-point grid used for each structure. Γ-

point harmonic vibrational frequencies were computed using finite difference with Phonopy v1.11.2.[204] Predicted room-temperature molar volumes are compared against the available experimental volumes from the CSD (Appendix A Section 4).

The structures for the fixed-cell optimizations from Ref 12 were optimized using the PBE density functional,[3] the D2 dispersion correction,[205] and ultrasoft pseudopotentials. See Ref 12 for details. Although the specific modeling parameters used there differ from those here, test calculations suggest that re-optimizing the structures with the B86bPBE-XDM protocol used here would impact the chemical shifts minimally. Constraining the optimization to retain the experimental lattice parameters reduces the impact of the optimization protocol on the resulting structure. See Appendix A Section 2.2 for details.

### 3.3.3 Chemical shielding calculations

Fragment-based NMR chemical shift calculations were carried out using the Hybrid Many-Body Interaction model (HMBI) v2.0 (available on github https://github.com/gberan/HMBI).[94–96] Chemical shielding calculations for individual fragments were carried out using Gaussian 09[206] with the PBE0 density functional.[207] This functional was chosen based on its excellent performance in earlier benchmarks.[12,104]

The chemical shielding calculations were carried out using the gauge-including atomic orbital (GIAO) approximation[208] and a locally-dense basis set.[209,210] All calculations employed a 6-311+G(2d,p) basis for atoms in the asymmetric unit, 6-311G(d,p) basis for neighboring atoms out to 4 Å, and a 6-31G basis for all atoms beyond 4 Å.[211–216] Previous work showed this mixed basis approach provides a noticeable speed-up on the computational cost with a minimal loss in chemical shift accuracy.[143] As described in previous work, a large

DFT integration grid consisting of 150 radial and 974 Lebedev angular points was used to approach rotational invariance and mitigate numerical noise in the fragment calculations.[143]

The CHELPG atom-centered charges[148] used to construct the electrostatic embedding environment were computed using Gaussian 09.[206] Atom-centered charges were computed using the same PBE0 functional a 6-311+G(2dp,p) basis set. Point charges were placed on all molecules within 30 Å of any atom in the asymmetric unit, and these were surrounded by an additional set of point charges designed to mimic the Madelung potential of the full crystal at the nuclear centers in the asymmetric unit, as described in Ref 12.

### 3.3.4 Chemical shift referencing

The chemical shifts reported here are referenced relative to neat TMS under magic angle spinning (MAS) conditions for $^{13}$C and solid $NH_4Cl$ under MAS for $^{15}$N. A linear regression scheme is used to map between the computed absolute chemical shieldings $\sigma_i$ and the experimentally observed chemical shifts $\delta_i$,

$$\delta_i = A\sigma_i + B \tag{3.9}$$

The regression parameters $A$ and $B$ were taken from an earlier study[12] which employed fixed, experimental lattice parameter optimizations on a set that includes all twenty crystals in the benchmark here plus many additional structures. For $^{13}$C, the regression line is $\delta_i = -0.9658\sigma_i + 179.48$, while for $^{15}$N, it is $\delta_i = -1.0106\sigma_i + 197.46$. Note that these same regression parameters determined using the experimental unit cells were employed for the No QHA, 0 K, and 300 K QHA structures.

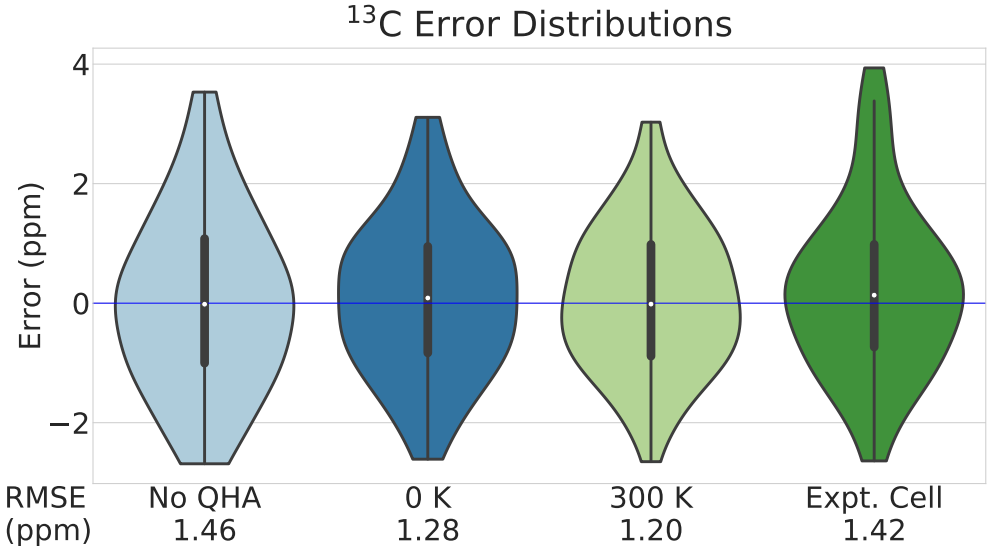



Figure 3.3: Distribution of errors relative to experiment over 68 $^{13}C$ chemical shifts for 12 molecular crystals. Overall RMS errors are displayed below each violin plot.

## 3.4 Results and Discussion

Section 3.4.1 examines how structure relaxation and subsequent thermal expansion impact the chemical shifts in the test set as a whole. Section 3.4.2 examines the performance of the QHA for reproducing experimental room temperature crystal volumes and how that relates to the accuracy of the chemical shifts. Section 3.4.3 considers the impact of thermal expansion on chemical shifts associated with different functional groups. Section 3.4.4 examines the impact of including thermal expansion in five pharmaceutical crystals: acetaminophen, ibuprofen, theophylline, carbamazepine, and naproxen, and Section 3.4.5 investigates how thermal expansion alters the ability to discriminate among several crystral structure prediction candidates for theophylline.

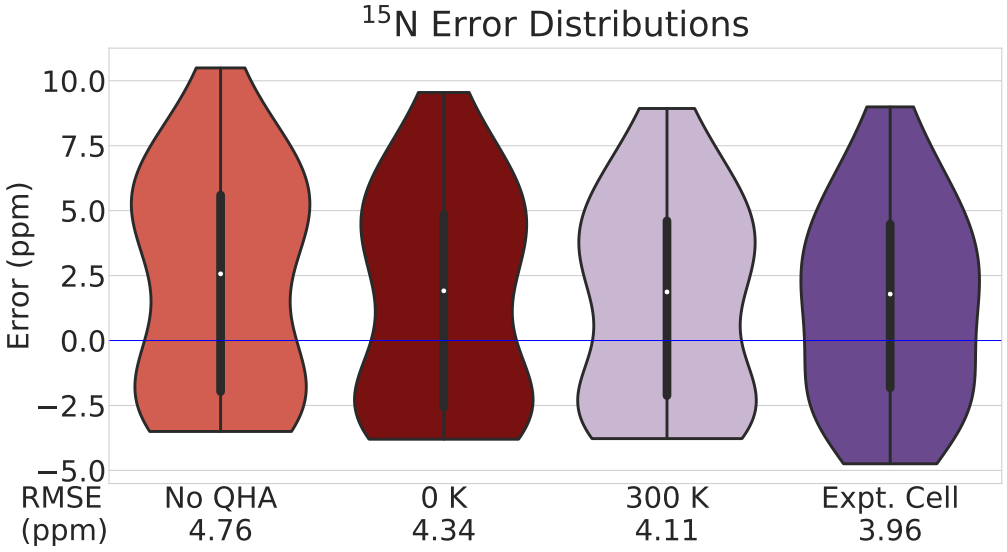



Figure 3.4: Distribution of errors relative to experiment over 28 $^{15}N$ chemical shifts for 14 molecular crystals. Overall RMS errors are displayed below each violin plot.

### 3.4.1 Accuracy of predicted chemical shifts

To begin, we first compare the distributions of errors in the predicted chemical shifts relative to experiment for the fixed-cell optimized (Expt Cell) and fully optimized (No QHA) geometries shown in Figures 3.3 and 3.4. For the $^{13}$C chemical shifts, using the experimental cell gives an overall root-mean-square (rms) error of 1.42 ppm. The error distribution exhibits a single maximum near zero error. For $^{15}$N, the rms error is larger at 4.0 ppm, and the error distribution is slightly bimodal. This bimodal shape is retained across all four sets of structural models and probably reflects the smaller number of data points in the nitrogen set (28 for nitrogen versus 68 for carbon). The larger errors obtained for nitrogen shifts than carbon ones are consistent with the larger $^{15}$N chemical shift range, its sensitivity to electrostatic environment, and prior chemical shift benchmarks.[104,217,218]

Fully relaxing the unit cells (i.e. without constraining the experimental lattice parameters) leads to a 3.9% volume decrease on average, with several crystals decreasing by 6–8% (Tables 3.1 and 3.2). Surprisingly, this full relaxation only increases the $^{13}$C isotropic shift error versus experiment trivially, from 1.42 to 1.46 ppm. However, this small change in the statistical error relative to experiment belies more significant changes to the individual $^{13}$C shifts. The average individual $^{13}$C chemical shift changes by half a ppm upon cell relaxation, and the largest change is 2.4 ppm. In other words, the statistically similar $^{13}$C rms chemical shift errors for the fixed cell and fully optimized structures is fortuitous. For $^{15}$N, full relaxation of the cell increases the errors more noticeably, from 4.0 to 4.8 ppm. The average individual $^{15}$N shift change is 1.1 ppm, with a maximum change of 3.7 ppm.

Accounting for the expansion arising from zero-point vibrational energy (ZPVE) contributions (0 K) and thermal expansion (300 K) via the QHA expands the crystals appreciably. It also has noticeable impacts on the chemical shifts. By 300 K, the average $^{13}$C chemical shift has changed by half a ppm, and chemical shifts in the upper quartile have changed by 0.7–1.2 ppm. The average $^{15}$N chemical shift has changed by 1.0 ppm, and those in the upper quartile have changed by 1.5–3.3 ppm. Notably, accounting for thermal expansion eliminates some of the largest error shifts in Figures 3.3 and 3.4. For example, whereas 18% of the fully relaxed No QHA $^{13}$C shifts exhibit errors greater than 2 ppm, only 10% are greater than 2 ppm in the 300 K QHA structures, and the maximum error drops from 3.5 ppm to 3.0 ppm. For $^{15}$N, the number of shift errors larger than 5 ppm decreases from 32% in the fully relaxed structures to 18% in the 300 K QHA ones. These improvements manifest in the smaller rms errors of 1.2 ppm for $^{13}$C and 4.1 ppm for $^{15}$N at 300 K, which

Table 3.1: For each crystal, the percent volume change and rmsd15 relative to the experimental structure are reported, along with the RMS error in predicted $^{13}C$ isotropic chemical shifts. These are reported for the fully optimized crystal (No QHA), the 0 K QHA structure, the 300 K QHA structure, and the structure optimized with the fixed experimental cell (Expt. Cell). For the Expt. Cell there is no change in volume, hence the percent volume change is not reported.

| Species | No QHA | | | 0 K QHA | | | 300 K QHA | | | Expt. Cell | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | % Volume change | rmsd15 (Å) | RMSE (ppm) | % Volume change | rmsd15 (Å) | RMSE (ppm) | % Volume change | rmsd15 (Å) | RMSE (ppm) | rmsd15 (Å) | RMSE (ppm) |
| HXACAN13 | -6.27 | 0.211 | 1.50 | -2.82 | 0.143 | 1.31 | 0.34 | 0.123 | 1.21 | 0.055 | 1.08 |
| GLYCIN03 | -1.83 | 0.069 | 1.69 | 1.32 | 0.086 | 1.43 | 3.74 | 0.124 | 1.30 | 0.067 | 1.98 |
| LALNIN12 | -2.63 | 0.107 | 2.04 | 0.43 | 0.116 | 1.90 | 2.02 | 0.153 | 1.91 | 0.027 | 2.38 |
| LSERIN01 | -2.31 | 0.106 | 1.37 | 0.24 | 0.116 | 1.15 | 1.74 | 0.133 | 1.07 | 0.063 | 1.77 |
| LTYROS11 | -3.12 | 0.127 | 1.51 | -0.18 | 0.058 | 1.25 | 1.43 | 0.061 | 1.15 | 0.024 | 1.38 |
| SUCROS04 | -3.24 | 0.154 | 1.36 | -0.45 | 0.152 | 1.07 | 0.95 | 0.170 | 0.97 | 0.079 | 0.86 |
| NAPHTA36 | -8.13 | 0.171 | 1.16 | -3.89 | 0.089 | 1.02 | 1.21 | 0.078 | 0.95 | 0.007 | 0.82 |
| ADENOS12 | -5.23 | 0.134 | 1.35 | -1.65 | 0.079 | 1.28 | 0.10 | 0.078 | 1.19 | 0.049 | 1.29 |
| LTHREO01 | -2.94 | 0.090 | 1.25 | -0.38 | 0.092 | 1.03 | 1.27 | 0.121 | 0.95 | 0.068 | 1.31 |
| GLUTAM01 | -2.52 | 0.103 | 1.63 | 0.06 | 0.084 | 1.56 | 1.44 | 0.090 | 1.54 | 0.068 | 1.83 |
| ASPARM03 | -2.82 | 0.120 | 1.66 | -0.22 | 0.114 | 1.55 | 1.30 | 0.123 | 1.52 | 0.030 | 1.57 |
| LCYSTN21 | -2.50 | 0.151 | 1.27 | -0.06 | 0.130 | 0.96 | 1.06 | 0.123 | 0.85 | 0.106 | 2.46 |

Table 3.2: For each crystal, the percent volume change and rmsd15 relative to the experimental structure are reported, along with the RMS error in predicted $^{15}$N isotropic chemical shifts.

| Species | No QHA | | | 0 K QHA | | | 300 K QHA | | | Expt. | Cell |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | % Volume change | rmsd15 (Å) | RMSE (ppm) | % Volume change | rmsd15 (Å) | RMSE (ppm) | % Volume change | rmsd15 (Å) | RMSE (ppm) | rmsd15 (Å) | RMSE (ppm) |
| LSERIN01 | -2.31 | 0.106 | 2.21 | 0.24 | 0.116 | 2.86 | 1.74 | 0.133 | 3.12 | 0.063 | 2.08 |
| GLYCIN03 | -1.83 | 0.069 | 2.21 | 1.32 | 0.086 | 3.07 | 3.74 | 0.124 | 3.56 | 0.067 | 1.49 |
| FUSVAQ01 | -6.74 | 0.148 | 3.55 | -4.13 | 0.118 | 2.93 | -2.32 | 0.092 | 2.52 | 0.048 | 2.08 |
| URACIL | -3.74 | 0.080 | 7.53 | -0.28 | 0.069 | 6.85 | 2.08 | 0.102 | 6.41 | 0.030 | 6.51 |
| BAPLOT01 | -4.66 | 0.193 | 5.83 | -1.50 | 0.192 | 5.29 | 1.72 | 0.212 | 5.05 | 0.131 | 4.66 |
| HXACAN13 | -6.27 | 0.211 | 7.04 | -2.82 | 0.143 | 7.52 | 0.34 | 0.123 | 7.31 | 0.055 | 7.07 |
| LHISTD02 | -3.32 | 0.102 | 5.03 | -0.28 | 0.063 | 4.46 | 1.26 | 0.079 | 4.32 | 0.052 | 4.10 |
| BITZAF | -4.04 | 0.191 | 4.48 | -0.94 | 0.208 | 3.80 | 2.25 | 0.274 | 3.20 | 0.142 | 4.74 |
| CYTSIN | -3.77 | 0.091 | 3.86 | -0.41 | 0.094 | 3.83 | 1.45 | 0.093 | 3.83 | 0.061 | 3.45 |
| LCYSTN21 | -2.50 | 0.151 | 1.88 | -0.06 | 0.130 | 2.75 | 1.06 | 0.123 | 3.13 | 0.106 | 4.56 |
| GLUTAM01 | -2.52 | 0.103 | 6.00 | 0.06 | 0.084 | 4.50 | 1.44 | 0.090 | 3.81 | 0.068 | 3.41 |
| THYMIN01 | -6.33 | 0.235 | 3.91 | -2.29 | 0.166 | 3.55 | 4.13 | 0.153 | 3.05 | 0.056 | 3.43 |
| ASPARM03 | -2.82 | 0.120 | 3.86 | -0.22 | 0.114 | 3.16 | 1.30 | 0.123 | 2.92 | 0.030 | 3.15 |
| ALUCAL04 | -3.08 | 0.110 | 3.47 | 0.08 | 0.113 | 3.73 | 1.95 | 0.157 | 3.77 | 0.028 | 3.11 |

are competitive with or better than the errors obtained with the experimental unit cells. Even just accounting for the ZPVE expansion captures a sizable fraction of the impact of the overall chemical shift correction due to expansion, which is consistent with earlier QHA studies[106,107] that found that ZPVE contributes a significant fraction of the overall expansion that separates the fully relaxed structures with no vibrational contributions and those modeled at room temperature via the QHA.

The fact that the 300 K QHA $^{13}$C rms error of 1.2 ppm is smaller than the fixed experimental cell error of 1.4 ppm deserves further examination. The largest improvement in the predicted $^{13}$C chemical shifts comes from L-cysteine. The experimental LCYSTN21 crystal structure used omits a hydrogen atom on the thiol group. In our previous work, this missing hydrogen atom was placed manually, and the structure was relaxed under fixed experimental cell conditions. However, upon full cell relaxation here, the hydrogen bonding network reconfigures to a more optimal arrangement. Apparently the combination of poor manual placement and constraining the unit cell parameters in the earlier relaxation prevented the necessary relaxation. This new arrangement is maintained in the 300 K QHA structures, and this reduces the error on the cysteine $\alpha$ and side-chain carbons by 2.5 and 1.5 ppm, respectively.

As described in Section 3.3.4, the regression parameters used here to map from chemical shielding to chemical shift are those determined from a larger set of fixed-cell $^{13}$C and $^{15}$N benchmarks used in an earlier study.[12] One might wonder how different the results would be if the regressions were determined directly by fitting the data for these specific systems against experiment separately for each structure set (No QHA, 0 K, 300 K and

experimental cell). Fitting the regressions to fixed-cell structures could conceivably lead to beneficial error cancellation in the fixed cell structures here that is lost in the fully relaxed and QHA structure sets. In practice, however, the changes in the $^{13}$C rms errors upon re-fitting the regression parameters specifically for each set are a negligible 0.01 ppm. For nitrogen, refitting the regression parameters would decrease the chemical shift errors by a few tenths of a ppm, but it would not significantly affect the general trends. In that scenario, the rms errors become 4.3, 3.9, and 3.7 ppm for the No QHA, 0 K, and 300 K $^{15}$N sets, respectively, versus 3.6 ppm for the experimental cell data.

### 3.4.2 Assessment of QHA thermal expansion

Having seen that accounting for thermal expansion improves the quality of the predicted chemical shifts relative to those obtained from fully optimized crystal structures, we now investigate the impact of expansion in more detail. In particular, how well do the room-temperature QHA structures reproduce the experimental ones, and to what extent are the error reductions correlated with the magnitude of the volume correction between the No QHA and 300 K structures? Table 3.1 lists the percent volume changes relative to the experimental structure, the rmsd15 metric, and the rms error in the chemical shifts relative to experiment for that structure. The rmsd15 metric measures the root-mean square deviations in the non-hydrogen atom positions between the optimized and experimental structure using a 15-molecule cluster.[5] Plots visualizing the QHA volume changes relative to the experimental volume are provided in Appendix A Section 5.

As noted earlier, the full No QHA relaxations reduce the molar volumes of the 20 molecular crystals tested here by an average of 3.9%. The unit cell volumes for five

crystals, naphthalene, acetaminophen, thymine, adenine trihydrate, and adenosine, contract by 5% or more. The relatively weak, non-specific van der Waals interactions between naphthalene molecules lead to a large 8.1% volume change between the fully optimized and experimental naphthalene crystal structures. Acetaminophen (6.3% change) and thymine (6.3%) form layered structures with low-dimensional hydrogen bond networks that allow facile expansion between layers. Adenosine exhibits a three-dimensional hydrogen bonding network, but the highly flexible intramolecular degrees of freedom may allow it to change volume more readily (5.2%), while the aforementioned issues with the room-temperature adenine trihydrate crystal structure creates the appearance of a large contraction upon relaxation.

Employing the QHA approximation allows the fully optimized crystals to expand back toward the room temperature structures. At 0 K, the predicted crystal volumes are on average 0.7% smaller than the room-temperature experimental ones (excluding adenine trihydrate). By 300 K, the predicted volumes have expanded too much, exceeding the experimental volumes by 1.6% on average. For adenine trihydrate, the QHA structure with the energetically preferred hydrogen bonding arrangement remains 2.3% smaller than the experimental FUSVAQ structure at room temperature. The outlier nature of FUSVAQ suggests that it might be worthwhile to redetermine its room-temperature crystal structure experimentally. The tendency for the B86bPBE-XDM quasi-harmonic protocol employed here to over-estimate the molar volume is consistent with our earlier work on several small-molecule crystals.[107] That study also found that further refining the crystal structures with higher-level wave function models frequently decreased the molar volumes relative to

the B86bPBE-XDM calculations, bringing them into closer agreement with experiment. However, such wave function-based calculations are somewhat more computationally expensive,[219] so they are not pursued here.

Looking at the rmsd15 values, one unsurprisingly finds excellent agreement between the fixed-cell relaxations and the experimental structures (mean rmsd15 value of 0.06 Å). The rmsd15 values are roughly twice as large for the fully optimized, 0 K QHA, and 300 K QHA structures, with mean values ranging from 0.11-0.13 Å. While these values represent generally good agreement between the predicted and experimental structures, the fact that the rmsd15 values do not improve monotonically along the progression from NoQHA to 0 K and 300 K structures further highlights the systematic over-estimation of the room-temperature volumes by the QHA model. In other words, employing the QHA to refine the fully optimized structures clearly improves the packing densities of the crystal structures relative to experiment. However, the improvement one would expect from heating the QHA model from 0 K to 300 K is not obvious.

Despite the ambiguity about whether the 300 K QHA structures are superior to the 0 K QHA ones in terms of structural metrics, the 300 K structures are clearly superior to the 0 K ones in terms of the accuracy of the predicted chemical shifts, as discussed in Section 3.4.1. Furthermore, the statistical similarity in the accuracy of the chemical shifts predicted for the 300 K QHA and fixed experimental cell structures suggests that, despite differences in the crystal structures, the QHA structures are reproducing the local chemical environments well. Examining the rms errors on a crystal-by-crystal basis in Tables 3.1 and 3.2, one sees that in a number of cases (especially for the $^{13}$C), the shift errors for the

300 K QHA structures are smaller than those for the fixed experimental cells. One might expect to find a direct relationship between good structural agreement and small errors in the predicted chemical shifts, but no such obvious correlation exists.

Several factors might contribute to absence of a clear correlation between the quality of the structure and the quality of the chemical shifts. First, chemical shifts are inherently governed by local electronic structure which is most sensitive to the local atomic environment. Perhaps the gross structural metrics employed here fail to capture more subtle local geometric refinements. Second, the experimental crystal structures may be imperfect. DFT can help refine or correct X-ray diffraction structures,[181,220] for example, and predicted chemical shifts from even high-quality neutron diffraction structures are improved by relaxing the atomic positions beforehand.[221] Third, even the reported experimental crystal unit cell volumes can vary considerably between different diffraction studies. For example, the CSD contains eight different crystal structures for form I acetaminophen at ambient temperature and pressure. These molar volumes for these structures span a range of 3%, with a standard deviation of $\pm$1.2%. Finally, there are of course the errors inherent in density functional theory and the basis sets used here, including perhaps the use of different functionals for the geometry optimization (the GGA B86bPBE-XDM) and chemical shift prediction (the hybrid functional PBE0) steps.

### 3.4.3 Functional group analysis

For further insight into how accounting for thermal expansion improves the predicted chemical shifts, Figures 3.5 and 3.6 plot how the error associated with each predicted chemical shift changes from the fully relaxed structure (arrow tail) to the 300 K QHA

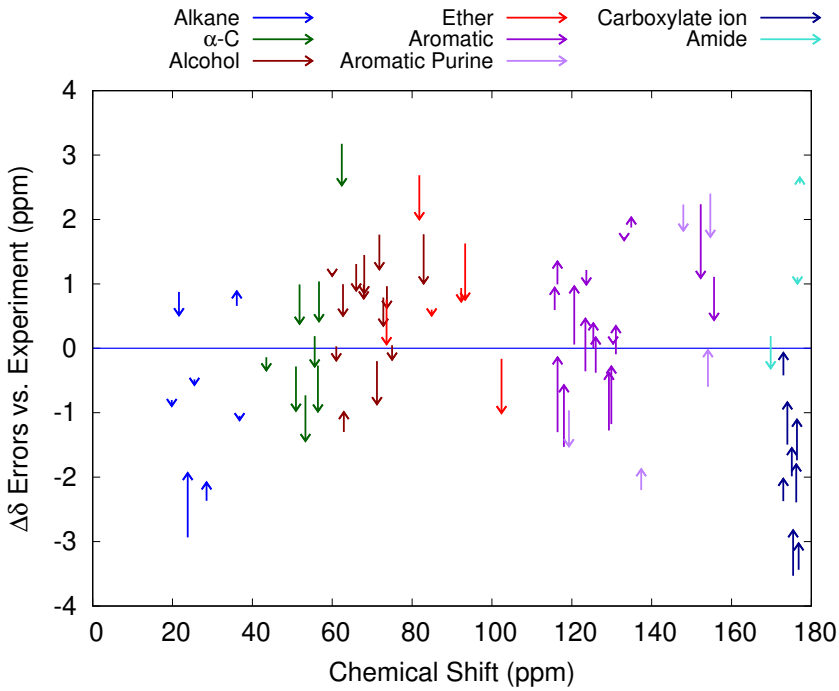



Figure 3.5: Impact of QHA expansion on the errors in individual predicted $^{13}$C chemical shifts relative to experiment. Arrows indicate the change in the error from the fully optimized geometry to the 300 K QHA one. The chemical shifts are color coded by their functional group.

structure (arrow head) for $^{13}$C and $^{15}$N chemical shifts, respectively. The shift changes are color-coded by functional group. See Appendix A Section 3.1 for assignment details. Ideally, accounting for the expansion would shift the errors toward zero.

Consider first the $^{13}$C chemical shifts in Figure 3.5. The magnitude of the average $^{13}$C chemical shift changes by 0.5 ppm between the No QHA and 300 K QHA structures. The smallest changes occur for alkyl and amide carbons (∼0.2-0.3 ppm), while the remaining functional groups exhibit average changes ranging ∼0.4–0.6 ppm, and the largest individual shift changes exceed 1 ppm. As can be seen from Figure 3.5, however, the shift changes do not always correspond to improved agreement with experiment. Defining the (arbitrary)

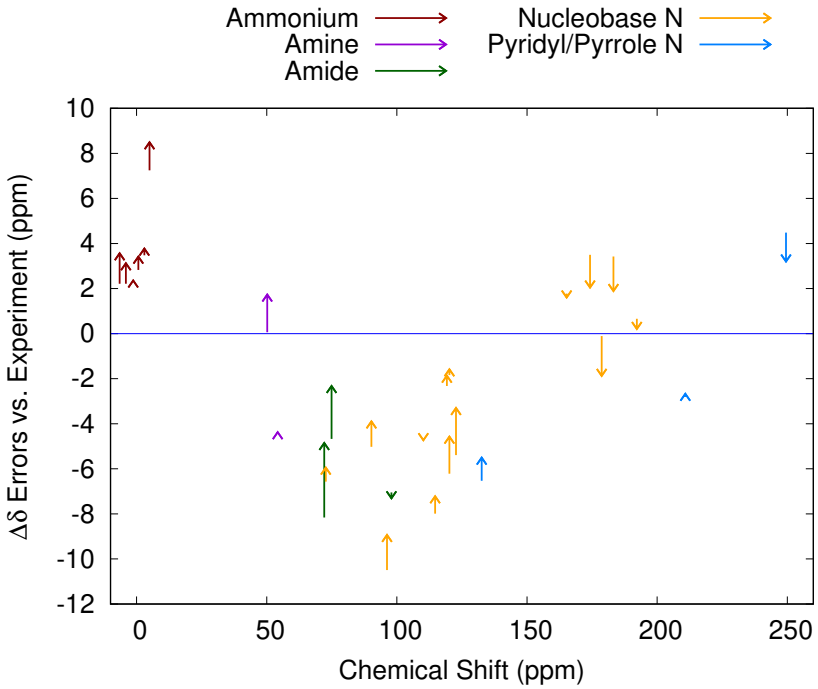



Figure 3.6: Impact of QHA expansion on the errors in individual predicted $^{15}$N chemical shifts relative to experiment. Arrows indicate the change in the error from the fully optimized geometry to the 300 K QHA one. The chemical shifts are color coded by their functional group.

threshold of 0.3 ppm for an appreciable change in the error relative to experiment, one finds that accounting or thermal expansion improves 47% of the $^{13}$C shifts by 0.3 ppm or more relative to experiment, it has little impact on 37% of the shifts (i.e. changes the error by less than 0.3 ppm), and it increases the errors by more than 0.3 ppm for 16% of the shifts.

With the exception of the methyl carbon in acetaminophen, whose shift is improved by 1.0 ppm, the changes in the alkyl shifts upon thermal expansion are small and do not significantly impact the agreement with experiment. These flexible groups are perhaps more impacted by factors such as dynamical motions that are not captured in the simple QHA model here. Thermal expansion corrections to the amide carbons have similarly small

impacts on the agreement with experiment. The test set contains quite a few amino acids, and interestingly, thermal expansion consistently pushes all of the $\alpha$-carbon shifts downfield by an average 0.6 ppm. However, this correction does not always prove beneficial, and the net agreement with experiment becomes negligibly worse (by 0.1 ppm). Mixed behavior is also seen for carbons adjacent to ether oxygens and in aromatic rings, but the net effects are toward improved agreement with experiment. The most interesting impact of thermal expansion occurs for the amino acid carboxylate carbons, which are uniformly improved by half a ppm on average. Experience suggests that the chemical shifts of carboxylate carbons are very sensitive to geometry, and the data here indicates that accounting for thermal expansion proves helpful.

Figure 3.6 plots the shift and error changes upon thermal expansion for the $^{15}$N chemical shifts. As noted earlier, accounting for thermal expansion has an average 1.0 ppm impact on the nitrogen chemical shifts, double what was observed for $^{13}$C. Using the same 0.3 ppm criterion as above, accounting for thermal expansion improves 50% of the 28 nitrogen shifts relative to experiment, has little impact on 25% of the shifts, and makes 25% of the shifts worse.

Interestingly, the shift changes for ammonium nitrogens are relatively modest and below average for the set (average 0.7 ppm), and they systematically shift toward worse agreement with experiment. These functional groups occur in the amino acid crystals and hydrogen bond with the carboxylate groups, whose $^{13}$C shifts were systematically improved via the inclusion of thermal expansion (though they still remain some of the largest errors in the $^{13}$C set). In reality, these functional groups are likely involved in a fast quantum

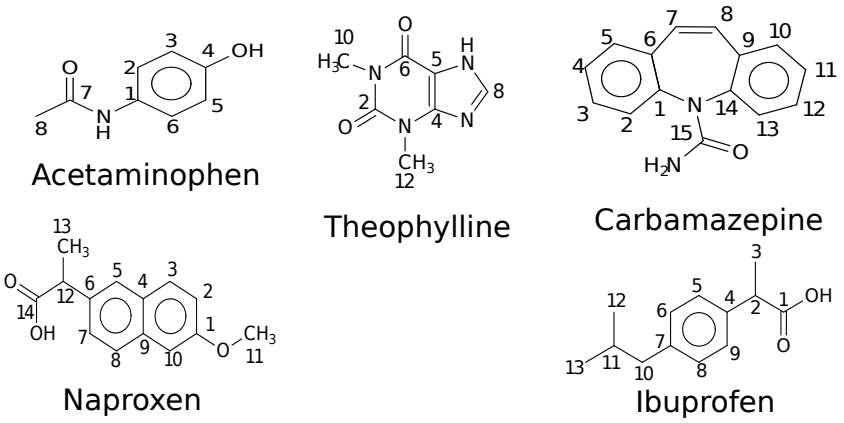



Figure 3.7: The five pharmaceutical species tested here and the carbon atom numbering for each.

mechanical proton exchange[222] between the carboxylate and ammonium groups that is not captured in the models here. It has also been demonstrated that DFT artificially stabilizes crystalline salts due to delocalization error,[223] and that might also lead to problematic treatment of these ammonium-carboxylate interactions.

Aside from the ammonium nitrogens, thermal expansion improves the remaining $^{15}$N shifts by an average 0.8 ppm relative to experiment. Most of the shifts that change appreciably with thermal expansion shift toward better agreement with experiment. The largest gains occur for the terminal amides in L-glutamine and L-asparagine monohydrate near 72–75 ppm. The two most notable exceptions to the general improvement occur for the primary amine in cytosine at 50.2 ppm and the imine nitrogen in theophylline at 178.7 ppm.

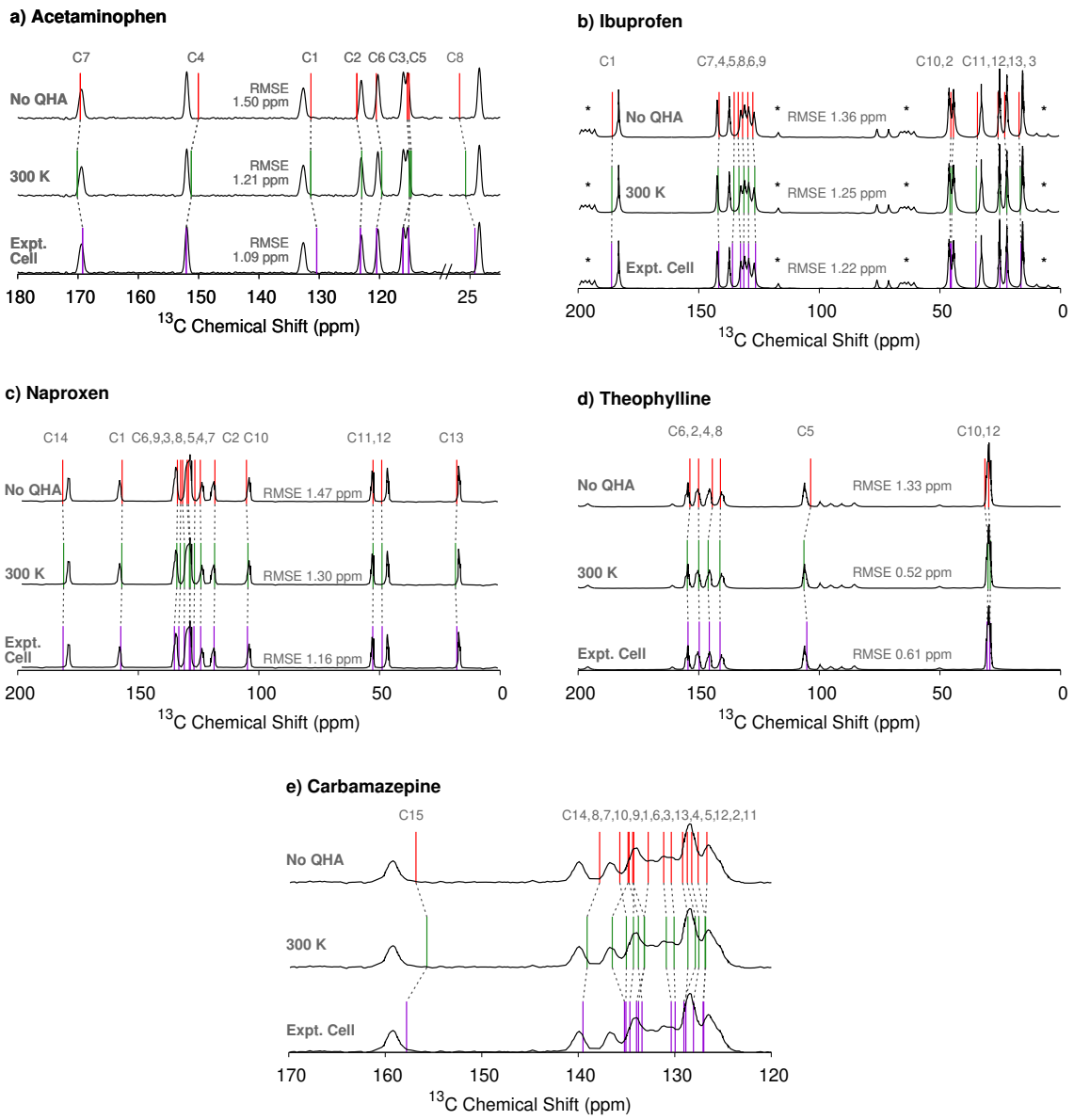



Figure 3.8: Impact of QHA structure refinement on the predicted $^{13}$ isotropic chemical shifts for a) acetaminophen, b) ibuprofen, c) naproxen, d) theophylline, and e) carbamazepine. Spinning side bands in the ibuprofen spectrum are marked with asterisks.

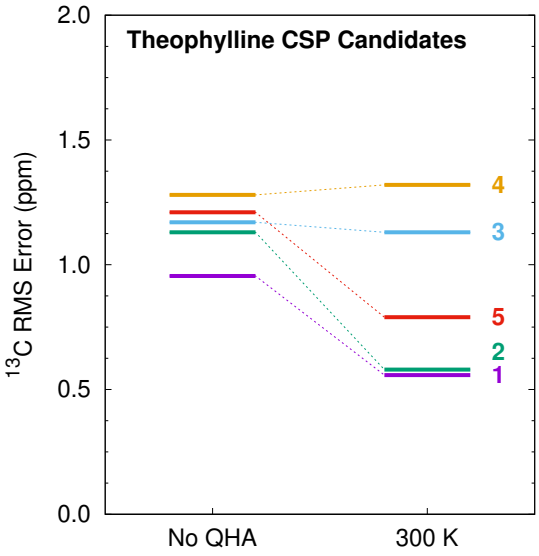



Figure 3.9: Impact of QHA structure refinement on the predicted $^{13}$C isotropic chemical shifts for five candidate structures of theophylline.

### 3.4.4 Refining Chemical Shift Predictions in Pharmaceutical Crystals

Next, we examine the impact of accounting for thermal expansion in predicting the chemical shifts for five pharmaceutical crystals: acetaminophen, ibuprofen, theophylline, carbamazepine, and naproxen (Figure 3.7). As in the previous sections, the crystal structures were fully optimized (No QHA) and then subsequentally expanded to 300 K using the QHA. Fixed-cell geometry optimizations and chemical shift calculations were performed for benchmark comparison. Note that while acetaminophen and theophylline were both included in the results presented above, ibuprofen, carbamazepine, and naproxen were not.

Figure 3.8 compares the predicted $^{13}$C isotropic chemical shifts to the experimental solid-state NMR spectra. See Figure 3.7 for the atom numbering in each species. Volume expansion curves for each structure are available in Appendix A Section 4 and the raw chem-

ical shift data is available in Appendix A Section 3.2. Root-mean-square errors (RMSE) between the predicted and experimental shifts are provided in each case. As was found for the crystals discussed in Section 3.4.2, the RMS error for each crystal generally improves from No QHA to 300 K. Once again, fully optimizing the crystal structure results in less accurate chemical shifts than using the fixed experimental cell structure. However more interesting details can be teased out by examining how each individual predicted chemical shift changes.

Acetaminophen, ibuprofen, and naproxen all show modest statistical improvement going from the fully optimized structure down to the fixed cell. For acetaminophen, all shifts except C3/C5 near 115 ppm improve progressively toward the experimental peaks upon application of the QHA. The fixed cell resolves the problem with the C3/C5, but it predicts a much worse shift for C1. Ibuprofen and naproxen exhibit relatively little temperature dependence in their spectra, though including thermal expansion does slightly improve the level of agreement between theory and experiment. For theophylline, the agreement is already quite good with the No QHA structure. However, performing QHA expansion to 300 K reduces the RMS error over half to only an exceptional 0.5 ppm, with every shift predicted in excellent agreement with experiment.

Finally, carbamazepine represents an interesting case. Most of the chemical shifts shown are concentrated in the aromatic region between ∼125–140 ppm. In that region, accounting for thermal expansion substantially refines and even re-orders many of the chemical shifts. This re-ordering could conceivably change how one might assign the experimental spectrum in this region. Unfortunately, the relatively low resolution of the experimental

spectrum prevents clear assignment of specific peaks or calculation of an rms error. Nevertheless, it is clear visually that thermal expansion noticeably improves agreement with all peaks save the amide group (C15). Indeed, the 300 K QHA $^{13}$C spectrum is arguably in better agreement with experiment than the shifts predicted from the fixed cell structure.

Overall these five pharmaceutical species re-iterate how QHA thermal expansion generally improves the quality of the predicted chemical shifts and brings them to a level of agreement that is competitive with what is obtained from the experimental fixed-cell structures. In cases like theophylline or carbamazepine, the QHA expansion has a large impact on the agreement and even the potential chemical shift assignment. In others such as ibuprofen or naproxen, the impact is quite small. Since the QHA treatment is computationally expensive and it is not currently obvious *a priori* in which cases the effects of thermal expansion on the chemical shifts will be large, it appears that one should likely consider performing QHA refinement of the structures primarily in cases where the agreement with the experimental chemical shifts is unsatisfactory or where one is having difficulty discriminating between structural candidates.

### 3.4.5 Improved resolution of crystal candidates

Finally, an important potential use case for QHA refinement of structures would be in the context of NMR crystallography, where one might try to discriminate among a number of predicted crystal structures. In 2013, Baias et al examined this approach for four different species, including theophylline.[134] While the combination of structure and NMR chemical shift prediction worked well for most of the crystals, neither $^{13}$C nor $^{1}$H chemical shifts provided clear discrimination among the 44 candidate structures of theophylline, with

many different crystal structures exhibiting similar errors in the predicted chemical shifts relative to experiment. Given the significant improvement in the theophylline chemical shifts upon accounting for thermal expansion observed in Section 3.4.4, it is interesting to revisit this discrimination problem here. Specifically, we perform a proof-of-concept investigation of how accounting for QHA thermal expansion impacts the chemical shifts for the five lowest-energy structures from Ref 134.

Table 3.3: Relative electronic and quasi-harmonic Gibbs free energies for the five candidate theophylline crystal structures after B86bPBE-XDM refinement. Energies are relative to Structure 1.

| Candidate Structure | Electronic Energy (kJ/mol) | Relative $\Delta G$(300 K) (kJ/mol) |
|---|---|---|
| 1 | 0.0 | 0.0 |
| 2 | 3.5 | 2.3 |
| 3 | -0.6 | 2.1 |
| 4 | 6.7 | 5.6 |
| 5 | 0.8 | 1.1 |

The structures in Ref 134 were ranked using a hybrid B3LYP/6-31G(d,p) intramolecular energies and a classical intermolecular potential. Relaxing the five lowest-energy structures with periodic B86bPBE-XDM here leads to a considerable energetic reranking (Table 3.3). Structure 1 corresponds to the experimental structure, with an rmsd15 value of 0.212 Å relative to the experimental structure BAPLOT01. At the electronic energy level, structure 3 actually becomes slightly more stable than structure 1 after DFT refinement. However, the QHA Gibbs free energy restores structure 1 as the most stable. Structure 5 is also stabilized appreciably, becoming the second most stable structure, while structures 2 and 3 are nearly degenerate in free energy. Structure 4 lies nearly 6 kJ/mol above structure 1, suggesting it is a less likely candidate.

Next, consider the chemical shifts. Ideally, accounting for thermal expansion would improve the accuracy of the chemical shifts predicted for the correct structure, while increasing the errors for those from the incorrect structures. Figure 3.9 shows how the rms errors in the $^{13}$C chemical shifts change with the inclusion of thermal expansion. Without any thermal expansion, the rms errors on all five structures span a range of only about 0.3 ppm, with structure 1 exhibiting marginally smaller errors than several others. Accounting for thermal expansion reduces the errors appreciably for structures 1, 2, and 5, while it has little impact on those for structures 3 and 4. This more than doubles the spread of rms chemical shift to 0.8 ppm. While the rms error for structure 5 is reduced, the difference between structures 1 and 5 remains about the same.

Interestingly, however, the 300 K QHA models predict nearly identical rms errors for structures 1 and 2. Examination of the crystal structures reveals that both crystal structures exhibit identical stacks of one-dimensional hydrogen bonded chains. The primary difference between the two structures lies in how these adjacent stacks of one-dimensional chains are oriented relative to one another in the third dimension. In structure 1, they are oriented at roughly 45° relative to one another, while in structure 2 they are oriented at 180°. In both cases, the inter-stack distances exceed 3 Å, suggesting a relatively weak interaction between the adjacent stacks. In other words, the difficulty in discriminating those two structures based on $^{13}$C reflects the high similarity in crystal packing motifs and local chemical environments of these two structures. Overall, although identifying the correct theophylline crystal structure based on $^{13}$C chemical shifts alone remains difficult,

accounting for thermal expansion does generally increase the differentiation among these structures.

## 3.5 Conclusions

This work has demonstrated how the quasi-harmonic approximation can be used to refine isotropic NMR chemical shift predictions for molecular crystals. Employing the QHA recovers most or all of the chemical shift error accrued by neglecting the thermal expansion of the crystal. Indeed, the accuracy of chemical shifts computed using quasi-harmonic room-temperature structures rivals what is obtained for shifts computed with the experimental unit cell parameters. In select cases, using the QHA to thermally expand a fully relaxed structure led to slightly better fidelity with experimental chemical shifts than a fixed-cell simulation.

The impact of QHA thermal expansion on chemical shifts can be quite variable, however. No clear relationship between the quality of the structural agreement and the chemical shift errors was found, nor does the magnitude of the change of the chemical shifts upon thermal expansion seem to correlate strongly with the amount of thermal expansion that occurs. We did find, however, that certain functional groups appeared to benefit more from the QHA treatment. In particular, functional groups which exhibit dynamic flexibility such as methyl or alkyl groups tended to benefit less from the QHA treatment than those associated with aromatic or other more rigid functionalities.

Although the QHA expansion calculations are considerably more expensive than traditional optimization and chemical shift prediction approaches, they are feasible for

small-molecule pharmaceutical crystals and other chemically interesting species. The five pharmaceutical examples considered here provide a sampling of the diverse behaviors that can be observed with QHA refinement. In cases like ibuprofen or naproxen, thermal expansion has very little effect on the predicted chemical shifts, while in others like theophylline and carbamazepine, the QHA expansion significantly improves the accuracy of the shifts and can even alter how one might assign them. Finally, it was also demonstrated how employing QHA thermal expansion could help increase the discrimination between candidate structures in a proof-of-concept examination of several theophylline structures that were generated via crystal structure prediction.

Given the modest statistical improvements to the chemical shifts resulting from the QHA thermal expansion models and the higher computational costs compared to conventional approaches, it may not worthwhile performing QHA expansion prior to predicting the chemical shifts on a routine basis. If the experimental lattice parameters are known, it is generally preferable to use that information in the chemical shift prediction. However, in situations where the experimental lattice parameters are unknown and better agreement with experiment is required, as in the context of crystal structure prediction, QHA refinement may prove helpful.

# Chapter 4

# Identifying pragmatic quasi-harmonic electronic structure approaches for modeling molecular crystal thermal expansion

Chapter 3 demonstrated how the quasi-harmonic approximation (QHA) can provide an economical route to modeling the temperature dependence of molecular crystal structures and properties. However, for the majority of the crystals modeled in Chapter 3, density functional theory (DFT) over-expanded the molecular crystal volume. Ideally one would like to use a higher-level electronic structure theory such as correlated wavefunction methods as several studies have demonstrated good performance of these models, at least for rigid molecules, when using fragment-based approaches. However due to the high com-

putational cost of computing harmonic phonons the quasi-harmonic approximation quickly becomes infeasible at this level of theory. In this chapter, we introduce a hierarchy of models in which the energies, geometries, and phonons are computed either with correlated methods or DFT are examined to identify which combinations produce useful predictions for properties such as the molar volume, enthalpy, and entropy as a function of temperature. The work presented in this chapter was originally published in Ref 107.

## 4.1 Introduction

Molecular crystals occur in many areas of chemistry, including pharmaceuticals, organic semi-conductor materials, foods, and explosives. Crystal structure and polymorphism, or the tendency for molecules to adopt multiple distinct crystalline packing motifs, can have substantial impacts on macroscopic properties of these materials. The difficulty in identifying or engineering crystal forms experimentally has generated considerable interest in predicting crystal structures and properties from first principles.[47,224,225] Substantial progress has been made toward successful crystal structure prediction, as evidenced by the successes in the last few Blind Tests of Crystal Structure Prediction.[52,53,162]

Most molecular crystal modeling focuses on the 0 K electronic (internal) energy rather than the finite-temperature free energy.[226] However, temperature can play an important role in many molecular crystal properties. Crystals typically undergo thermal expansion, often at a rate of 0.8–2.5% expansion per 100 K.[63] Even larger expansion occurs in weakly bound crystals. As the crystal expands, its mechanical properties change. Bulk

moduli often decrease by 50% and shear moduli by 30–50% at temperatures approaching the melting point.[63]

The Gibbs free energy also varies with temperature, with room-temperature values typically differing by 1–4 kJ/mol from the 0 K electronic energy.[63,106] Thermal expansion typically softens the lattice vibrations, resulting in a larger Helmholtz vibrational entropy, while simultaneously decreasing the lattice energy. Early investigations suggested that entropic effects were seldom large enough to reverse polymorph stability.[227] Though a significant fraction of the finite-temperature contribution cancels when considering polymorph energy differences, a more recent survey noted that at least 20% of polymorphs pairs are enantiotropically related and reverse their relative thermodynamic stabilities upon heating.[63] Many other properties, including vibrational[129] and nuclear magnetic resonance spectroscopic properties vary with the unit cell size, as well.

Various strategies exist for incorporating finite temperature contributions into the model. Molecular dynamics techniques provide a natural means of treating these effects.[116–125] However, obtaining force fields capable of describing the subtle energetic balances that occur in molecular crystal problems is not always easy. Alternatively, one can employ statistical thermodynamic harmonic or quasi-harmonic corrections to static models.[129] Harmonic phonon contributions are increasingly being employed in crystal structure prediction,[162] and they capture a significant fraction of the vibrational contribution to the free energy.[62,63] Harmonic vibrational contributions impact the polymorph stability ordering in glycine,[111,112], oxalyl dihydrazide,[102] aspirin,[113] and modern drug targets,[114] for example.

However, to capture thermal expansion effects, one must address how the phonons change with crystal volume. The quasi-harmonic approximation provides a straightforward means for doing so and for therefore predicting how the Gibbs free energy depends on temperature and pressure. The quasi-harmonic approximation has been used long been used in force field studies.[182,228–232] Recent DFT work has examined how free energy impacts structure rankings in crystal structure prediction.[130,233] It has sought to reproduce structural and thermochemical properties of crystalline ammonia,[129] urea,[234] and carbamazepine (using density functional tight binding theory).[235]

These DFT-based quasi-harmonic studies have been successful, and wavefunction-based techniques are often too computationally demanding for quasi-harmonic crystal studies. However, fragment approaches such as the hybrid many-body interaction (HMBI) model, enable application of high-level, correlated wavefunction techniques that would otherwise be computationally prohibitive to molecular crystal problems. HMBI combines a quantum mechanical (QM) treatment of the unit cell monomers and their short-range dimer interactions with a classical polarizable force field treatment of longer-range and many-body effects.[97,101,103,236] We have demonstrated that the quasi-harmonic HMBI[237] with large-basis second-order Møller-Plesset perturbation theory (MP2) and/or coupled clusters singles and doubles with perturbative triples (CCSD(T)) allows prediction of small-molecule structures and properties in excellent agreement with experiment. For example, it predicts the thermal expansion of carbon dioxide phase I to within 2–3%, the sublimation enthalpy to within 1–2 kJ/mol, and the sublimation temperature to within 3 K.[105,106] By comparing and contrasting the excellent agreement between theory and experimental structures and

Raman spectra for several different crystalline phases of carbon dioxide, we argued that the long-accepted structure for phase III carbon dioxide is incorrect, and that phases III and VII are actually identical.[26] Furthermore, these same techniques enabled successful prediction of crystalline methanol thermochemistry.[108] More significantly, they allow *ab initio* prediction of the polymorph phase diagram with accuracy corresponding to Gibbs free energy errors of ∼0.5 kJ/mol or less.[109] Other fragment-based studies have similarly employed the quasi-harmonic approximation to study thermochemical properties in a much broader range of crystals.[127,128]

However, even with fragment techniques, the studies employing correlated wavefunction methods are computationally demanding. Predicting the methanol phase diagram entirely with MP2 and CCSD(T) required several hundred thousand central processing unit hours, for example. Performing the same level calculations on a typical pharmaceutical crystal would be impractical. Is there a useful middle ground, which achieves most of the accuracy found above, albeit at much lower computational cost?

A quasi-harmonic calculation involves three main ingredients: energies, crystal geometries, and phonon frequencies. This study investigates how accurately one must compute each of those pieces to predict how crystal structures vary with temperature accurately. Can one, for example, replace a computationally expensive MP2 treatment of the phonon frequencies with a less-demanding one from periodic DFT? What about the geometries? Or, taken from the opposite perspective in which dispersion-corrected DFT models are the baseline: How much do the predicted finite-temperature structures or thermochemistry benefit by investing the additional computational effort to replace some or all of these quasi-

harmonic approximation components with results obtained from correlated wave function methods? Even inexpensive semi-empirical density functional tight binding models reproduce the thermal expansion in carbamazepine fairly well,[235] for example.

Here, to improve our understanding of how the performance of correlated wavefunction and DFT methods compare and the potential benefits of combining them in different ways, we examine the thermal expansion in several small-molecule crystals, systematically replacing the MP2 (or better) treatment of the energies, structures, and/or phonons with a dispersion-corrected DFT one. We find that while portions of the calculation can often be performed with periodic density functional theory, there are clear benefits to including contributions from wavefunction techniques. Based on the results from the small molecule systems, quasi-harmonic calculations are performed on the pharmaceutical acetaminophen (paracetamol) to predict both the thermal expansion and to investigate a discrepancy between two experimentally reported heats of sublimation.

## 4.2 Theory

### 4.2.1 Quasi-Harmonic Approximation

The Gibbs free energy is required to model molecular crystals at finite temperatures and pressures. From statistical thermodynamics, the Gibbs free energy combines the electronic internal energy $U_{el}$, the Helmholtz vibrational free energy $F_{vib}$, and a pressure-volume ($PV$) contribution,

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

In crystals at ambient pressure, the $PV$ term contributes negligibly.

The internal electronic energy $U_{el}$ is computed here via the fragment-based hybrid many-body interaction model (HMBI). HMBI treats individual molecules in the unit cell and their short-range pairwise interactions quantum mechanically (QM), while the longer-range and many-body interactions are typically approximated using a classical molecular mechanics (MM) polarizable force field.

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

In some cases, the MM terms will be computed from periodic Hartree-Fock (HF) instead, in which case HMBI is equivalent to the method of increments.[98–100]

The Helmholtz vibrational free energy is computed from standard harmonic oscillator vibrational partition functions as,

$$F_{vib}(T) = N_a \sum_i \left( \frac{\hbar\omega_i}{2} + k_b T \ln \left[ 1 - \exp\left( -\frac{\hbar\omega_i}{k_b T} \right) \right] \right) \tag{4.3}$$

where $N_a$ is Avogadro's number, $\hbar$ is Plank's constant, $k_b$ is the Boltzmann constant, and $\omega_i$ is the vibrational frequency of mode $i$. The first term corresponds to the zero-point vibrational contribution, while the second gives the thermal vibrational contribution. Here, phonons are evaluated only at the $\Gamma$ point.

To circumvent the high computational cost of computing the phonons repeatedly while minimizing $G(T, P)$ for the given thermodynamic conditions, the vibrational contribution can be approximated via the quasi-harmonic approximation, which estimates how the phonon frequencies and $F_{vib}$ vary with unit cell volume. Mode-specific Grüneisen parameters $\gamma_i$ are employed to estimate how each individual phonon frequency varies with

volume,

$$\gamma_i = -\left(\frac{\partial \ln \omega_i}{\partial \ln V}\right) \tag{4.4}$$

Integrating this equation yields,

$$\omega_i = \omega_i^{ref}\left(\frac{V}{V^{ref}}\right)^{-\gamma_i} \tag{4.5}$$

The reference volumes and phonon frequencies are obtained for the crystal structures which minimize the electronic energy $U_{el}$, and the Grüneisen parameters are obtained via finite difference of the phonon frequencies with respect to changes in unit cell volume about that reference structure.

### 4.2.2 Hierarchy of Approximations

Table 4.1: The four model tiers considered in this work. Tier 1 is the most computationally demanding, and higher tiers subsequently reduce the computational cost by replacing portions of calculations at the "High" HMBI level with faster ones at the "Low" level of theory, DFT. HMBI corresponds to employing MP2 or other correlated wavefunction methods for the monomer and dimer treatment plus AMOEBA or periodic HF many-body contributions.

| Tier | Energies | Structures | Frequencies |
|---|---|---|---|
| 1 | HMBI | HMBI | HMBI |
| 2 | HMBI | HMBI | DFT |
| 3 | HMBI | DFT | DFT |
| 4 | DFT | DFT | DFT |

The key question addressed in this study is how accurately one needs to compute each ingredient in the quasi-harmonic model: the energies, crystal geometries, and the phonon frequencies. Previous work has demonstrated that evaluating all three ingredients with HMBI using large-basis MP2 often performs very well (subject to the well-known limitations of MP2[100]), and even better results are obtained if the single-point energies are

refined with a CCSD(T) treatment of the HMBI monomers and dimers. However, large-basis correlated wave function method calculations can be computationally expensive. Here, a systematic hierarchy of four different tiers of approximation is adopted, as summarized in Table 4.1.

In Tier 1, all calculations are performed at the "High" level of theory—optimizing the structures, computing the phonons, and in some cases refining the single-point energies with correlated wavefunction methods via HMBI. Single-point energy refinement might include using larger basis sets, employing CCSD(T) or MP2C instead of MP2, or calculating the many-body contributions from periodic HF instead of AMOEBA. AMOEBA models electrostatics with atom-centered multipole moments up to quadrupoles, polarization with self-consistently induced dipoles and a Thole damping model, and van der Waals via a buffered 14-7 potential.[238] Although Tier 1 performs well,[26,105,106,108,109] it is computationally demanding and quickly becomes impractical beyond several dozen atoms per unit cell (assuming $P2_1/c$ symmetry or similar).

Tier 2 replaces the single most computationally demanding portion of the calculation, the HMBI phonon frequency evaluation, with phonons at the "Low" level, periodic DFT. Because harmonic frequency calculations require that the system geometry be at a stationary point of the potential energy surface with respect to the atomic positions, Tier 2 structures optimized with HMBI then have their atomic positions relaxed with DFT while the lattice parameters being held fixed at their HMBI values. This ensures real vibrational frequencies while maintaining the (hopefully) more reliable HMBI unit cell. Because

the unit cell is constrained, the DFT relaxation typically alters the atomic positions only slightly.

A similar approach was previously used to predict pressure-dependent Raman spectra for several phases of carbon dioxide in excellent agreement with experiment.[26] In that instance, the unit cell was first predicted at the quasi-harmonic HMBI MP2/CBS limit, and then atomic positions were relaxed and Raman spectrum predicted at the HMBI MP2/aug-cc-pVDZ level to reduce the computational cost of the phonon evaluation. Constraining the cell at the dimensions predicted with the large basis set helped ensure the key lattice phonon modes were reproduced correctly despite the small basis set.

Tier 3 seeks to reduce the computational costs further, performing all structure optimizations and phonon frequency evaluations with periodic DFT. Only single-point energies are computed from wave function methods with HMBI. Depending on the electronic structure method used for the monomers and dimers in the HMBI single-point energies, the computational cost of Tier 3 can be only moderately higher than a pure DFT calculation (Tier 4).

Our recent studies of crystalline methanol,[108,109] which considered a complete basis set (CBS) CCSD(T) HMBI treatment on geometries from either MP2/CBS HMBI (Tier 1) or periodic PBE-D3 calculations (Tier 3). In the initial work,[108] it was unclear how much better Tier 1 was than Tier 3. For example, the sublimation enthalpy was more accurate when computed from MP2 structures, while thermal expansion agreed more closely with experiment when using the DFT structures. Subsequently, we recognized that AMOEBA was overestimating the many-body contribution. Replacing that AMOEBA contribution

with one computed from periodic HF in the final single-point energies significantly improved the quality of the predictions using the MP2 geometries, while those using the PBE-D3 geometries became slightly worse.[109] In the present work, examining several additional crystals will provide further insights into the performance of Tier 3 models.

Finally, Tier 4 performs all calculations with periodic DFT. Many quasi-harmonic studies in the literature already employ this approach. Here, Tier 4 provides a baseline for assessing what, if any, improvement one finds upon incorporating energies, structures, or phonons from correlated wavefunction techniques on top of DFT. Of course, many different density functionals and dispersion corrections exist, and the results will depend somewhat on the particular model choices.

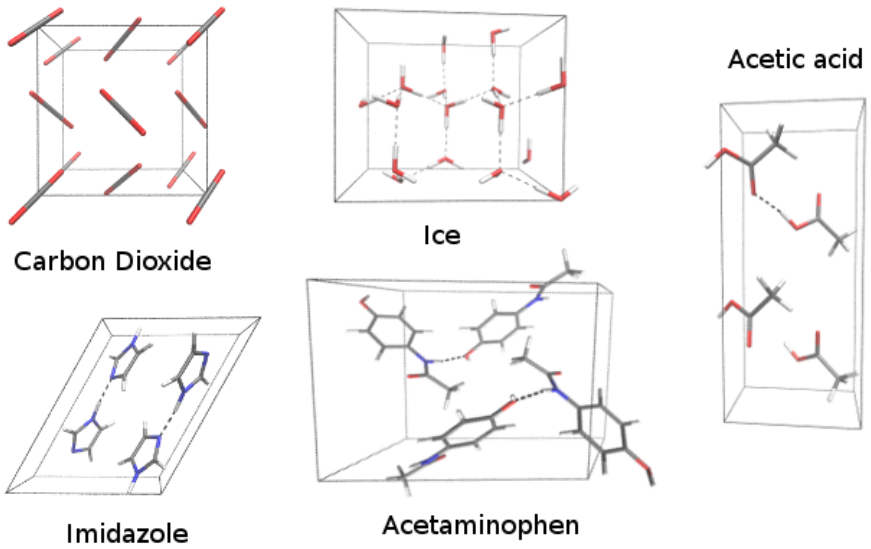



Figure 4.1: The crystals modeled in this study.

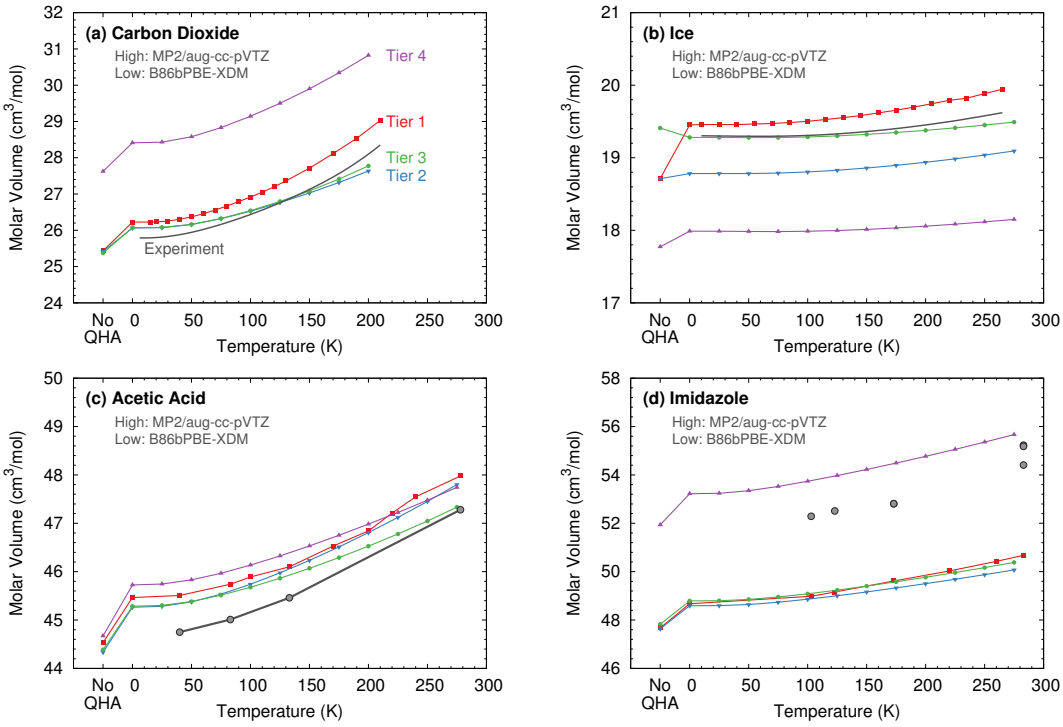



Figure 4.2: Comparison of predicted molar volumes from Tiers 1–4 approximations for (a) carbon dioxide, (b) ice, (c) acetic acid, and (d) imidazole using MP2/aug-cc-pVTZ + AMOEBA for the high-level calculations and B86bPBE-XDM for the low level ones. The No QHA data refers to the electronic energy minimum, with no vibrational contribution.

### 4.2.3 Computational details

To evaluate the performance of the different hybrid tiers described in Section 4.2.2, we return to the four small-molecule crystals whose thermal expansion was previously investigated with Tier 1:[105,106] carbon dioxide phase I, ice Ih, acetic acid (orthorhombic phase), and $\alpha$ imidazole (Figure 4.1). Initial crystal structures were taken from a low-temperature experimental crystal structure of carbon dioxide,[239] a zero net dipole 16-molecule supercell of ice,[240] a 278 K single crystal x-ray diffraction structure of acetic acid (Cambridge Crystal Structure Database (CSD) reference code ACETAC01),[241] and a 103 K neutron scattering structure of imidazole (IMAZOL06).[242] To investigate the performance on larger species,

results are also presented for acetaminophen form I (monoclinic form) starting from a 20 K neutron scattering structure (HXACAN13)[11]. Experimental temperature-dependent volume data was found in the literature for carbon dioxide,[243] ice,[244] acetic acid,[241,245], and imidazole.[242,246,247] Enthalpy and entropy data for carbon dioxide,[248–251] ice,[252–254] acetic acid,[253,255,256] and imidazole[257–259] were taken or derived from experimental data, as described previously.[106]

All Tier 1 results here were taken from our earlier work in Ref 106. Those were obtained by minimizing the Gibbs free energy (Eq 4.1) on the fly at each temperature with quasi-harmonic evaluation of the phonon contribution. In that work, the Grüneisen parameters were computed by isotropically expanding and contracting the lattice by 10 $Å^3$ in order to approximate the derivative in Eq 4.4 via finite difference. Ice and carbon dioxide do in fact expand nearly isotropically. The other two crystals exhibit slightly more anisotropic expansion, so one might prefer an approach that estimates the Grüneisen parameters based on anisotropic volume changes. Nevertheless, agreement between the predicted and experimental volumes are fairly reasonable. Note that the $PV$ term contributes negligibly at ambient pressure and was neglected in all cases. Also, slight irregularities occur in some of the Tier 1 thermal expansion results due to numerical convergence issues on the generally flat energy surfaces.

For the new Tier 2–4 results here, a slightly different approach is taken. First, energy versus volume curves $E(V)$ were mapped out as a function of volume, by minimizing the HMBI or DFT energy of the cell under positive (cell compression) or negative pressure

(cell expansion). The resulting $E(V)$ curve was fitted to the Murnaghan equation of state,

$$E(V) = E_0 + \frac{B_0 V}{B_0'}\left[\frac{(V_0/V)^{B_0'}}{B_0' - 1} + 1\right] - \frac{B_0 V_0}{B_0' - 1} \tag{4.6}$$

where $E_0$, $V_0$, $B_0$, and $B_0'$ are the fit parameters. $E_0$ gives the electronic energy at the minimum, $V_0$ is the molar volume at the minimum energy, $B_0$ is the bulk modulus, and $B_0'$ is the first derivative of the bulk modulus with respect to pressure.

Next, the reference phonon frequencies and Grüneisen parameters were computed. Instead of isotropically scaling the crystal volumes in the finite difference estimate of Eq 4.4, frequencies were computed on the positive and negative pressure structures surrounding the electronic energy minimum. This anisotropic estimate for the Grüneisen parameters should provide a better description for how the phonons vary with volume. Using Eqs 4.3 and 4.5, the Helmholtz vibrational free energy $F_{vib}$ was calculated as a function of volume for various temperatures and cubically splined. Care was taken to ensure that explicit data points extend outside the actual volume range to avoid spline artifacts in the region of interest. Summing the fitted $E(V)$ and splined $F_{vib}$ contributions (neglecting the $PV$ term at ambient pressure) gives the Gibbs free energy as a function of volume and temperature. Minimizing $G(V,T)$ with respect to volume provides the optimal unit cell structure at that temperature. Selected crystalline energy-volume curves, Helmholtz vibrational free energies, and Gibbs free energies are provided in Appendix B Figures 1–4.

Enthalpy and entropy were computed from the same ingredients using the standard statistical mechanical expression. See Ref 105 for details. Similarly, standard ideal gas, rigid rotor, and harmonic oscillator partition function expressions were used to evaluate thermochemical contributions for the gas phase species when computing sublimation

enthalpies and entropies.

For HMBI, density-fitted MP2 were carried out in the Dunning aug-cc-pVXZ basis sets (X = D, T, or Q, and abbreviated aXZ in some places) using Molpro v2012.1.[260] Because MP2 has well-known problems with van der Waals dispersion, we also refined some single-point energies with MP2C,[261,262] also using Molpro. MP2C replaces the uncoupled HF treatment of dispersion found in MP2 with an improved coupled Kohn-Sham treatment, and it performs very well for non-covalent interactions.[263] Because the MP2C dispersion correction is less sensitive to basis set than MP2,[264] the correction was computed in the aug-cc-pVTZ basis set and then combined with MP2 in various basis sets (aug-cc-pVTZ or larger). Extrapolation to the complete basis set limit from the triple and quadruple zeta basis sets was performed in the usual fashion.[265,266] All dimer calculations in HMBI employed a counterpoise correction for basis set superposition error.

Most of the HMBI calculations used the AMOEBA force field for the long-range and many-body contributions. These were evaluated using Tinker 7.1.[267] Existing force field parameters were used for water, acetic acid, and imidazole.[238] Poltype[268] was used to generate parameters for carbon dioxide[105] and acetaminophen. Only the intermolecular force field parameters are needed in the context of HMBI, since intramolecular contributions are treated quantum mechanically.

In select cases, single-point energies were refined with periodic Hartree-Fock and the pob-TZVP basis,[269] which is a variant of def2-TZVP adapted for periodic calculations, and were performed using CRYSTAL09.[270,271] Basis set superposition error is a potential problem in Gaussian basis set calculations on periodic crystals, but addressing it in the

context of the many-body calculations here is challenging. This basis was chosen here because an earlier study found that many-body contributions in small molecular clusters computed in the pob-TZVP basis set without counterpoise correction compared well against those from in large basis sets.[109] A $10^{-7}$ a.u. tolerance was used for the Coulomb overlap threshold, Coulomb penetration threshold, and exchange overlap threshold, and tolerances of $10^{-12}$ and $10^{-30}$ a.u. for the pseudo-overlaps (see Ref 271 for details). Monkhorst-Pack and Gilat shrinking factors of 8 were used for the four smaller crystals, and 4 for acetaminophen.

DFT calculations on carbon dioxide, ice, acetic acid, and imidazole were performed using the periodic boundary planewave/pseudopotential and projector augmented wave (PAW) approaches with an 80 Rydberg energy cutoff and 5×5×5 Monkhorst-Pack k-point grid. For the larger acetaminophen crystal, a lower plane wave energy cutoff of 60 Ry and $1 \times 3 \times 3$ k-point grid were used. DFT energies and gradients were computed with Quantum Espresso v6.1[1,272] and Γ-point frequencies were produced by Phonopy v1.11.2.[204] The BLYP and B86BPBE PAW functionals for H, C, N and O were produced using A. Dal Corso's Atomic code v6.1. Most DFT results reported here were obtained using the B86bPBE functional[2,273] with the Becke-Johnson's exchange-hole dipole method (XDM) dispersion correction,[4] which has performed well in earlier molecular crystal studies.[130,233,274,275] Selected results with other functionals, such as BLYP[276,277] and PBE[273] with the D2[278] or XDM dispersion correction are also presented.

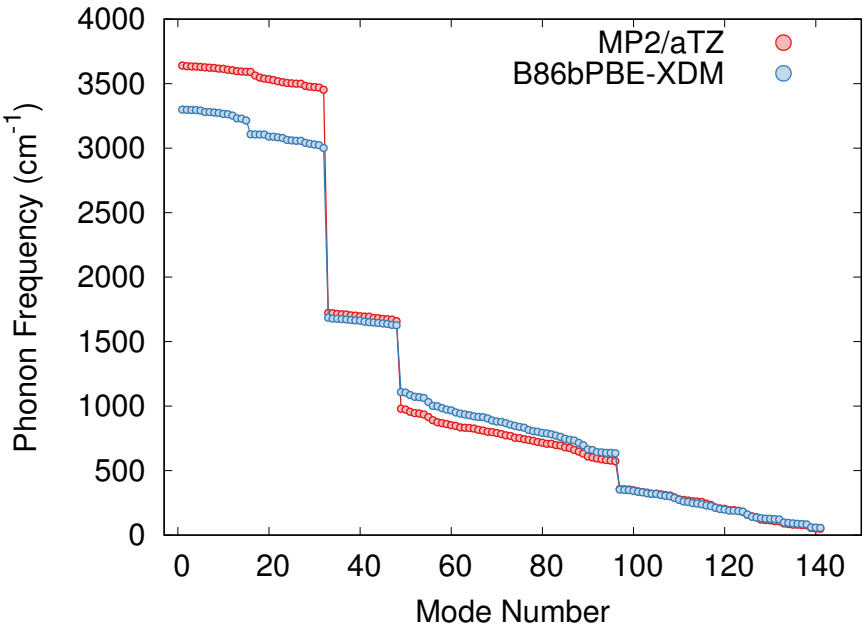



Figure 4.3: Comparison of predicted ice phonon frequencies predicted with MP2/aug-cc-pVTZ + AMOEBA or from periodic B86bPBE-XDM.

## 4.3 Results and Discussion

The work here primarily focuses on phase I carbon dioxide, ice Ih, the orthorhombic phase of acetic acid, and the $\alpha$ form of imidazole. These crystals exhibit diverse non-covalent interactions, ranging from strongly cooperative hydrogen bonding to $\pi$ systems with significant van der Waals dispersion interactions. Ice and acetic acid are dominated by hydrogen bonded interactions. Ice has the familiar hexagonal pattern and each molecule accepting two and donating two hydrogen bonds (Bernal-Fowler rules), while the acetic acid molecules form one-dimensional chains. Carbon dioxide is bound by a mixture of electrostatics and dispersion, and the $\pi$-conjugated imidazole rings form one-dimensional hydrogen-bonded chains with strong dispersion interactions between them.

### 4.3.1 Molar volume and thermal expansion

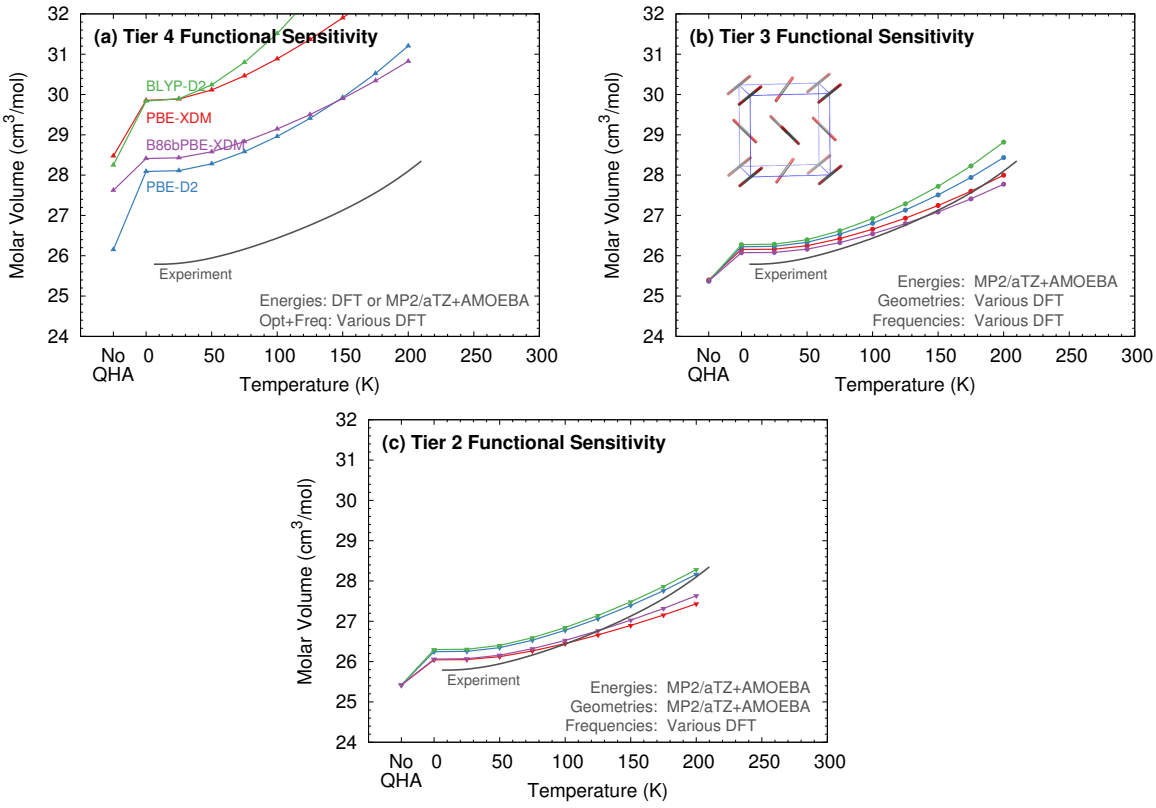



Figure 4.4: (a) While different density functionals predict somewhat different molar volumes for carbon dioxide (Tier 4), the sensitivity of the predicted volumes decreases considerably when (b) MP2 single-points are used on the DFT geometries (Tier 3), and (c) it decreases further when only the DFT phonons are used (Tier 2).

To begin, we examine the performance of Tiers 1–4 on carbon dioxide, ice, acetic acid, and imidazole. We previously examined these crystals with MP2 and various basis sets (Tier 1).[106] With counterpoise corrected dimer energies, smaller basis sets typically underestimate non-covalent attractions. For these crystals modeled with HMBI, this translates to molar volumes being overestimated. In carbon dioxide or ice, for example, systematically increasing the basis set from aug-cc-pVDZ all the way to the complete basis set limit

shrinks the molar volume, with the thermal expansion curves in the different basis sets being relatively parallel to one another.[106]

The present study focuses on Tier 1 results at the MP2/aug-cc-pVTZ level for a couple reasons. First, Tier 1 calculations require the energies, geometries, and phonons all be computed at the MP2 level, which is computationally expensive. The aug-cc-pVTZ basis set provides a practical compromise between basis-set completeness and computational efficiency, making Tier 1 calculations feasible on species like acetic acid and imidazole. Second, as shown in Figure 4.2, this level of theory predicts volumes for carbon dioxide, ice, and acetic acid that are (fortuitously) in reasonably good agreement with experiment. MP2 performs worse for imidazole, since it exaggerates the van der Waals attractions between the molecules, which translates to the underestimated volume seen in Figure 4.2. The Tier 1 results also generally reproduce the experimentally observed rate of thermal expansion.

Tier 2 replaces the MP2 phonons with ones calculated from DFT (B86bPBE-XDM here). The MP2 phonon frequencies tend to differ moderately from the DFT ones, as shown for ice in Figure 4.3. Most notably, MP2 predicts excessively large intramolecular frequencies in the O-H stretching region. In inelastic neutron scattering experiments,[279,280] these modes occur in the range ∼3000–3500 $cm^{-1}$, while MP2 predicts them to be several hundred wave numbers higher in energy.[281] DFT calculations provide a much better description of those high-frequency modes. On the other hand, the slightly lower 500–1100 $cm^{-1}$ range of the MP2 librational modes is in better agreement with experiment than the DFT ones, which are shifted about 100 $cm^{-1}$ higher. Both models predict generally similar frequencies for the HOH bending modes near 1600 $cm^{-1}$ and pseudotranslational models below 400 $cm^{-1}$.

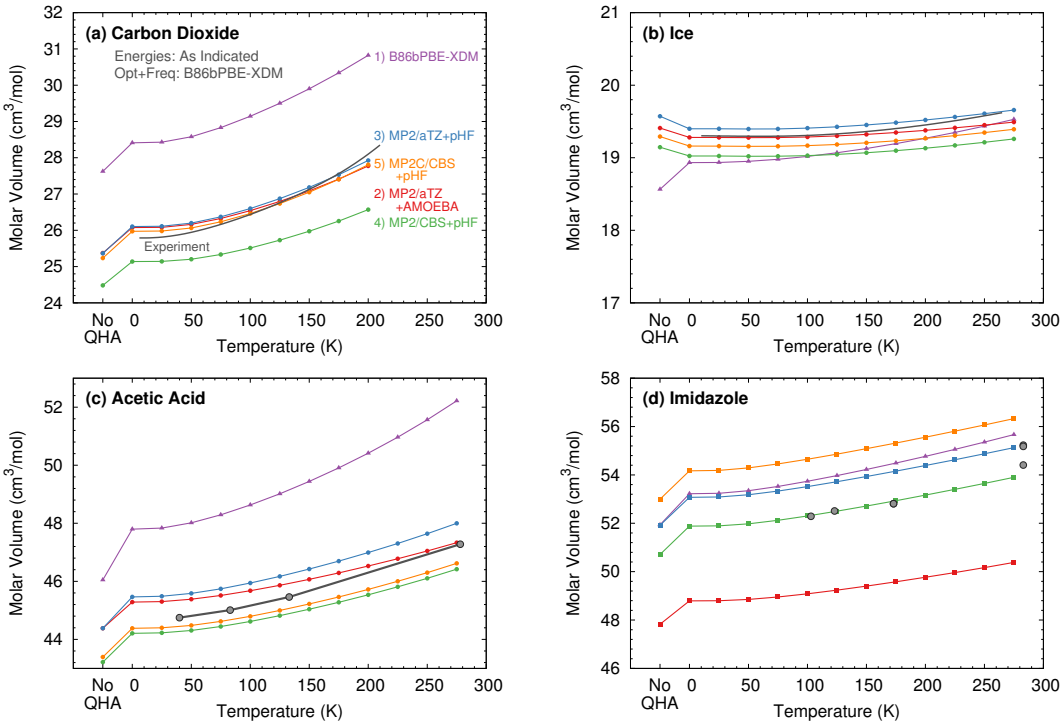



Figure 4.5: Impact of single-point HMBI energy refinement on the DFT predictions for (a) carbon dioxide, (b) ice, (c) acetic acid, and (d) imidazole. Starting with (1) pure B86bPBE (Tier 4), we (2) first refine to Tier 3 with MP2/aug-cc-pVTZ + AMOEBA. Further Tier 3 refinements are made by (3) replacing the AMOEBA many-body treatment with a periodic HF one, (4) extrapolating MP2 to the complete basis set limit, and (5) correcting MP2 with MP2C.

See Refs 281 and 282 for more discussion of MP2 and DFT prediction of inelastic neutron scattering spectra in ice. MP2 and DFT phonon frequencies exhibited similar qualitative behaviors relative to experiment in crystalline methanol.[108] Errors in the high-frequency modes will impact the zero-point vibrational energy contribution, but they have a much smaller effect on thermal expansion. Differences in low frequency modes will have a bigger impact on the thermal expansion.

When applied to quasi-harmonic thermal expansion, replacing the MP2 phonon frequencies with DFT ones in Tier 2 introduces only a modest decrease in the quality of the

predicted molar volumes. In ice, for example, moving from Tier 1 to Tier 2 with B86bPBE-XDM phonons shrinks the molar volume by a few tenths of a $cm^3$/mol. Increasing the MP2 basis set further would lead to further volume underestimation.[106] For carbon dioxide, 0 K volumes are similar with Tier 1 and 2, but the thermal expansion is underestimated at Tier 2.

The performance of Tier 3, which uses geometries and phonons computed with B86bPBE-XDM, is generally very similar to Tier 2. The largest molar volume differences occur for ice, where the two models differ by up to $\sim$1 $cm^3$/mol. For acetic acid, they agree very closely at low temperature, but differ by $\sim$0.5 $cm^3$/mol near room temperature. The differences between Tiers 2 and 3 are even smaller in the other two crystals.

However, Tier 3 ice exhibits the unusual feature that adding the zero-point vibrational energy contribution (i.e. from "No QHA" to 0 K in Figure 4.2b) actually leads to a 0.1 $cm^3$/mol lattice contraction, instead of the typical expansion. This does not occur for Tier 4, which uses the same phonons. As shown in Appendix B Figure 2, the quasi-harmonic B86bPBE-XDM vibrational free energy exhibits a shallow minimum at 18 $cm^3$/mol. For Tier 4 B86bPBE-XDM, the electronic energy minimum occurs at smaller volume, where $F_{vib}$ has negative slope, so adding the zero-point contribution leads to expansion. On the other hand, the Tier 3 MP2 electronic energy minimum occurs above 19 $cm^3$/mol, in the regime where the B86bPBE-XDM vibrational free energy has a positive slope, and adding zero-point vibrational energy drives an initial contraction. At higher temperatures, the $F_{vib}$ minima shift toward larger volumes, allowing expansion to occur in Tier 3. Interestingly, this minimum in $F_{vib}$ does not occur for MP2 phonons (Tier 1). In Tier 2, the minimum

occurs at larger volumes and does not cause the contraction seen in Tier 3. In other words, this behavior appears to be an unfortunate artifact of the Tier 3 combination of MP2 energies with DFT geometries and phonons. Nevertheless, the actual predicted molar volumes are reasonable.

Finally, employing pure B86bPBE-XDM (Tier 4) also predicts the molar volumes rather accurately. It significantly overestimates the molar volume in carbon dioxide, but it performs well for the other three crystals. Notably, B86bPBE-XDM reproduces the experimental molar volume in imidazole fairly well, where MP2 significantly overbinds the crystal and underestimates the volume due to its well-known problems describing van der Waals dispersion.[283] That MP2 problem will be addressed below.

Interestingly, the Tier 2 and 3 (and to a lesser extent Tier 4) results here tend to exhibit less thermal expansion than Tier 1. One possible explanation could be that the former included only Γ-point phonons, while the Tier 1 results from Ref 106 include lattice dynamical phonon dispersion. We previously showed that including phonon dispersion increased the predicted rate of thermal expansion.[105] Supercell phonon calculations quickly become computationally demanding in DFT. With the fragment HMBI approach, on the other hand, they can be performed with very little additional cost over the Γ-point only calculation, since the unique additional contributions in the supercell are handled at the MM level. Other more subtle factors involving the interplay between the $E(V)$ curves and $F_{vib}$ contributions likely contribute as well. In acetic acid, for instance, the Tier 2 model expands more rapidly with temperature than Tier 3, despite both neglecting phonon dispersion.

Ice Ih is unusual in that it exhibits negative thermal expansion, with the volume contracting by around 0.06% between 10 K and 70 K, before expanding. Most of the models studied here reproduce this trend qualitatively, though fewer are quantitatively correct. We previously found that Tier 1 MP2/CBS + AMOEBA predicts a slight contraction of only 0.1% and at at a lower 40 K. Interestingly, a different MP2-based fragment study[284] also predicted the contraction in the aug-cc-pVDZ and aug-cc-pVTZ basis sets, though they found that MP2 exaggerated the contraction and predicted the minimum at higher temperatures. The discrepancy between those results and our earlier ones might result from the use of counterpoise correction in our work, which leads to underbinding of the crystal, versus no counterpoise correction in Ref 284, which will lead to overbinding. Counterpoise correction will impact the balance between the key hydrogen bonding phonon modes and hydrogen stretching modes.

Here, we examine how the hybrid approaches behave. Appendix B Figure 5 plots the relative molar volumes at low temperature for several representative methods from Tiers 2, 3, and 4. While Tier 2 MP2/aug-cc-pVTZ + AMOEBA demonstrates negligible contraction of the lattice, the Tier 3 and Tier 4 results perform somewhat better. They correctly predict the location of the minimum near 60–70 K, but they underestimate the amount of contraction at only 0.02–0.03%.

Taken together, these results reiterate that dispersion-corrected density functionals like B86bPBE-XDM can predict thermal expansion well. Tier 1 MP2 molar volumes and thermal expansion often differ from the DFT ones, but refining the single-point energies with MP2 already corrects for much of that difference. Such single-point energy refinements can

be achieved with relatively modest additional computational cost. The further advantages of Tier 2 or Tier 1 for molar volumes are somewhat smaller, and will not be computationally worthwhile for predicting structural parameters in many applications.

Of course, these results represent only the B86bPBE-XDM functional. Additional insight is obtained by considering the behavior of Tiers 2–4 with several different density functionals and either the XDM or D2 dispersion correction. Figure 4.4a plots the predicted thermal expansion with four different functionals: BLYP-D2, PBE-D2, PBE-XDM, and B86bPBE-XDM. The functionals overestimate the 0 K molar volume by ∼10–15%, and the rate of thermal expansion differs considerably between the D2 and XDM dispersion corrections. Models corrected with D2 generally predict larger thermal expansion than those employing XDM. Because they derive their dispersion coefficients directly from electronic structure dispersion models like XDM, Tkatchenko-Scheffler (TS),[71] or many-body dispersion (MBD),[73–75] are generally expected to be superior to a purely empirical correction like D2.[68] The D3 or D4 models,[69,70] which adapts the dispersion coefficients based on the chemical environment would provide a better comparison, but those are not currently implemented in Quantum Espresso.

Regardless, focus not on the performance of specific density functionals and/or dispersion corrections, but rather what happens when ingredients from those functionals are used in Tier 2 or 3 calculations. As shown in Figures 4.4b, simply refining the calculations with MP2 single point energies (Tier 3) dramatically reduces the sensitivity of the calculations to the density functional used to optimize the structures and compute the phonons. At Tier 2, where only the phonons are obtained from DFT, the sensitivity of the

predicted molar volumes to the functional decreases further. Additional sensitivity data is provided in Appendix B Figures 6 and 7. These results suggest that although the DFT geometries defining the $E(V)$ curve clearly differ from each other and the MP2/aug-cc-pVTZ ones, single-point energy refinement renormalizes much of the variations among them and gives $E(V)$ curves in generally good agreement with one another. In this context, it is also worth noting that density functional tight binding proved successful in modeling the thermal expansion in carbamazepine, and this could provide another route to low-cost phonons for Tiers 2 and 3.[235]

Given the good performance and relatively low computational cost of using HMBI-refined single-point energies with DFT optimizations and phonons (Tier 3), the next question is to see how reliably one can predict the molar volumes by further improving the quality of the HMBI single-point energies. These energies can be refined in three ways. First, one might increase the basis set size from aug-cc-pVTZ to the complete basis set limit via a triple-zeta/quadruple-zeta extrapolation. The counterpoise-corrected energies tend to underestimate the interactions in small basis sets, leading to overestimation of the molar volume.[106]

Second, one could replace the AMOEBA many-body treatment with one from periodic HF. Recent work found that AMOEBA overestimates many-body polarization associated with the cooperative hydrogen bonding in crystalline methanol, and refining the energetics with periodic HF proved important for obtaining the correct phase diagram.[109] Similar over-polarization is found for AMOEBA treatments of ion-water interactions.[285]

Strong polarization effects are present in several of the crystals here, so there could be appreciable benefits to periodic HF refinement here as well.

Third, MP2 exhibits well-known problems with the treatment of van der Waals dispersion.[283] Among the four crystals here, this is most apparent for imidazole, where MP2 substantially overbinds the lattice energy[101,103] and underestimates the molar volume (see Figure 4.2d).[106] Ideally, one would correct this problem by replacing MP2 with CCSD(T). However, dispersion-corrected MP2C[261,262] provides a more pragmatic approach which corrects most of the MP2 error in these crystals at much lower computational cost.[264]

Figure 4.5 examines this hierarchy of refinements for the four crystals. As seen in the earlier discussion of Figure 4.2, switching from (1) pure B86bPBE-XDM (Tier 4) to (2) Tier 3 MP2/aug-cc-pVTZ + AMOEBA single-point energies leads to a substantial change in the predicted molar volumes, usually toward smaller unit cells (the slight volume increase in ice is the exception). (3) Replacing the AMOEBA many-body contribution with periodic HF increases the molar volume. For carbon dioxide, ice, and acetic acid, the effect is fairly small. For imidazole, on the other hand, it increases the volume by ∼9%. As expected, (4) increasing the MP2 basis set toward the CBS limit induces a volume contraction of ∼0.4 $cm^3$/mol in ice and around 1 $cm^3$/mol in the other three crystals. With the exception of imidazole, the MP2/CBS + periodic HF predicted volumes are slightly smaller than experiment. Surprisingly, in imidazole, the MP2/CBS + periodic HF volumes agree very well with experiment, despite MP2's overestimation of the lattice energy. In all four cases, (5) MP2C weakens the binding and shifts toward larger molar volumes. Unsurprisingly, the

largest corrections are observed for carbon dioxide and imidazole, while hydrogen bonded ice and acetic acid are less affected by the dispersion correction.

Overall, the combination of MP2C/CBS + periodic HF energies on B86bPBE-XDM geometries and phonons predicts the highest-temperature experimental molar volumes of carbon dioxide, ice, and acetic acid to within 1–2%. For imidazole, the errors are 2% or 3.5%, depending on which high-temperature experimental structure one compares against. Finally, notice that the different possible Tier 3 single-point energy refinements considered here all predict similar thermal expansivity (curves (2)–(5) in Figure 4.5 are largely parallel). The most significant variations manifest as shifts in the 0 K molar volume. This reflects that while the minimum of the $E(V)$ curves shifts upon single-point energy refinement, changes to the potential energy well curvature are apparently minor (see Appendix B Figure 8).

### 4.3.2 Thermochemical data

Table 4.2: Root-mean-square error (in kJ/mol) between predicted and experimental sublimation enthalpies over the temperatures for which experimental data is available (see in Figure 4.6). Errors were computed by splining the data curves and taking differences between them at 1 K intervals.

| Crystal | Temperature Range | Tier 3 MP2/CBS+pHF | Tier 3 MP2C/CBS+pHF | Tier 4 B86bPBE-XDM |
|---|---|---|---|---|
| Carbon Dioxide | 0–195 K | 1.7 | 2.1 | 4.2 |
| Ice | 10–265 K | 2.5 | 1.9 | 8.9 |
| Acetic Acid | 50–278 K | 1.3 | 1.0 | 5.6 |
| Imidazole | 90–283 K | 9.2 | 0.9 | 2.6 |

A key reason for modeling thermal expansion in molecular crystals is to predict temperature-dependence of other crystal properties. Accordingly, we consider the enthalpy and entropy of sublimation for these four crystals. Experimental values were taken and/or

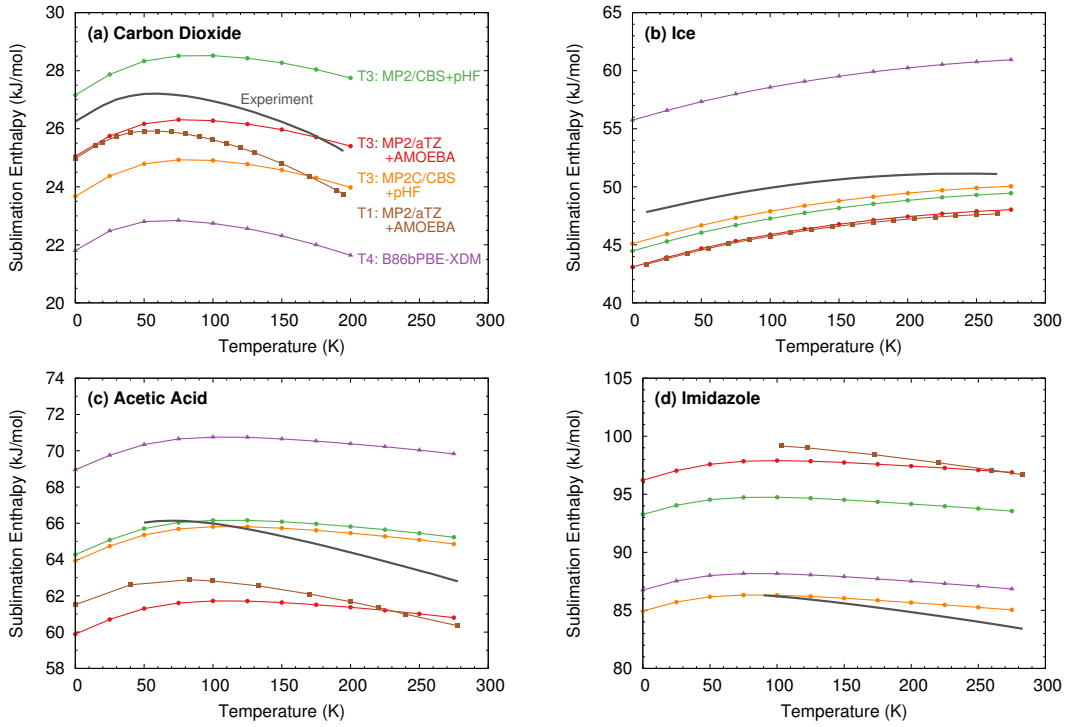



Figure 4.6: Comparison of predicted sublimation enthalpies for (a) carbon dioxide, (b) ice, (c) acetic acid, and (d) imidazole using Tier 1 (T1) MP2/aug-cc-pVTZ + AMOEBA, Tier 4 (T4) B86bPBE-XDM, and Tier 3 (T3) with several different energy refinements.

derived from the literature as described in Refs 106 and 105. Quantitative experimental uncertainties are not readily available for the thermochemical properties presented here. The temperature-dependent results were frequently derived using data from a variety of sources that did not always report uncertainties. Moreover, different studies do not always agree (e.g. reported room-temperature sublimation enthalpies for imidazole differ by several kJ/mol[286]). Crystalline defects in the experimental solids can also play a role. Forming a point defect is typically an endothermic process, but it also leads to a gain in entropy. Quantifying the impact of the various factors here is difficult, but one should probably allow for uncertainties of up to a few kJ/mol in the sublimation enthalpies, for example.

Figure 4.6 compares sublimation enthalpies as predicted with several of the techniques described in the previous section against experiment. Given the relatively small improvements in molar volume provided by Tier 2 over Tier 3, the discussion here focuses primarily on Tiers 3 and 4. Unless otherwise mentioned, all DFT results in this section employ B86bPBE-XDM.

As temperature increases, the sublimation enthalpy $\Delta H_{sub} = H_{gas} - H_{crystal}$ typically first increases before reaching a maximum and then decreasing. Heating increases $H_{gas}$ via destabilizing translational, rotational, and $PV$ contributions. $H_{crystal}$ derives from a balance between the thermal expansion and the vibrational contributions, but it is typically dominated by lattice energy destabilization due to thermal expansion. Little thermal expansion occurs in the crystal at low temperatures, and the $\Delta H_{sub}$ increases due primarily to the gas contribution in that temperature regime. At higher temperatures, however, the larger thermal expansion in the crystal eventually leads to the increase in $H_{crystal}$ becoming more significant than the changes in $H_{gas}$, which produces the maximum and subsequent decrease in $\Delta H_{sub}$ seen for carbon dioxide, acetic acid, and imidazole. In ice, this maximum does not occur below the melting point due to the low thermal expansivity. See Ref 106 for more detailed discussion of the enthalpy behaviors.

Earlier Tier 1 work[105,106] demonstrated that changing the electronic structure method and basis set shifts the entire sublimation curves vertically but has minor impact on the curvature. Larger basis sets in particular shift the enthalpy toward larger values. Obtaining the proper maximum and subsequent decrease of the sublimation enthalpy at higher temperatures requires capturing thermal expansion correctly. A simple harmonic

model that neglects thermal expansion agrees well with the quasi-harmonic result at low temperatures, but it overestimates the enthalpy at higher temperatures by about 1 kJ/mol in these crystals. Ice proves the exception: since it exhibits comparatively little thermal expansion, the error introduced in $\Delta H_{sub}$ by neglecting that expansion at 273 K is negligible.

In the present work, the Tier 2-3 results frequently underestimate the amount of thermal expansion compared to Tier 1. In Figure 4.6, this underestimated thermal expansion manifests in the Tier 3 MP2/aug-cc-pVTZ + AMOEBA sublimation enthalpy (red curve) being too large at higher temperatures, analogously to what was observed for the harmonic model.[106] Nevertheless, those Tier 3 results generally mimic the Tier 1 ones to within 1–2 kJ/mol across the full temperature range.

Refining Tier 1 by increasing the basis set size to the CBS limit, applying the MP2C dispersion correction, and replacing AMOEBA with periodic HF does not resolve the issue of the erroneous slope in $\Delta H_{sub}$ at higher temperatures. Nevertheless, MP2C/CBS + periodic HF results (orange curve) provide excellent agreement with experiment across the temperature range, with root-mean-square (rms) errors ranging 0.9–2.1 kJ/mol for the four crystals (Table 4.2). Applying the dispersion correction in MP2C has minimal impact on hydrogen-bonded ice and acetic acid, but it alters the sublimation enthalpies in carbon dioxide and imidazole considerably (compare green vs. orange curves). In our earlier work, Tier 1 MP2/CBS and CCSD(T)/CBS both overestimated the sublimation enthalpy for carbon dioxide, with CCSD(T) predicting a slightly smaller $\Delta H_{sub}$.[105,106] Here, it appears that MP2C over-corrects the MP2 result, with the net result of slightly increasing the rms

error from 1.7 to 2.1 kJ/mol. On the other hand, MP2C performs very well for imidazole, reducing the rms error in $\Delta H_{sub}$ from 9.2 kJ/mol (MP2) to 0.8 kJ/mol (MP2C).

Compared to the best Tier 3 MP2C results, the errors for B86bPBE-XDM are 2–3 times larger for carbon dioxide and imidazole, where van der Waals dispersion contributes appreciably, and 5–6 times larger for acetic acid and ice, which are dominated by hydrogen bonding interactions (Table 4.2). Notably, B86bPBE-XDM performs very well for imidazole, which is problematic for MP2. Overall, the sublimation enthalpy data here bolsters the case for Tier 3 single-point energy refinement with MP2C/CBS + periodic HF on top of the DFT structures and phonons.

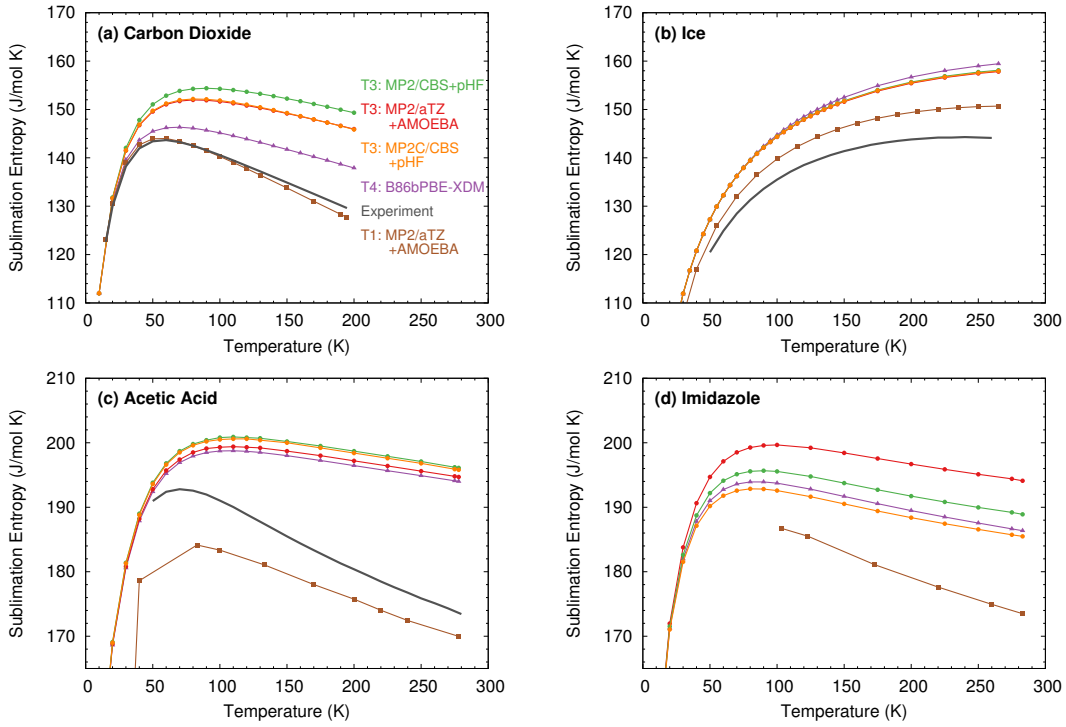



Figure 4.7: Comparison of predicted sublimation entropies for (a) carbon dioxide, (b) ice, (c) acetic acid, and (d) imidazole using Tier 1 MP2/aug-cc-pVTZ + AMOEBA, Tier 4 B86bPBE-XDM, and Tier 3 with several different energy refinements. Experimental entropy data is unavailable for imidazole.

Next, consider prediction of the sublimation entropy. Whereas $H_{crystal}$ is dominated by the lattice energy and therefore sensitive to the final single-point energy, $S_{crystal}$ is governed by the phonon frequencies. Tiers 2–4 all compute the phonons with DFT. The improved energies in Tier 2 or Tier 3 will alter the predicted volume at a given temperature, which in turn impacts the phonons via their volume dependence (Eq 4.5).

Like for the enthalpy, the temperature dependence of the sublimation entropy $\Delta S_{sub}(T) = S_{gas}(T) - S_{crystal}(T)$ arises from a competition between the entropy of the gas phase and the crystal phase.[106] At low temperatures, $S_{gas}(T)$ rises more quickly with temperature. At higher temperatures, the situation reverses, with $S_{crystal}(T)$ increasing more quickly due to the phonon contribution. Together, these produce the overall concave shape of $\Delta S_{sub}(T)$. If thermal expansion is not accounted for, the phonon frequencies tend to be too large, and $S_{crystal}(T)$ will be too small. This causes overestimation of the entropy at high temperatures.

Here, the thermal expansion is frequently underestimated relative both to experiment and Tier 1 calculations. This translates to consistent overestimation of $\Delta S_{sub}(T)$ by up to 10–12% relative to experiment at high temperatures (Figure 4.7). In cases like ice and acetic acid, where the volume is not too sensitive to specific method used for Tier 3 energy refinement, the variation in predicted sublimation entropies is fairly small across the different energy models. Larger variations are observed among the different Tier 3 models for imidazole and and carbon dioxide, commensurate with the greater sensitivity of the molar volumes to the modeling approach in those crystals. Overall, in marked contrast to the molar volumes and sublimation enthalpies, Tier 3 energy refinement does not clearly

improve the quality the underlying DFT entropies. The same can be said for Tier 2 (not shown), since the Tier 2 and 3 sublimation entropies differ by less than 1%.

Calculating phase transition temperatures provides another means of assessing predicted thermochemistry values, since the transition temperature can be sensitive to subtle changes in the Gibbs free energy.[105,106,109] Computing the Gibbs free energy accurately requires balance between the enthalpic and entropic components, which may or many not be maintained in the hybrid approaches.

Experimentally, carbon dioxide sublimes at 194.7 K and atmospheric pressure.[249] Table 4.3 summarizes predicted sublimation temperatures with several different methods. Large-basis Tier 1 MP2 and CCSD(T) perform excellently, predicting the sublimation temperature of 5 degrees Kelvin of experiment. In smaller basis sets, it underestimates the sublimation temperature. At the other extreme, B86bPBE-XDM underestimates the sublimation temperature by almost 40 K. Tier 3 refinement at the MP2 level improves the predicted sublimation temperature somewhat, though the Tier 3 temperatures are $\sim$10 K lower than the corresponding Tier 1 ones. As noted earlier, MP2C seemingly over-corrects MP2, to the detriment of the predicted sublimation temperature. The $\sim$30 K error in the MP2C/CBS + periodic HF sublimation temperature is only modestly better than that of B86bPBE-XDM.

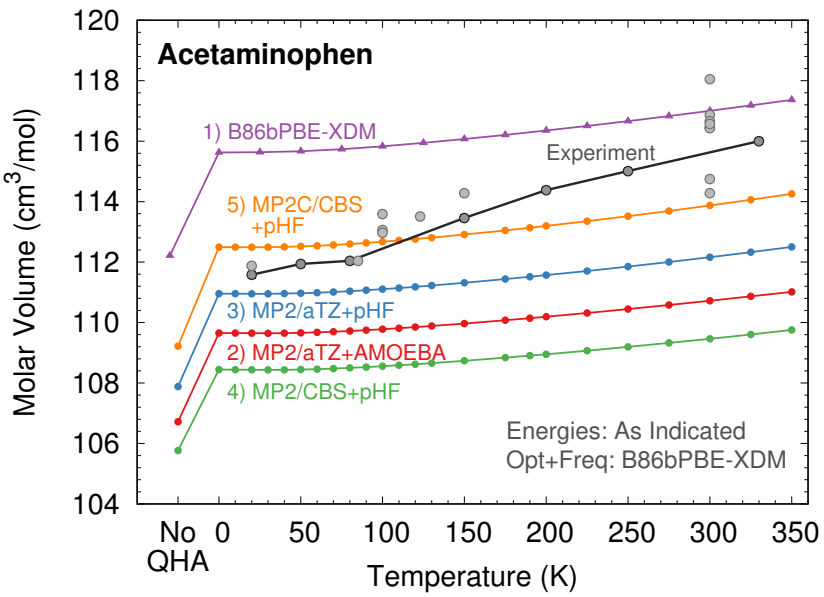



Figure 4.8: Predicted acetaminophen molar volumes versus temperature. The connected dark gray points indicate data from a temperature-dependent neutron scattering study,[11] while light gray points represent other experimental data points found in the CSD.

### 4.3.3 Acetaminophen form I

To investigate how well these models perform in a pharmaceutically relevant species, Figure 4.8 plots thermal expansion data for acetaminophen. Twenty two experimental crystal structures of ambient pressure form I have been reported in the CSD. While most of the reported molar volumes exhibit reasonable agreement, a wide $\sim$4 cm$^3$/mol scatter exists in the reported room temperature values. Seven of the structures (those connected with a line in Figure 4.8) come from a single temperature-dependent neutron scattering study.[11] The discussion below focuses primarily on this consistent set of neutron scattering data.

Given the large size of acetaminophen ($C_8H_9NO_2$ and $Z = 4$ in form I), calculations were only performed at Tiers 3 and 4. Tiers 1 and 2 would require substantially

more computational effort. With a predicted molar volume of 115.6 $cm^3$/mol at 20 K, Tier 4 B86bPBE-XDM overestimates the experimental molar volume of 111.6 $cm^3$/mol (HXACAN13[11]) by 3.6%. It also significantly underestimates the thermal expansivity, with the B86bPBE-XDM molar volume increasing only 1.3% versus 3.8% experimentally. The combination of overestimated volume at low temperature and underestimated thermal expansivity leads to much smaller errors in the molar volumes at high temperatures, e.g. only 1% error at 330 K (HXACAN19[11]).

Switching to Tier 3 with MP2/aug-cc-pVTZ + AMOEBA leads to a considerable reduction in the molar volume to 109.65 $cm^3$/mol, which is in somewhat better agreement with experiment (1.7% error at 20 K). However, the underestimated thermal expansivity persists, and the volume at 330 K is underestimated by 4.4%. Moving toward the CBS limit, replacing MP2 with MP2C and AMOEBA with periodic HF both shift back toward larger molar volumes, but they have little impact on the thermal expansivity. In the end, the MP2C/CBS + periodic HF volumes are fairly close to experiment, especially at low temperatures (0.8% error at 20 K), but they underestimate the volume by 1.6% at 330 K due to the low thermal expansivity.

Next, we turn to the form I acetaminophen sublimation enthalpy and entropy. Perlovich and co-workers measured $\Delta H_{sub}(298\text{K}) = 117.9 \pm 0.7$ kJ/mol from vapor pressure measurements,[287,288] while Picciochi at al subsequently obtained 129.9±1.4 kJ/mol from a combination of vapor pressure measurements and microcalorimetry experiments (and they proposed a slight refinement of Perlovich's value to 118.9 ± 1.6 kJ/mol).[289] The ∼10 kJ/mol discrepancy between the experimental values is significantly larger than the reported 1–2

kJ/mol experimental uncertainties. The quasi-harmonic calculations here can address this discrepancy.

Table 4.4 compares the predicted and experimental room-temperature sublimation enthalpies. Tier 4 B86bPBE-XDM predicts 131.0 kJ/mol, which lies within the stated uncertainty in Picciochi et al's value of 129.9 ± 1.4 kJ/mol. MP2 predicts a sublimation enthalpy that is up to 23 kJ/mol higher, but employing MP2C/CBS + periodic HF reduces this enthalpy to 138.0 kJ/mol. This is 8 kJ/mol larger than the Picciochi et al experimental value. Some of this overestimation may stem from the underpredicted thermal expansion, as seen for acetic acid and imidazole earlier. In any case, both the MP2C and B86bPBE-XDM results support the larger 129.9 ± 1.4 experimental sublimation enthalpy instead of the smaller 117.9 ± 0.7 kJ/mol value.

Finally, Perlovich et al reported a sublimation entropy of 190 ± 2 J/(mol K) at 298 K. Tier 4 B86bPBE-XDM and various Tier 3 refinements all predict much larger values of 240–246 J/(mol K). Overestimation of the sublimation entropy is expected based on the results for the other crystals described above and by how much the predictions underestimate the thermal expansion. Still, this ∼25% disagreement between theory and experiment is significantly larger than the ∼10–12% seen in the smaller crystals. On the other hand, it is not clear how reliable the reported experimental entropy value is, since it was derived from the sublimation enthalpy measurement[287] which is suspect in light of the Picciochi et al experiments[289] and our calculations here.

## 4.4 Conclusions

The combination of fragment-based correlated wavefunction electronic structure techniques and quasi-harmonic approximations can be very successful in predicting thermal expansion and other properties, at least for rigid-molecule crystals. On the other hand, dispersion-corrected DFT models also often perform well at appreciably lower computational cost. By comparing a hierarchy of models ranging from pure DFT to pure MP2, we have demonstrated that introducing correlated wavefunction energies on top of DFT geometries and phonons can appreciably improve predicted molar volumes and enthalpies. One must ensure, however, that the single-point energies are well-chosen, which in practice means employing large basis sets and accounting for any deficiencies in the models (e.g. by replacing MP2 with MP2C for the monomer and dimer contributions or AMOEBA with HF for the many-body contributions).

In the small molecule crystals, the hybrid Tier 3 approach led to predicting molar volumes to within a few percent and sublimation enthalpies to within 1–2 kJ/mol of experiment over broad temperature ranges. Though more computationally expensive than pure DFT approaches, such single-point energy refinement can be performed on crystals containing dozens of atoms per molecule, especially when crystal symmetry can be exploited. Using these techniques, we demonstrated that the predicted sublimation enthalpy for acetaminophen agrees better with the more recent experimental value of Picciochi et al than the earlier reported value.

Issues remain, however. The hybrid approaches that combine DFT and wavefunction techniques do not appreciably improve the predicted entropies, since those are largely

governed by phonon contributions. We also found that the hybrid calculations here generally underestimate the amount of thermal expansion. In the smaller crystals the impact was modest, but it becomes more significant in acetaminophen. Acetaminophen also exhibited considerably larger 8 kJ/mol errors in the predicted sublimation enthalpy than for the smaller crystals. Accounting for phonon dispersion might help, but anharmonicity or other factors may also play an important role. More testing in larger, more flexible molecules is needed to assess the performance of the quasi-harmonic approach. Efficient and effective strategies for addressing anharmonicity would be especially valuable.

Moreover, though the performance of the hybrid approaches is quite good, it may not always be sufficient. The imbalances between the enthalpy and entropy appreciably increased the error in the predicted carbon dioxide sublimation temperature, for example. More dramatically, in our recent study of the methanol phase diagram,[109] even a half kJ/mol error in the relative free energies between two polymorphs shifts the phase transition temperature by $\sim$100 K. For such challenging properties, more reliable and computationally demanding Tier 1 approaches may be needed.

Table 4.3: Comparison of experimental and predicted sublimation temperatures for carbon dioxide.

| Tier | Method | $T_{sub}$ (K) |
|---|---|---|
| | Experiment[a] | 194.7 |
| Tier 1 | MP2/aDZ + AMOEBA[b] | 163.6 |
| Tier 1 | MP2/aTZ + AMOEBA[b] | 185.3 |
| Tier 1 | MP2/aQZ + AMOEBA[b] | 193.4 |
| Tier 1 | MP2/CBS + AMOEBA[b] | 199.2 |
| Tier 1 | CCSD(T)/CBS + AMOEBA[b] | 201.0 |
| Tier 3 | MP2/aTZ + AMOEBA | 174.2 |
| Tier 3 | MP2/CBS + pHF | 186.6 |
| Tier 3 | MP2C/CBS + pHF | 165.7 |
| Tier 4 | B86bPBE-XDM | 157.5 |

[a] Ref 249 [b] Ref 105

Table 4.4: Comparison of experimental and predicted 298 K sublimation enthalpies (kJ/mol) and entropies (J/(mol K)) for acetaminophen form I.

| Method | Source | $\Delta H_{sub}$ (298 K) | $\Delta S_{sub}$ (298 K) |
|---|---|---|---|
| Perlovich et al | Ref 287,288 | 117.9 $\pm$0.7 | 190 $\pm$ 2 |
| Picciochi et al | Ref 289 | 129.9 $\pm$ 1.4 | |
| B86bPBE-XDM | Tier 4 | 131.0 | 242.1 |
| MP2/aTZ+AMOEBA | Tier 3 | 146.4 | 245.6 |
| MP2/aTZ+pHF | Tier 3 | 147.4 | 244.8 |
| MP2/CBS+pHF | Tier 3 | 154.1 | 246.3 |
| MP2C/CBS+pHF | Tier 3 | 138.0 | 243.9 |

# Chapter 5

# Investigating the phase transition of $\alpha-$ and $\beta-$ Resorcinol

In Chapter 4 a hierarchy of models for calculating the quasi-harmonic approximation (QHA) was introduced. It was found that we can compute the geometries and phonons of the crystal using density functional theory (DFT) and perform single point energy corrections to find the volume of the crystal using a more robust method (most promisingly: dispersion-corrected second-order Møller-Plesset perturbation theory (MP2) and periodic Hartree-Fock with large basis sets). Compared to performing the entire calculation using pure MP2, this leads to minor loss in chemical accuracy and provides a necessary increase in the speed of computation. Additionally, employing this method increases the system size we can feasibly simulate to over 100 atoms per central unit cell, allowing us to simulate pharmaceutically-relevant molecular crystals. In this chapter employ this new tiered-QHA to predict the phase transition properties of $\alpha$- and $\beta$-resorcinol.

## 5.1 Introduction

Resorcinol is a long-studied crystal with the first two phases ($\alpha-$ and $\beta-$phase) resolved by Robinson et al in 1936 using X-ray diffraction.[290] Both phases are stable at ambient pressure and share the same $Pna2_1$ space group, with $\alpha$ found to be the thermodynamically stable polymorph. The $\beta$ form, despite being the denser polymorph, is considered to be meta-stable to the $\alpha$ form, and it and will eventually transform back into the $\alpha$ polymorph over time. The $\alpha \rightarrow \beta$ phase transition occurs upon heating of the $\alpha$ polymorph at ambient pressure, with the most reliable transition temperature at 369 $\pm$ 6 K reported by Ebisuzaki et al in 1987.[291] This transition can also be achieved by compressing the $\alpha-$polymorph at pressures $\sim 0.5$ GPa.[292] The transition temperature depends strongly both on the rate of heating and the rate of pressurization implying that kinetics plays a large role in this transition.[291–294] Kichanov et al have reported the existence of the $\alpha-$polymorph in pressures reaching up to 4.0 GPa and the $\beta-$polymorph as high at 5.6 GPa.[292] A plethora of experimental information regarding the Raman spectra exists as well as a few studies on thermodynamic properties of each polymorph.[291–293,295–297] Two other high-pressure polymorphs had been reported for resorcinol ($\gamma$ and $\delta$) however neither of these crystal structures have been resolved.[292,294]

Until very recently, it was thought that only $\alpha-$ and $\beta-$ polymorphs could be observed at atmospheric pressure. However, in 2016 Zhu et al published a joint computational crystal structure prediction and experimental study confirming the existence of one additional phase of resorcinol at atmospheric conditions ($\epsilon-$phase) and suggested there might be a fourth as well.[298] They reported that the $\epsilon-$phase is always grown concomitantly with

other phases and is always polycrystalline, which made experimental synthesis challenging. They also reported that, despite the relative simplicity of the molecule, the modeling was challenging, requiring both a re-parameterized force-field (as was previously needed by Chatchawalsaisin et al) and harmonic phonons to achieve correct stability ranking of these polymorphs.[299]

Recently, Červinka and Beran conducted a first-principles study on the methanol phase diagram which achieved the best agreement to date between predicted and experimental phase diagrams.[109] However, computing the methanol phase diagram required resolving free energy differences of ∼0.5 kJ/mol accuracy and had a computational cost of about 300,000 CPU hours. Here we attempt employ the tiered-QHA method that was developed in Chapter 4 to model the phase-transitions of resorcinol. This system will be an ideal test of whether we can retain the sub-kJ/mol accuracy required for phase diagram prediction for moderately sized organic molecules. Ideally, our goal would be to model the full phase diagram including the confirmed $\epsilon-$phase and the as-yet unconfirmed $P2_1$ structure to establish the stability region of each of these polymorphs. To begin, we first focus on phase transition of $\alpha-$ and $\beta-$resorcinol. As was forewarned by Zhu et al, the system is surprisingly challenging. We report the most accurate modeling to date (to our knowledge) of the unit cell volumes for both polymorphs at low pressures. However, inconsistencies in simulated thermodynamic properties suggests that further work must be done to resolve the relative free energies of these crystals.

## 5.2 Theory

For this study we will use the hybrid Tier 3 QHA approach that was established in Chapter 4 to simulate the Gibbs free energies. Specifically, we combine DFT structures and harmonic phonons with and single-point energy (SPE) corrections that refine the electronic energy surface up to the desired level of theory. In the small molecule crystals, the hybrid Tier 3 approach led to predicting molar volumes to within a few percent and sublimation enthalpies to within 1–2 kJ/mol of experiment over broad temperature ranges. Refer to Chapter 4 for further details.

Previously the electronic energy surfaces were fitted to a Murnaghan equation of state curve in order to minimize the amount of volumes that are needed to properly sample the Gibbs free energy. While this approach readily captured the basin of the electronic energy curve, thermal expansion to high temperatures were found to introduce up to 2 kJ/mol error in the predicted minima of the Gibbs free energy surface. To correct this error, we instead choose to explicitly sample more of the electronic energy surface, increasing the volumes that need to be sampled explicitly from 10 to 20-30.

We wrote a MATLAB script to compute the Helmholtz vibrational free energy at every volume sampled on the electronic energy surface using mode-specific Grüneisen parameters. The Gibbs free energy curve is computed by summing the electronic energy curve, the Helmholtz vibrational energy curve, and the PV term at any given pressure and temperature. Since the volume which minimizes the free energy for a given temperature and pressure is unlikely to correspond to one of the explicitly sampled volumes, each free energy curve is fitted to a weighted double-Murnaghan equation of state. The Murnaghan

equation of state is given by,

$$G(V) = G_0 + \frac{B_0 V}{B_0'}\left[\frac{(V_0/V)^{B_0'}}{B_0' - 1} + 1\right] - \frac{B_0 V_0}{B_0' - 1} \tag{5.1}$$

where $G_0$, $V_0$, $B_0$, and $B_0'$ are the fit parameters. $G_0$ gives the free energy at the minimum, $V_0$ is the molar volume at the minimum energy, $B_0$ is the bulk modulus, and $B_0'$ is the first derivative of the bulk modulus with respect to pressure. This method identifies the free-energy minimum effectively while avoiding artifacts that can be caused by simpler equation of state fits or splines.[108]

## 5.3 Computational details

Structures for both $\alpha-$ and $\beta-$resorcinol (RESORA03 and RESORA08 respectively) were taken from the Cambridge structural database (CSD). Each structure was initially optimized to its electronic minimum using DFT. An electronic energy surface was then mapped out using isotropic pressure, and single point energy corrections were then employed to corrected the potential energy surface up to the desired level of theory. Each structure was then subsequently expanded to a chosen temperature and pressure using the QHA protocol outlined in section 5.2.

### 5.3.1 DFT structure optimizations

For the calculations performed here, the crystals were first optimized with periodic boundary/planewave DFT in Quantum Espresso v6.1, using the B86bPBE density functional and the exchange-hole dipole method (XDM) dispersion correction.[2–4,201] Core electrons were treated according to the projector augmented wave (PAW) approach, and

PAW potentials for H, C, and O were produced using A. Dal Corso's Atomic code v6.1.[202] Optimizations were carried out using a 50 Ry planewave energy cutoff. Reciprocal space k-points were placed with a Monkhorst-Pack grid with a 1x1x3 grid for $\alpha-$resorcinol and a 3x1x3 grid for $\beta-$resorcinol.[203] $\Gamma$-point harmonic vibrational frequencies were computed using finite difference with Phonopy v1.11.2.[204] Predicted room-temperature lattice parameters are compared against the available experimental parameters from Reference 292 (Appendix C Section 1).

### 5.3.2 Single-point energy corrections

Single-point energy corrections were calculated using using the Hybrid Many-Body Interaction model (HMBI) v2.0 (available on github https://github.com/gberan/HMBI).[94–96] Density-fitted MP2 was carried out in the Dunning aug-cc-pVXZ basis sets (X = D, T, or Q) using Molpro v2012.1[260] Because MP2 has well-known problems with van der Waals dispersion, we also refined the single-point energies with MP2C, also using Molpro.[261,262] MP2C replaces the uncoupled HF treatment of dispersion found in MP2 with an improved coupled Kohn-Sham treatment, and it performs very well for non-covalent interactions.[263] Because the MP2C dispersion correction is less sensitive to basis set than MP2, the correction was computed in the aug-cc-pVTZ basis set and then combined with MP2 in various basis sets (aug-cc-pVTZ or larger).[264] Extrapolation to the complete basis set limit from the triple and quadruple zeta basis sets was also performed.[265,266] All dimer calculations in HMBI employed a counterpoise correction to account for basis set superposition error.

The long-range and many-body correction in the HMBI energies were calculated with either the atomic multipole optimized energetics for biomolecular applications (AMOEBA)

force field or periodic Hartree-Fock (pHF) and the pob-TZVP basis.[269] The AMOEBA contributions were evaluated using Tinker 7.1 and the pHF calculations were done using CRYSTAL17.[267,270,300] Poltype was used to generate force field parameters for both forms of resorcinol.[268] For the AMOEBA calculations only the intermolecular force field parameters are needed in the context of HMBI, since intramolecular contributions are treated quantum mechanically. The pob-TZVP basis is a variant of def2-TZVP adapted for periodic calculations, and was chosen here because an earlier studies found that many-body contributions in small molecular clusters computed in the pob-TZVP basis set without counterpoise correction compared well against those from in large basis sets.[107,109] Basis set superposition error is a potential problem in Gaussian basis set calculations on periodic crystals, but addressing it in the context of the many-body calculations here is challenging. A $10^{-7}$ a.u. tolerance was used for the Coulomb overlap threshold, Coulomb penetration threshold, and exchange overlap threshold, and tolerances of $10^{-12}$ and $10^{-30}$ a.u. for the pseudo-overlaps (see Ref 300 for details). Monkhorst-Pack and Gilat shrinking factors of 6 6 6 were used.

## 5.4 Results and Discussion

Section 5.4.1 examines the effect correcting the potential energy surface up to different electronic structure theory levels has on the predicted molar volumes for each structure. Section 5.4.2 investigates the the predicted phase transition and the predicted thermodynamic properties at ambient pressure. Section 5.4.3 more closely investigates the predicted free energies of both polymorphs calculated at the MP2C/CBS+pHF level and the sensitivity of this phase transition.

### 5.4.1 Assessment of thermal expansion

To begin, we investigate the quality of the predicted molar volumes for both polymorphs relative to experiment at different potential energy surfaces. Upon applying the quasi-harmonic approximation, both structures can be expanded to any desired pressure and temperature. First we examine the quality of the predicted volumes at a fixed pressure. Figure 5.1 plots the thermal expansion of both structures. For the $\alpha$ phase, the volumes are computed at 0.0 GPa which is comparable to ambient pressure ($10^{-4}$ GPa) while the $\beta$ volumes are computed at 0.09 GPa which is consistent with the experimental conditions.[292]

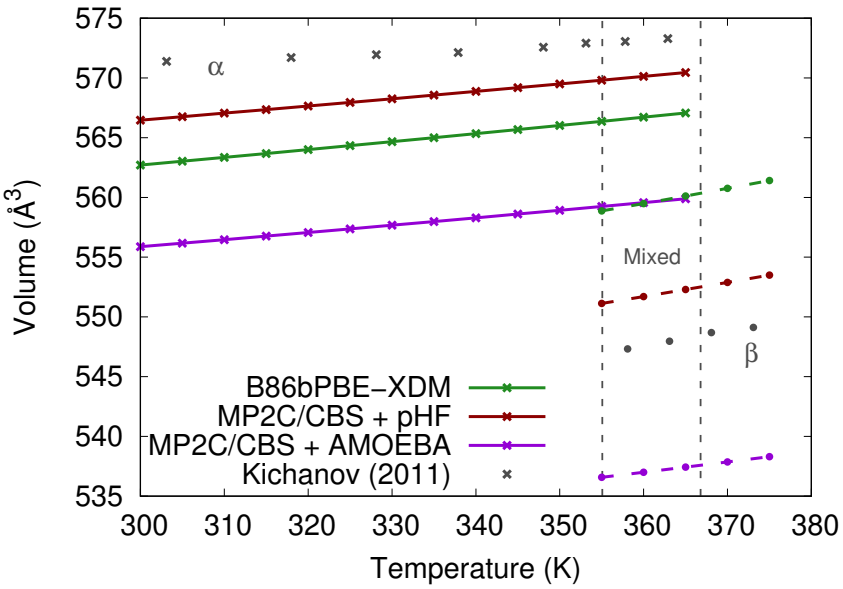



Figure 5.1: Predicted volume-expansion for $\alpha-$ and $\beta$-resorcinol. The $\alpha-$polymorph is displayed in x's while the $\beta$-polmorph is displayed in dots.

Looking first at the $\alpha$ polymorph, it is clear that all the predicted volumes are too small. It is well known that $\Gamma-$point phonons do not achieve sufficient sampling of the

low-frequency modes which drives the thermal expansion of the crystal at high temperatures. Accounting for phonon dispersion has been shown previously to help address some of the systematic underexpansion that is seen in modeling small molecule crystals.[106,107] Typically this is addressed by calculating a supercell of the molecular crystal which increases the amount of reciprocal space that is evaluated directly. Unfortunately these calculations are computationally intensive and convergence of the acoustic modes can be challenging to achieve. Nonetheless, MP2C/CBS+pHF clearly has the best agreement with the experimental volumes with the largest error only at 0.9%. Volumes computed at the B86bPBE-XDM level (i.e. no correction to the potential energy surface) also performs acceptably well with a maximum error only 1.5% relative to experiment. Surprisingly, MP2C/CBS+AMOEBA performs the worst with a maximum over 2.7% error in the predicted volumes.

For the $\beta$-polymorph both DFT and MP2C/CBS+pHF predict volumes that are too large, while MP2C/CBS+AMOEBA again systematically underestimates molar volume. This leads to maximum errors in the predicted volumes of 2.1%, 0.7%, and 2.0% for DFT and MP2C/CBS+pHF, and MP2C/CBS+AMOEBA respectively. While the MP2C/CBS+AMOEBA volumes would benefit from better sampling of the phonon modes, it is likely that the DFT and MP2C/CBS+pHF would deviate further away from experiment.

Overall, for both polymorphs the volumes calculated at the MP2C/CBS+pHF level achieve the closest agreement with experimental volumes. Achieving 0.7%-0.9% accuracy with experiment is on par with previous studies which employed coupled cluster theory calculations on small molecules.[105,106] However, those calculated at the MP2C/CBS+AMOEBA

level perform poorly. Given that the only difference in these potential energy surfaces is the many-body contribution this could indicate a failure to capture some of the higher-order interactions accurately. While the shape of the expansion is qualitatively correct for both polymorphs the narrow temperature window makes it difficult to say if this was fortuitous or not.

Next we investigate the quality of the predicted volumes at a fixed temperature. Figure 5.2 maps out the predicted volumes for both polymorphs in a pressure window from 0 to 5 GPa with the temperature fixed at 300 K. Experimental values are pulled from reference 292 and we compare our predictions against a previous study done by Drużbicki et al (Ref 295) which simulated the volumes using planewave PBE-TS. For the work done by Drużbicki only harmonic corrections were employed, hence their structures can be thought of as being simulated at 0 K.

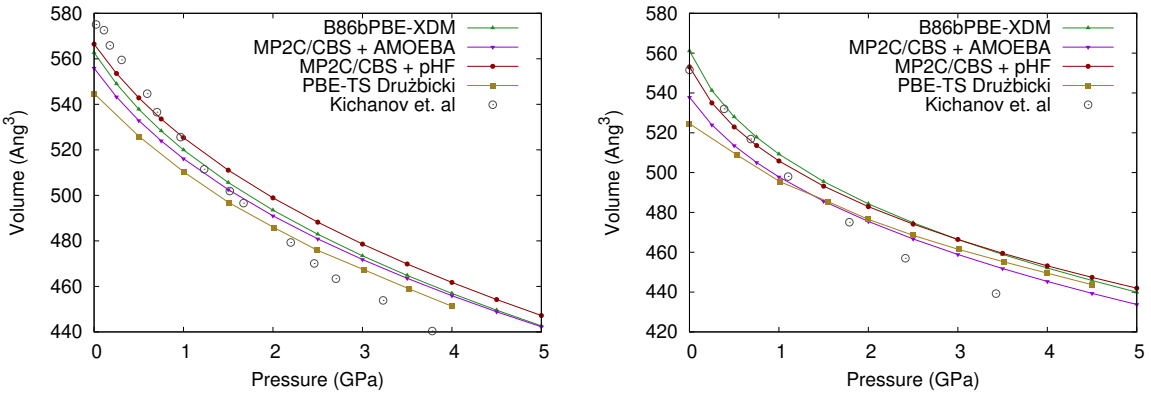



Figure 5.2: Equation of state curves for a) $\alpha$-resorcinol and b) $\beta$-resorcinol.

Looking first at the overall trends for both polymorphs in the low pressure regime (0-1 GPa), once again the volumes calculated at the MP2C/CBS+pHF level achieve the

best agreement with experiment with both polymorphs predicted to be only slightly underestimated. The volumes calculated at the B86bPBE-XDM level also perform well, despite the inconsistent prediction with the $\alpha$ polymorph too small and the $\beta$ polymorph too large. The volumes predicted at the MP2C/CBS+AMOEBA level are underestimated with the $\beta$ polymorph contracted by almost 15 Å$^3$ at 0 GPa. While these volumes are no worse than the volumes previously predicted using PBE-TS, the inclusion of thermal expansion typically yields better performance than the harmonic approximation.

Beyond the 0-1 GPa pressure regime all predicted volumes eventually become too large. Furthermore, near 4 GPa it is clear that all predicted volumes, regardless of the method, begin to converge to the same value. This is consistent with previous results showing that inclusion of thermal expansion matters less in the high-pressure regimes.[105] Given how consistent the volume errors are at high pressures, this is either an inherent limitation of the method or the experimental equation of state curves need to be revisited. Overall none of the current methods are able to qualitatively preserve the shape of the equation of state curves. It is possible that inclusion of phonon dispersion could help correct the volumes for both polymorphs at low pressures. Nonetheless all three of our current methods achieve better agreement with experiment in the 0-1 GPa than PBE-TS.

Given that the phonon contributions are computed solely at the DFT level, the primary differences in the predicted volumes shown here come from the electronic energy surfaces. It is expected that the different potential energy surfaces will shift the electronic minima and, at times, affect the curvature of the expansion and compression branches. Figure C.1 plots the electronic energy surface for both polymorphs at all simulated levels of

theory. Poorly sampled electronic energy surfaces can introduce noise when attempting to predict the minimum volume on each Gibbs surface. For this reason a fine grid of volumes was sampled for both polymorphs to ensure that the free energy basin is free of artifacts. In Appendix C.3, we show for both polymorphs that the Gibbs free energy surface is smooth and continuous over a wide range of pressures and temperatures.

### 5.4.2 Predicted thermodynamic properties

Having seen that we have reasonable agreement in our molar-volume predictions at low pressures, we now investigate the predicted phase-transition in the 0-1 GPa regime. In Figure 5.3 we plot predicted phase transition temperatures from 0 to 1 GPa at various levels of theory. Our predicted phase-transition line is compared against experimental data pulled from reference 292. The phase transition was produced by monitoring the proton spin-lattice relaxation time using Free Induction Decay (FID) amplitudes. In the experiment they report the existence of a mixed state where the $\beta-$polymorph nucleates in the $\alpha-$polymorph due to a temperature or pressure induced transition. Eventually this seed spreads throughout the rest of the crystal to complete the phase transition. However, the mixed state is somewhat long-lived with a mixed $\alpha + \beta$ state existing for 120 minutes at 315 K. Note that our prediction is purely a thermodynamic transition and hence will always exhibit a sharp phase boundary. Prediction of the experimentally observed mixed state would require accounting for the kinetics of the phase transition.

Surprisingly, despite consistently giving the best prediction of the molar volumes, MP2C/CBS+pHF performs poorly, overestimating the phase transition temperature at 0

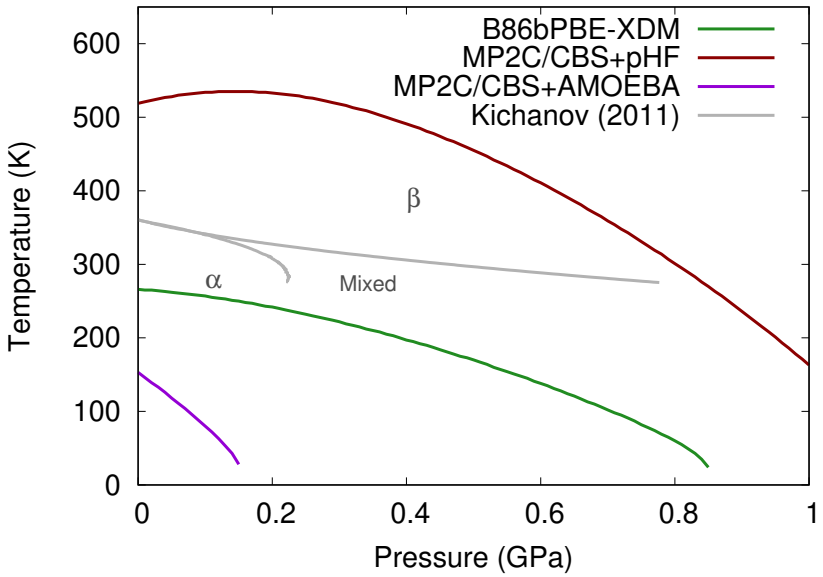



Figure 5.3: Predicted phase-transition between $\alpha-$ and $\beta$-resorcinol.

GPa by 150 K. The predicted phase-transition temperature remains surprisingly high at all pressures, indicating that the free energy of the $\alpha$ polymorph is over-stabilized relative to the $\beta$-polymorph. The predicted phase transition for B86bPBE-XDM is 50 K closer to experiment with a predicted phase transition temperature of 266 K relative to Ebisuzaki's 369 K at 0 GPa. Unsurprisingly, MP2C/CBS+AMOEBA gives the worst phase-transition curve with the $\beta$ polymorph grossly over-stabilized, causing the phase transition curve to only exist between 0 and 0.15 GPa. As there have been $\alpha$ to $\beta$ transitions reported up to 3.2 GPa, this is prediction is clearly inaccurate.

A closer examination of the thermodynamics of transition adds further confusion to this story (see Table 5.1). Looking first at the predicted enthalpy of transition, MP2C/CBS + AMOEBA surprisingly performs the best relative to experiment despite

Table 5.1: Thermodynamic data for the $\alpha$ to $\beta$ phase transition of resorcinol

| Method | $\Delta H_{\alpha-\beta}$ (kJ/mol) | $\Delta S_{\alpha-\beta}$ (J/(mol K)) | Temp. (K) |
|---|---|---|---|
| Ebisuzaki et.al.[291] | 1.370 ± 0.007 | 3.71 ± 0.05 | 369 ± 6 |
| B86bPBE-XDM | 2.353 | 8.82 | 266 |
| MP2C/CBS + pHF | 7.756 | 14.93 | 519 |
| MP2C/CBS + AMOEBA | 0.737 | 4.80 | 153 |

having the worst agreement in the transition temperature. B86bPBE-XDM also provides a reasonable enthalpy of transition with only 1.0 kJ/mol error relative to experiment. Unfortunately, MP2C/CBS+pHF appears to be surprisingly wrong with over 6 kJ/mol error. Given the over-stabilization of the $\alpha$ phase that is seen using this method some error was expected. However, previous studies have shown MP2C at the complete-basis set limit normally gives energy descriptions that are on par with coupled cluster benchmarks.[301] In the methanol paper Červinka and Beran reported an error of only 0.23 kJ/mol in the $\alpha \rightarrow \beta$ enthalpy of transition relative to experiment using benchmark calculations.[109] A recent study on acetamidomenzamide using MP2C/CBS+pHF also reported accurate predicted enthalpies of transitions for the $\beta \rightarrow \alpha$ transition (-2.4 kJ/mol, vs -1.9 and -2.9 experimentally).[302] Given that acetamidomenzamide is a far more complex system involving intra- and inter-molecular interactions, this error in the MP2C/CBS+pHF enthalpy of transition is disconcertingly large.

Looking next at the predicted entropy of transition, MP2C/CBS+AMOEBA again significantly outperforms the other methods with only a 1 J/(mol K) error relative to experiment. B86bPBE-XDM doubles this error and MP2C/CBS+pHF more than quadruples it. Since the entropy of transition is governed by harmonic phonons it is expected that this value can only feasibly be improved by better phonon sampling or a drastic change in the

predicted volumes. However, given that the entropy of transition at ambient temperatures (T*dS) is ∼0.3 kJ/mol and chemical accuracy is considered to be 2 kJ/mol, this value is well below what most methods report confidence in, hence any agreement with experiment here can be thought of as fortuitous.

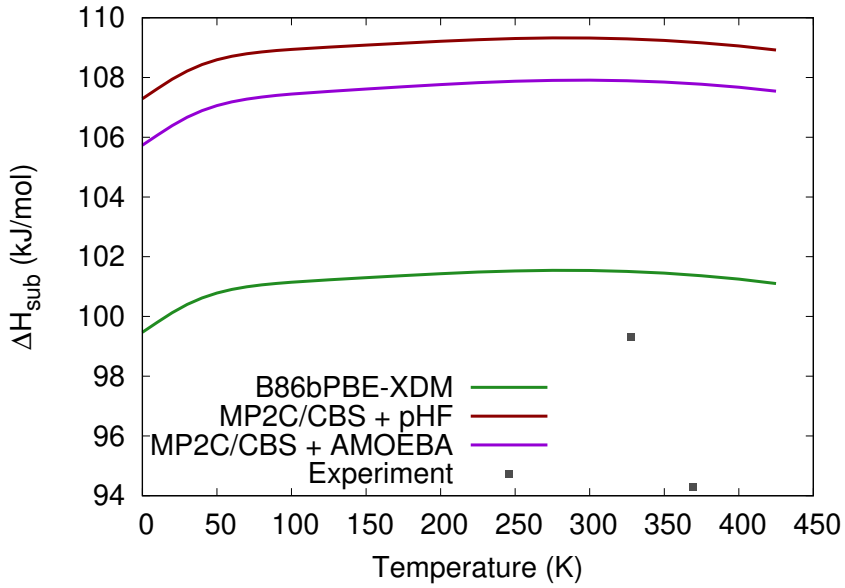



Figure 5.4: Predicted enthalpy of sublimation for $\alpha-$resorcinol.

Finally, enthalpies of sublimation were computed for $\alpha-$resorcinol using the standard statistical mechanical expression (see Figure 5.4). See Ref 105 for details. Experimental values are taken from references 297 and 296. Standard ideal gas, rigid rotor, and harmonic oscillator partition function expressions were used to evaluate thermochemical contributions for the gas phase species.

The sublimation enthalpy is derived from $\Delta H_{sub} = H_{gas} - H_{crystal}$. Upon increasing temperature the sublimation enthalpy typically first increases before reaching a

maximum and then decreasing. At low temperatures, little thermal expansion occurs in the crystal and the $\Delta H_{sub}$ increases due primarily to the gas contribution in that temperature regime. At higher temperatures, the larger thermal expansion in the crystal eventually leads to the $H_{crystal}$ becoming larger than $H_{gas}$, which produces the maximum and subsequent decrease in $\Delta H_{sub}$. B86bPBE-XDM once again performs consistently well, achieving almost 3 kJ/mol error relative to the more recent experimental value found by Goncalves et al.[296] The higher errors seen for MP2C/CBS+pHF and MP2C/CBS+AMOEBA indicate that the crystal form is being overstabilized relative to the gas phase hence requiring more energy to force the sublimation to occur.

Overall, B86bPBE-XDM performs consistently better on predicted thermodynamic properties than MP2C/CBS+pHF despite having a poorer prediction in the molar volumes. One obvious question is whether this could be a breakdown of the tiered-QHA method. Put another way, does the tiered-QHA give good prediction of volumes but not energies due to the mixing of methods. Previous predictions on small molecules using the hybrid-Tier 3 method have shown that enthalpies of sublimation can be predicted within 1-2 kJ/mol relative to both experiment. Furthermore, direct comparison of calculating the QHA using pure MP2/aTZ+AMOEBA (Tier 1) with the DFT structures SPE-corrected up to the same level (Tier 3) showed almost perfect agreement between the methods for both volumes and enthalpies of sublimation making this option unlikely.[107] Instead it seems more likely that there exists a problem in the MP2C/CBS+pHF electronic energy surface.

### 5.4.3 Sensitivity of the MP2C/CBS+pHF phase-transition line

Given that MP2C/CBS+pHF gave the best agreement in the predicted molar volumes, it is somewhat surprising to see it perform so poorly on the predicted phase transition temperatures. To examine the cause of this further we plot the relative change in the Gibbs free energies and enthalpies for both the $\alpha$ and $\beta$ polymorph at 0.0, 0.5, and 1.0 GPa (see Figure 5.5). All energies are plotted relative to the $\alpha$ polymorph.

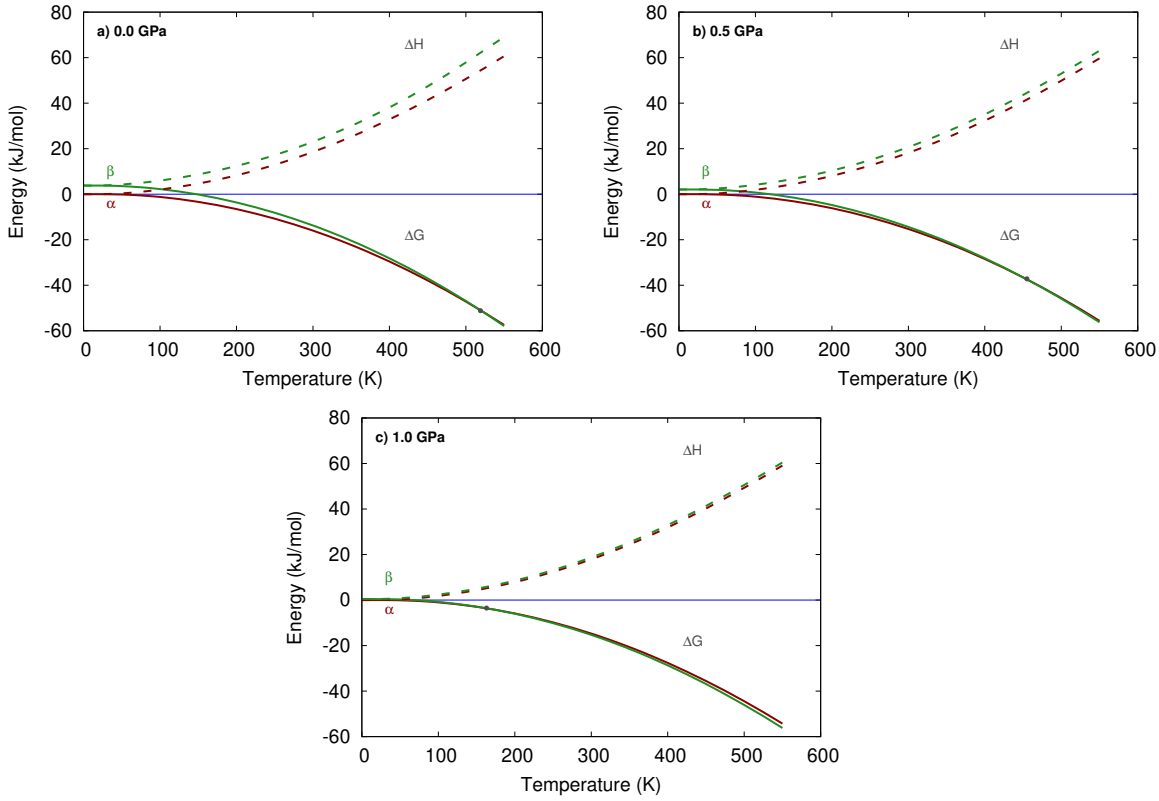



Figure 5.5: Relative free energies at MP2C/CBS+pHF for $\alpha-$ and $\beta$-resorcinol at a) 0.0 GPa, b) 0.5 GPa, and c) 1.0 GPa.

At 0 GPa (approximately ambient pressure) there is a large offset of 3.79 kJ/mol between the Gibbs free energy of the $\alpha$ and $\beta$ polymorphs. The enthalpies also appear to

have a large offset which grows increasingly larger as the temperature increases. The enthalpies never cross, indicating that it is an entropy-driven phase transition. This explains the higher temperatures that are needed to overcome both the initial 0 K enthalpy differences and the branching enthalpies at higher temperatures. At 0.5 GPa the offset in the free energies between the two polymorphs narrows to 2.08 kJ/mol lowering the predicted transition temperature from 519 K to 455 K. By 1 GPa there is already almost no difference in the Gibbs free energy (0.44 kJ/mol) at low temperatures leading to an early transition temperature of 163 K. The enthalpies are also almost degenerate over the entire temperature range although they do branch away from each other slightly at higher temperatures.

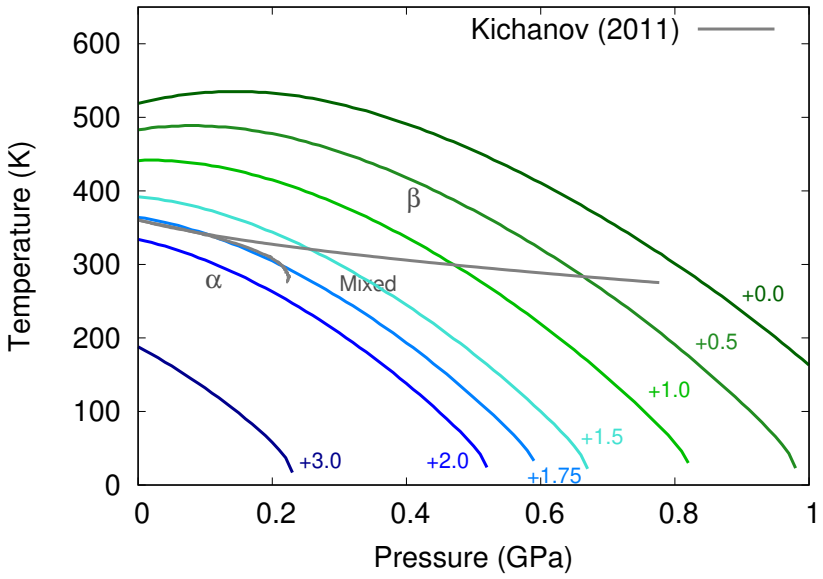



Figure 5.6: Sensitivity of the predicted phase-transition between $\alpha-$ and $\beta$-resorcinol calculated at the MP2C/CBS+pHF. Indicated values are the amount by which the free energy of the $\alpha$ polymorph is destabilized.

Given these results it is clear that a more accurate phase transition temperature would be obtained if the initial 0 K enthalpy difference were smaller. To estimate the level

of accuracy required to correctly resolve this phase transition we can examine the sensitivity of the phase transition. In Figure 5.6 we examine the effect of artificially destabilizing the $\alpha$-polymorph and the effect this has on the predicted phase transition temperature. We find that by destabilizing the $\alpha$ polymorph by a mere 1.75 kJ/mol we are able to achieve almost perfect agreement with the experimental prediction! While this is most-likely achievable by inclusion of more high-level electronic structure theory, pursuing this route can be cost-prohibitive.

## 5.5 Future directions

Unfortunately resorcinol is a challenging system to model as was first noted by Zhu et al.[298] While the results presented here represent a good starting point, inconsistencies in which methods predict accurate volumes and those that predict accurate thermodynamic properties relative to known experimental values indicate that more work is needed. DFT has consistently performed acceptably well, predicting the enthalpy of transition, sublimation enthalpies, and the phase boundary all with reasonable errors relative to experiment. However, this performance is somewhat fortuitous, with consistent volume errors indicating a deeper problem in prediction. Furthermore, benchmark dimer calculations show that MP2C performs considerably better than B86bPBE-XDM (by  2 kJ/mol for one key dimer).

Single-point energy correcting the electronic energy surface up to MP2C/CBS+pHF gives a better prediction of the molar volumes but a poorer prediction of thermodynamic properties. The question remains as to what must be done to further correct the electronic energy surface in the post-DFT models. As the sensitivity analysis in Figure 5.6 has

shown, accurate prediction of the phase transition temperatures requires a method capable of resolving <2 kJ/mol errors in the free energy surface.

Ideally one would simply calculate the electronic energy surface using coupled cluster theory with perturbative triples (CCSD(T)) at the complete basis set limit as this is currently the gold standard in modeling. However, this is not practical given that the computational cost of these calculations scale steeply with system size, and the large number of dimers (5,000+) that would need to be calculated this is undesirable. Instead, predicting the dimers using MP2.5 (the average of MP2 and MP3) may be more computationally tractable as this has been found to provide a more balanced treatment of both many-body polarization and dispersion, reproducing CCSD(T) benchmarks faithfully.[262,303,304]

As the many-body correction seemed to play a significant role in the predicted phase transition temperature, further correction of the many-body contribution may be necessary (i.e. either Hartree-Fock and/or the pob-TZVP basis set may not adequately capture the many-body interactions). One way to address this issue is to forcibly expand the amount of interactions that are calculated at the high-level energy surface by including trimer interactions. Podeszwa et al used SAPT-DFT to predict the lattice-energy of crystalline benzene and found that 3-body dispersion interactions contributed ∼3 kJ/mol to the lattice energies.[305] As resorcinol is structurally similar to benzene, it is expected that this could also be a relevant correction for our system. Work investigating the role of 3-body dispersion interactions is on-going. Initial estimates suggest that the missing 3-body dispersion contributions could account for much of the remaining discrepancy between the predicted MP2C/CBS+pHF and experimental sublimation enthalpies. The DFT calcula-

tions we performed here neglect these contributions, lending further evidence that there is something fortuitous with the B86bPBE-XDM predictions.

Entropy is expected to play a large role in the predicted phase transition properties, but the low-frequency phonons which contribute most to this term are currently under-sampled as we have only computed harmonic phonons for the central unit cell. As the entropic contribution to the Helmoholtz vibrational energy drives the thermal expansion at high-temperatures this can also help address some of the slightly smaller volumes that are seen for the $\alpha-$polymorph. Typically, one accounts for phonon dispersion by performing a lattice dynamics calculation on a molecular crystal supercell which increases the amount of reciprocal space that is evaluated directly. However, the acoustic modes tend to become imaginary bands if a supercell of insufficient size is computed. As the relative free energy differences at finite temperatures are governed primarily by low frequency modes it is essential to converge these acoustic bands.

Hoja and Tkatchenko had previously found that a supercell of at least 15 Å in each direction is required to resolve all imaginary bands.[306] We attempted the direct-supercell approach for supercells of size 2x2x2 and 2x2x3 representing a supercell with at least 10 Å and 15 Å. Unfortunately despite the larger size a number of imaginary modes remain. While it is possible that computing a 3x3x3 supercell would converge these imaginary modes, a supercell of this size pushes the limit of what can reasonably be modeled and hence has not been attempted in this study.

Predicting the phase diagram of any system is challenging. There are a number of factors that go into the prediction (geometries, phonons, electronic energies) and there are

small tolerances for errors in prediction. The successful prediction of the methanol phase diagrams required that all these factors come together to give with less than 0.5 kJ/mol error in the relative free energies. Sensitivity analysis showed that the predicted phase transition boundaries showed drastic changes with small scaling of the phonons.[109] Our fragment-based calculations have clearly improved key parts of the modeling relative to the DFT calculations, but in doing so, it appears that some of the error cancellation has been disrupted. This leads to a worse prediction of the sublimation enthalpy phase transition temperature.

# Chapter 6

# Conclusions

In summary we have demonstrated how crystal structure prediction methods can be used to support arguments that carbon dioxide phases III and VII are actually the same phase. The crystal structure prediction USPEX was able to successfully generate phases I, II, and VII, along with 22 other potential structures within 10 kJ/mol of the most stable one. However, none of the other structures has a simulated powder X-ray diffraction pattern that is plausibly consistent with experimental phase III. In fact, the best match amongst the generated structures came from the phase VII one. While the possibility that the crystal structure prediction search simply failed to generate phase III cannot be ruled out, the absence of other viable phase III structures combined with other predicted structural and spectroscopic data suggest that phase III and VII are in fact the same.

We have also demonstrated how the quasi-harmonic approximation (QHA) can be used to refine isotropic nuclear magnetic resonance (NMR) chemical shift predictions for molecular crystals. Employing the QHA recovers most or all of the chemical shift error

accrued by neglecting the thermal expansion of the crystal. Indeed, the accuracy of chemical shifts computed using quasi-harmonic room-temperature structures rivals what is obtained for shifts computed with the experimental unit cell parameters. However, the impact of QHA thermal expansion on chemical shifts can be quite variable. No clear relationship between the quality of the structural agreement and the chemical shift errors was found, nor does the magnitude of the change of the chemical shifts upon thermal expansion seem to correlate strongly with the amount of thermal expansion that occurs. Certain functional groups appeared to benefit more from the QHA treatment. In particular, functional groups which exhibit less dynamic flexibility (such are aromatic or carboxylate groups) tend to benefit more from the QHA treatment. While it is difficult to know *a priori* what effect employing QHA thermal expansion can have, it has been shown in select cases to increase the discrimination between candidate structures that were generated via crystal structure prediction and can even alter how one might assign predicted chemical shifts.

We have also introduced a tiered-QHA which compared a hierarchy of models ranging from pure DFT to pure MP2. We have demonstrated that introducing correlated wavefunction energies on top of DFT geometries and phonons (Tier 3) can appreciably improve predicted molar volumes and enthalpies. One must ensure, however, that the single-point energies are well-chosen, which in practice means employing large basis sets and accounting for any deficiencies in the models (e.g. by replacing MP2 with MP2C or MP2D[115] for the monomer and dimer contributions or AMOEBA with HF for the many-body contributions). In the small molecule crystals, the hybrid Tier 3 approach led to predicting molar volumes to within a few percent and sublimation enthalpies to within 1–2

kJ/mol of experiment over broad temperature ranges.

Finally, we have attempted to employ this hybrid Tier 3 approach to simulate the phase-transition properties of $\alpha$ and $\beta-$resorcinol. We report the most accurate modeling to date (to our knowledge) of the unit cell volumes for both polymorphs at low pressures. However, the thermodynamic properties are inconsistently modeled using this method, indicating that further corrections are required. In order to achieve almost perfect agreement with the experimental prediction, the phase transition requires a sub-2 kJ/mol energy correction, indicating that further refinement of the models will be required. Correction of the method to include trimer interactions and potentially CCSD(T) dimer calculations will most likely be needed to fully converge the electronic energy surface.

The hybrid approaches that combine DFT and wavefunction techniques do not appreciably improve the predicted entropies, since those are largely governed by phonon contributions. We also found that the hybrid calculations generally underestimate the amount of thermal expansion. Accounting for phonon dispersion has been shown previously to help address some of the systematic underexpansion that is seen in modeling small molecule crystals. Typically this is addressed by calculating a supercell of the molecular crystal which increases the amount of reciprocal space that is evaluated directly. However, the acoustic modes tend to become imaginary bands if a supercell of insufficient size is computed. As the relative free energy differences at finite temperatures are governed primarily by low frequency modes it is essential to converge these acoustic bands. Hoja and Tkatchenko had previously found that a supercell of at least 15 Å in each direction is required to resolve all imaginary bands.[306] Especially for electronic structure methods a

supercell of this size pushes the limit of what can reasonably be modeled.

Occasionally it becomes necessary to evaluate effects beyond the quasi-harmonic approximation. For example, anharmonicity and dynamic sampling of the system may also play an important role in accurate thermodynamical modeling of these crystals. However, efficient and effective strategies for addressing anharmonicity are difficult to come by. Hoja and Tkatchenko have previously estimated anharmonic vibrational free energies by substituting the harmonic oscillators with Morse oscillators which changed the relative stabilities by 1 kJ/mol on average.[306] More accurate anharmonic free energies can be obtained by using vibrational self-consistent field approach[307–309] or path-integral molecular dynamics.[222] As the computational cost for evaluating these effects grows substantially with system size it is recommend to evaluate said effects only when it is necessary to resolve sub-kJ/mol energy differences or when attempting to accurately model nuclear quantum effects.

Overall, the new tiered-QHA method has demonstrated excellent performance for predicting the molar volumes of each crystal at any desired level of theory at a fraction of the cost.[107] This opens up a number of exciting applications in the field of Raman and solid state NMR studies. While there is still the question as to how well it can predict thermodynamic properties, the problems with resorcinol appear stem from not predicting higher order interactions at the quantum mechanical level rather than a problem with the tiered method. Provided these problems can be resolved, this model should be applied to systems where the full phase diagram is unknown.

One future application of this method could be to conduct a crystal structure prediction study and use this method to predict the thermodynamic stability of each structure.

While one such attempt has already been done on the stable polymorphs of nitrogen, application to pharmaceutically relevant molecules would be particularly interesting.[310] Perhaps another future application could be participation in a future crystal blind test. One application that should be explored is attempting to simulate systems that are expected to break this tiered-QHA method in order to establish clear boundaries of where and where not to apply this method. For example, an application on molecules that exhibit large amounts of intra-molecular bonding interactions and/or conformational flexibility are expected to break this model.

Like all methods there are areas where improvement is needed. Currently, the electronic potential energy surface is generated by optimizing the reference structure under isotropic pressure to allow the unit cell volume to deform naturally. However, as it is difficult to know *a priori* what volume each pressure will give generate this often requires a number of extra calculations which can be wasteful. Future work implementing constant volume optimization techniques (like the ones that exist in the VASP program) would go a long way towards addressing this problem.[311]

Evaluation of harmonic phonons beyond the $\Gamma-$point is a clear challenge. Currently, evaluation of supercells is extremely RAM intensive, making this calculation a significant bottleneck in all future endeavors. One possibility that should be explored is the use of more approximate levels of theory to compute phonons such as density-functional tight binding (DFTB). While some attention to detail may be needed to ensure a smooth transition from the phonons computed using DFT and those computed using DFTB, provided such a method was developed, evaluation of extended supercells at the DFTB level

would likely still be cheaper than the evaluation of the $\Gamma-$point phonons at the DFT level. However, whether such techniques can retain the sub-kJ/mol accuracy required for phase diagram prediction remains to be seen.

# Bibliography


[1] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni, J. Phys. Condens. Mat. **29**, 465901 (2017).

[2] A. D. Becke, J. Chem. Phys. **38**, 7184 (1986).

[3] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[4] A. Otero-De-La-Roza and E. R. Johnson, J. Chem. Phys. 174109 (2012).

[5] J. A. Chisholm and W. D. S. Motherwell, J. Appl. Crystall. **38**, 228 (2005).

[6] H. Olijnyk and A. P. Jephcoat, Phys. Rev. B **57**, 879 (1998).

[7] F. Datchi, B. Mallick, A. Salamat, G. Rousse, S. Ninet, G. Garbarino, P. Bouvier, and M. Mezouar, Phys. Rev. B **89**, 144101 (2014).

[8] F. Datchi, V. M. Giordano, P. Munsch, and A. M. Saitta, Phys. Rev. Lett. **103**, 185701 (2009).

[9] V. M. Giordano and F. Datchi, Europhys. Lett. **77**, 46002 (2007).

[10] K. Aoki, H. Yamawaki, M. Sakashita, Y. Gotoh, and K. Takemura, Science **263**, 356 (1994).

[11] R. Boese, B. D, R. Latz, and A. Bäumen, Z. Kristallogr. **215**, 693 (2000).

[12] J. D. Hartman, A. Balaji, and G. J. Beran, J. Chem. Theory Comput. **13**, 6043 (2017).

[13] M. Vasileiadis, A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, and C. C. Pantelides, Acta Crystallographica Section B Structural Science **68**, 677 (2012).

[14] S. P. Thomas and M. A. Spackman, Australian Journal of Chemistry **71**, 279 (2018).

[15] M. Tan, A. G. Shtukenberg, S. Zhu, W. Xu, E. Dooryhee, S. M. Nichols, M. D. Ward, B. Kahr, and Q. Zhu, Faraday Discussions **211**, 477 (2018).

[16] J. Nyman, L. Yu, and S. M. Reutzel-Edens, CrystEngComm **21**, 2080 (2019).

[17] A. J. Cruz-Cabeza and J. Bernstein, Chemical Reviews **114**, 2170 (2014).

[18] M. Habgood, I. J. Sugden, A. V. Kazantsev, C. S. Adjiman, and C. C. Pantelides, Journal of Chemical Theory and Computation **11**, 1957 (2015).

[19] J. D. Dunitz and A. Gavezzotti, Crystal Growth & Design **5**, 2180 (2005).

[20] T. Li, P. W. Ayers, S. Liu, M. J. Swadley, and C. Aubrey-Medendorp, Chemistry - A European Journal **15**, 361 (2009).

[21] E. H. Lee, Asian Journal of Pharmaceutical Sciences **9**, 163 (2014).

[22] S. L. Price, Faraday Discussions **211**, 9 (2018).

[23] D.-K. Bučar, R. W. Lancaster, and J. Bernstein, Angewandte Chemie International Edition **54**, 6972 (2015).

[24] D.-K. Bučar, G. M. Day, I. Halasz, G. G. Z. Zhang, J. R. G. Sander, D. G. Reid, L. R. MacGillivray, M. J. Duer, and W. Jones, Chemical Science **4**, 4417 (2013).

[25] R. Tandon, N. Tandon, and R. K. Thapar, Pharmaceutical Patent Analyst **7**, 59 (2018).

[26] W. Sontising, Y. N. Heit, J. L. McKinley, and G. J. O. Beran, Chem. Sci. **8**, 7374 (2017).

[27] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris, Pharmaceutical Research **18**, 859 (2001).

[28] G. Hasegawa, T. Komasaka, R. Bando, Y. Yoshihashi, E. Yonemochi, K. Fujii, H. Uekusa, and K. Terada, International Journal of Pharmaceutics **369**, 12 (2009).

[29] T. Stolar, S. Lukin, J. Požar, M. Rubčić, G. M. Day, I. Biljan, D. Š. Jung, G. Horvat, K. Užarević, E. Meštrović, and I. Halasz, Crystal Growth & Design **16**, 3262 (2016).

[30] A. L. Grzesiak, M. Lang, K. Kim, and A. J. Matzger, Journal of Pharmaceutical Sciences **92**, 2260 (2003).

[31] M. Sacchetti, Journal of Thermal Analysis and Calorimetry **63**, 345 (2001).

[32] S. Cherukuvada, R. Thakuria, and A. Nangia, Crystal Growth & Design **10**, 3931 (2010).

[33] M. Yoshino, K. Takahashi, Y. Okuda, T. Yoshizawa, N. Fukushima, and M. Naoki, The Journal of Physical Chemistry A **103**, 2775 (1999).

[34] A. Cesaro and G. Starec, The Journal of Physical Chemistry **84**, 1345 (1980).

[35] A. Torrisi, C. K. Leech, K. Shankland, W. I. F. David, R. M. Ibberson, J. Benet-Buchholz, R. Boese, M. Leslie, C. R. A. Catlow, and S. L. Price, The Journal of Physical Chemistry B **112**, 3746 (2008).

[36] E. Badea, I. Blanco, and G. Della Gatta, The Journal of Chemical Thermodynamics **39**, 1392 (2007).

[37] X. Alcobé, E. Estop, A. E. Aliev, K. D. Harris, J. Rodrguez-Carvajal, and J. Rius, Journal of Solid State Chemistry **110**, 20 (1994).

[38] F. P. A. Fabbiani and C. R. Pulham, Chemical Society Reviews **35**, 932 (2006).

[39] V. V. Boldyrev, Journal of Materials Science **39**, 5117 (2004).

[40] F. Cansell, D. Fabre, and J. Petitet, The Journal of Chemical Physics **99**, 7300 (1993).

[41] Y. V. Seryotkin, T. N. Drebushchak, and E. V. Boldyreva, Acta Crystallographica Section B Structural Science, Crystal Engineering and Materials **69**, 77 (2013).

[42] S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, S. Spanton, W. Dziki, W. Porter, J. Quick, P. Bauer, J. Donaubauer, B. A. Narayanan, M. Soldani, D. Riley, and K. McFarland, Organic Process Research & Development **4**, 413 (2000).

[43] S. L. Morissette, S. Soukasene, D. Levinson, M. J. Cima, and O. Almarsson, Proceedings of the National Academy of Sciences **100**, 2180 (2003).

[44] G. J. O. Beran, Angewandte Chemie International Edition n/a (2014).

[45] A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein, Chemical Society Reviews **44**, 8619 (2015).

[46] J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schutz, and G. K.-L. Chan, Science **345**, 640 (2014).

[47] S. L. Price, Chem. Soc. Rev. **43**, 2098 (2014).

[48] M. A. Neumann, F. J. J. Leusen, and J. Kendrick, Angewandte Chemie International Edition **47**, 2427 (2008).

[49] B. P. van Eijck and J. Kroon, Acta Crystallographica Section B Structural Science **56**, 535 (2000).

[50] G. M. Day, W. D. S. Motherwell, and W. Jones, Physical Chemistry Chemical Physics **9**, 1693 (2007).

[51] M. A. Neumann and J. van de Streek, Faraday Discussions **211**, 441 (2018).

[52] G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. K. Wolf, and B. Schweizer, Acta Cryst. B **65**, 107 (2009).

[53] D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. M. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. J. J. Leusen, A. V. Maleev, A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti, and I. K. Zhitkov, Acta Cryst. B **67**, 535 (2011).

[54] A. M. Reilly *et al.*, Acta Cryst. B **72**, 439 (2016).

[55] C. J. Pickard and R. J. Needs, Journal of Physics: Condensed Matter **23**, 053201 (2011).

[56] C. Collins, G. R. Darling, and M. J. Rosseinsky, Faraday Discussions **211**, 117 (2018).

[57] A. R. Oganov, A. O. Lyakhov, and M. Valle, Acc. Chem. Res. **44**, 227 (2011).

[58] F. Curtis, T. Rose, and N. Marom, Faraday Discussions **211**, 61 (2018).

[59] Q. Tong, L. Xue, J. Lv, Y. Wang, and Y. Ma, Faraday Discussions **211**, 31 (2018).

[60] V. L. Deringer, D. M. Proserpio, G. Csányi, and C. J. Pickard, Faraday Discussions **211**, 45 (2018).

[61] L. Iuzzolino, P. McCabe, S. L. Price, and J. G. Brandenburg, Faraday Disc. **211**, 275 (2018).

[62] J. Nyman and G. M. Day, CrystEngComm **17**, 5154 (2015).

[63] J. Nyman and G. M. Day, Phys. Chem. Chem. Phys. **18**, 31132 (2016).

[64] G. J. O. Beran, CrystEngComm **21**, 758 (2019).

[65] W. Kohn and L. J. Sham, Physical Review **140**, A1133 (1965).

[66] F. M. Bickelhaupt and E. J. Baerends, *Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry* (PUBLISHER, ADDRESS, 2007), pp. 1–86.

[67] V. Barone, M. Casarin, D. Forrer, M. Pavone, M. Sambi, and A. Vittadini, Journal of Computational Chemistry **30**, 934 (2009).

[68] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, Chem. Rev. **116**, 5105 (2016).

[69] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[70] E. Caldeweyher, C. Bannwarth, and S. Grimme, J. Chem. Phys. **147**, 034112 (2017).

[71] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).

[72] A. M. Reilly and A. Tkatchenko, Chemical Science **6**, 3289 (2015).

[73] R. A. DiStasio, O. A. von Lilienfeld, and A. Tkatchenko, Proc. Nat. Acad. Sci. **109**, 14791 (2012).

[74] A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, 236402 (2012).

[75] A. Ambrosetti, A. M. Reilly, R. A. DiStasio, and A. Tkatchenko, J. Chem. Phys. **140**, 18A508 (2014).

[76] A. M. Reilly and A. Tkatchenko, The Journal of Chemical Physics **139**, 024705 (2013).

[77] A. M. Reilly and A. Tkatchenko, The Journal of Physical Chemistry Letters **4**, 1028 (2013).

[78] A. Otero-de-la Roza and E. R. Johnson, The Journal of Chemical Physics **137**, 054103 (2012).

[79] N. Marom, R. A. DiStasio, V. Atalla, S. Levchenko, A. M. Reilly, J. R. Chelikowsky, L. Leiserowitz, and A. Tkatchenko, Angewandte Chemie International Edition **52**, 6629 (2013).

[80] A. G. Shtukenberg, Q. Zhu, D. J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk, A. Tkatchenko, A. R. Oganov, A. L. Rohl, M. E. Tuckerman, and B. Kahr, Chem. Sci. **8**, 4926 (2017).

[81] S. Hammes-Schiffer, Science (2017).

[82] K. A. Lyssenko, I. S. Bushmarinov, M. G. Medvedev, J. P. Perdew, and J. Sun, Science (2017).

[83] P. G. Karamertzanis, G. M. Day, G. W. A. Welch, J. Kendrick, F. J. J. Leusen, M. A. Neumann, and S. L. Price, The Journal of Chemical Physics **128**, 244708 (2008).

[84] J. P. Perdew, in *AIP Conference Proceedings* (AIP, ADDRESS, 2001), Vol. 577, pp. 1–20.

[85] J. Klimeš, D. R. Bowler, and A. Michaelides, Physical Review B **83**, 195131 (2011).

[86] L. Goerigk and S. Grimme, Physical Chemistry Chemical Physics **13**, 6670 (2011).

[87] Y. Zhao and D. G. Truhlar, Journal of Chemical Theory and Computation **1**, 415 (2005).

[88] L. M. LeBlanc, S. G. Dale, C. R. Taylor, A. D. Becke, G. M. Day, and E. R. Johnson, Angewandte Chemie International Edition **57**, 14906 (2018).

[89] P. Hobza and K. Muller-Dethlefs, *Non-Covalent Interactions*, *Theoretical and Computational Chemistry Series* (Royal Society of Chemistry, Cambridge, 2009), pp. P001–P226.

[90] P. Hobza and J. ezáč, Chemical Reviews **116**, 4911 (2016).

[91] C. D. Sherrill, T. Takatani, and E. G. Hohenstein, J. Phys. Chem. A **113**, 10146 (2009).

[92] A. E. DePrince and C. D. Sherrill, J. Chem. Theory Comput. **9**, 293 (2013).

[93] C. D. Sherrill, Acc. Chem. Res. **46**, 1020 (2013).

[94] G. J. Beran, J. Chem. Phys. **130**, 1 (2009).

[95] G. J. O. Beran and K. Nanda, J. Phys. Chem. Lett **1**, 3480 (2010).

[96] S. Wen and G. J. O. Beran, J. Chem. Theory Comput. **7**, 3733 (2011).

[97] K. Nanda and G. J. O. Beran, J. Chem. Phys. **137**, 174106 (2012).

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

[99] C. Müller and B. Paulus, Phys. Chem. Chem. Phys. **14**, 7605 (2012).

[100] G. J. O. Beran, Chem. Rev. **116**, 5567 (2016).

[101] S. Wen and G. J. O. Beran, J. Chem. Theory Comput. **7**, 3733 (2011).

[102] S. Wen and G. J. O. Beran, J. Chem. Theory Comput. **8**, 2698 (2012).

[103] G. J. O. Beran and K. Nanda, J. Phys. Chem. Lett. **1**, 3480 (2010).

[104] J. D. Hartman, R. A. Kudla, G. M. Day, L. J. Mueller, and G. J. O. Beran, Phys. Chem. Chem. Phys. **18**, 21686 (2016).

[105] Y. N. Heit, K. D. Nanda, and G. J. Beran, Chem. Sci. **7**, 246 (2016).

[106] Y. N. Heit and G. J. O. Beran, Acta Cryst. B **72**, 514 (2016).

[107] J. L. McKinley and G. J. O. Beran, Faraday Disc. **211**, 181 (2018).

[108] C. Červinka and G. J. O. Beran, Phys. Chem. Chem. Phys. **19**, 29940 (2017).

[109] C. Červinka and G. J. Beran, Chem. Sci. **9**, 4622 (2018).

[110] M. Born, K. Huang, and M. Lax, American Journal of Physics **23**, 474 (1955).

[111] S. A. Rivera, D. G. Allis, and B. S. Hudson, Cryst. Growth Des. **8**, 3905 (2008).

[112] N. Marom, R. A. DiStasio, V. Atalla, S. Levchenko, A. M. Reilly, J. R. Chelikowsky, L. Leiserowitz, and A. Tkatchenko, Angew. Chem. Int. Ed. **52**, 6629 (2013).

[113] A. M. Reilly and A. Tkatchenko, Phys. Rev. Lett. **113**, 055701 (2014).

[114] D. E. Braun, J. A. McMahon, L. H. Koztecki, S. L. Price, and S. M. Reutzel-Edens, Cryst. Growth Des. **14**, 2056 (2014).

[115] J. ezáč, C. Greenwell, and G. J. O. Beran, Journal of Chemical Theory and Computation **14**, 4711 (2018).

[116] P. Raiteri, R. Martonák, and M. Parrinello, Angew. Chem. Int. Ed. **44**, 3769 (2005).

[117] P. G. Karamertzanis, P. Raiteri, M. Parrinello, M. Leslie, and S. L. Price, J. Phys. Chem. B **112**, 4298 (2008).

[118] T. Zykova-Timan, P. Raiteri, and M. Parrinello, J. Phys. Chem. B **112**, 13231 (2008).

[119] A. M. Reilly, S. Habershon, C. A. Morrison, and D. W. H. Rankin, J. Chem. Phys. **132**, 134511 (2010).

[120] T.-Q. Yu and M. Tuckerman, Phys. Rev. Lett. **107**, 015701 (2011).

[121] E. Schneider, L. Vogt, and M. E. Tuckerman, Acta Cryst. B **72**, 542 (2016).

[122] M. J. Schnieders, J. Baltrusaitis, Y. Shi, G. Chattree, L. Zheng, W. Yang, and P. Ren, J. Chem. Theory Comput. **8**, 1721 (2012).

[123] I. J. Nessler, J. M. Litman, and M. J. Schnieders, Phys. Chem. Chem. Phys. 36 (2016).

[124] E. C. Dybeck, N. P. Schieber, and M. R. Shirts, J. Chem. Theory Comput. acs.jctc.6b00397 (2016).

[125] E. C. Dybeck, N. S. Abraham, N. P. Schieber, and M. R. Shirts, Crystal Growth and Design **17**, 1775 (2017).

[126] R. P. Stoffel, C. Wessel, M.-W. Lumey, and R. Dronskowski, Angew. Chem. Int. Ed. **49**, 5242 (2010).

[127] C. Červinka, M. Fulem, R. P. Stoffel, and R. Dronskowski, J. Phys. Chem. A **120**, 2022 (2016).

[128] C. Červinka and M. Fulem, J. Chem. Theory Comput. **13**, 2840 (2017).

[129] J. Hoja, A. M. Reilly, and A. Tkatchenko, WIREs Comput. Mol. Sci. **7**, e1294 (2017).

[130] S. R. Whittleton, A. Otero-de-la Roza, and E. R. Johnson, J. Chem. Theory Comput. **13**, 441 (2017).

[131] N. S. Abraham and M. R. Shirts, Journal of Chemical Theory and Computation **14**, 5904 (2018).

[132] A. Erba, J. Chem. Phys. **141**, 124115 (2014).

[133] E. Salager, G. M. Day, R. S. Stein, C. J. Pickard, B. Elena, and L. Emsley, J. Am. Chem. Soc. **132**, 2564 (2010).

[134] M. Baias, C. M. Widdifield, J. N. Dumez, H. P. Thompson, T. G. Cooper, E. Salager, S. Bassil, R. S. Stein, A. Lesage, G. M. Day, and L. Emsley, Phys. Chem. Chem. Phys. 8069 (2013).

[135] K. Kalakewich, R. Iuliucci, and J. K. Harper, Cryst. Growth Des. **13**, 5391 (2013).

[136] B. G. Caulkins, B. Bastin, C. Yang, T. J. Neubauer, R. P. Young, E. Hilario, Y.-m. M. Huang, C.-e. A. Chang, L. Fan, M. F. Dunn, M. J. Marsella, and L. J. Mueller, J. Am. Chem. Soc. **136**, 12824 (2014).

[137] M. Dračínský and P. Hodgkinson, RSC Advances **5**, 12300 (2015).

[138] C. J. Pickard and F. Mauri, Phys. Rev. B **63**, 245101 (2001).

[139] J. R. Yates, C. J. Pickard, and F. Mauri, Phys. Rev. B **76**, 024401 (2007).

[140] C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azaïs, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri, and C. J. Pickard, Chem. Rev. **112**, 5733 (2012).

[141] S. E. Ashbrook and D. McKay, Chem. Commun. **52**, 7186 (2016).

[142] J. D. Hartman and G. J. O. Beran, J. Chem. Theory Comput. **10**, 4862 (2014).

[143] J. D. Hartman, S. Monaco, B. Schatschneider, and G. J. O. Beran, J. Chem. Phys. **143**, 102809 (2015).

[144] J. D. Hartman, G. M. Day, and G. J. Beran, Cryst. Growth Des. **16**, 6479 (2016).

[145] D. Stueber, F. N. Guenneau, and D. M. Grant, J. Chem. Phys. (2001).

[146] D. Stueber, Conc. Magn. Resn. A (2006).

[147] D. Solís, M. B. Ferraro, and J. C. Facelli, J. Mol. Struct, (2002).

[148] C. M. Breneman and K. B. Wiberg, J. Comp. Chem. **11**, 361 (1990).

[149] A. J. Stone, Chem. Phys. Lett. (1981).

[150] A. Stone and M. Alderton, Mol. Phys. **56**, 1047 (1985).

[151] A. J. Stone, J. Chem. Theory Comput. (2005).

[152] P. P. Ewald, Annalen der Physik (1921).

[153] J. D. Hartman and G. J. Beran, Solid State Nuclear Magnetic Resonance **96**, 10 (2018).

[154] C. S. Yoo, Phys. Chem. Chem. Phys. **15**, 7949 (2013).

[155] V. Iota and C. S. Yoo, Phys. Rev. Lett. **86**, 5922 (2001).

[156] C. Yoo, H. Kohlmann, H. Cynn, M. Nicol, V. Iota, and T. LeBihan, Phys. Rev. B **65**, 1 (2002).

[157] J.-H. Park, C. S. Yoo, V. Iota, H. Cynn, M. F. Nicol, and T. Le Bihan, Phys. Rev. B **68**, 014107 (2003).

[158] S. A. Bonev, F. Gygi, T. Ogitsu, and G. Galli, Phys. Rev. Lett. **91**, 065501 (2003).

[159] S. Gohr, S. Grimme, T. Söhnel, B. Paulus, and P. Schwerdtfeger, J. Chem. Phys. **139**, 174501 (2013).

[160] C.-S. Yoo, A. Sengupta, and M. Kim, High Press. Res. **31**, 68 (2011).

[161] C. S. Yoo, H. Cynn, F. Gygi, G. Galli, V. Iota, M. Nicol, S. Carlson, D. Häusermann, and C. Mailhiot, Phys. Rev. Lett. **83**, 5527 (1999).

[162] A. M. Reilly *et al.*, Acta Cryst. B **72**, 439 (2016).

[163] A. R. Oganov, J. C. Schön, M. Jansen, S. M. Woodley, W. W. Tipton, and R. G. Hennig, in *Modern Methods of Crystal Structure Prediction*, edited by A. R. Oganov (Wiley, Weinheim, Germany, 2011), pp. 223–231.

[164] C. Martineau, J. Senker, and F. Taulelle, 1 (2014).

[165] M. D. Eddleston, K. E. Hejczyk, E. G. Bithell, G. M. Day, and W. Jones, Chem. Eur. J. **19**, 7874 (2013).

[166] S. Hirata, K. Gilliard, X. He, J. Li, and O. Sode, Acc. Chem. Res. **47**, 2721 (2014).

[167] F. Gygi, Comp. Mat. Sci. **10**, 63 (1998).

[168] J. Li, O. Sode, G. A. Voth, and S. Hirata, Nature Commun. **4**, 2647 (2013).

[169] J. Li, O. Sode, G. A. Voth, and S. Hirata, Nature Commun. **6**, 8907 (2015).

[170] Q. Zhu, A. R. Oganov, C. W. Glass, and H. T. Stokes, Acta Cryst. B **68**, 215 (2012).

[171] J. W. Ponder, TINKER v6.3, 2014, `http://dasher.wustl.edu/tinker/`, accessed August 10, 2015.

[172] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[173] C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek, and P. A. Wood, J. Appl. Cryst. **41**, 455 (2008).

[174] J. L. McKinley and G. J. O. Beran, J. Chem. Theory Comput. In Review (2019).

[175] M. Rajeswaran, T. N. Blanton, N. Zumbulyadis, D. J. Giesen, C. Conesa-Moratilla, S. T. Misture, P. W. Stephens, and A. Huq, J. Am. Chem. Soc. **124**, 14450 (2002).

[176] R. K. Harris, Solid State Sci. **6**, 1025 (2004).

[177] J. K. Harper and D. M. Grant, Cryst. Growth Des. **6**, 2315 (2006).

[178] M. Baias, J.-N. Dumez, P. H. Svensson, S. Schantz, G. M. Day, and L. Emsley, J. Am. Chem. Soc. **135**, 17501 (2013).

[179] L. J. Mueller and M. F. Dunn, Acc. Chem. Res. 2008 (2013).

[180] C. Yang, L. Zhu, R. A. Kudla, J. D. Hartman, R. O. Al-Kaysi, S. Monaco, B. Schatschneider, A. Magalhaes, G. J. O. Beran, C. J. Bardeen, and L. J. Mueller, CrystEngComm **18**, 7319 (2016).

[181] V. L. Deringer, V. Hoepfner, and R. Dronskowski, Cryst. Growth Des. **12**, 1014 (2012).

[182] G. Filippini and C. M. Gramaccioli, Acta Cryst. A **37**, 335 (1981).

[183] S. T. Holmes, R. J. Iuliucci, K. T. Mueller, and C. Dybowski, J. Chem. Phys. **141**, 164121 (2014).

[184] S. T. Holmes, R. J. Iuliucci, K. T. Mueller, and C. Dybowski, J. Chem. Theory Comput. **11**, 5229 (2015).

[185] J. K. Harper, R. Iuliucci, M. Gruber, and K. Kalakewich, CrystEngComm **15**, 8693 (2013).

[186] V. G. Malkin, O. L. Malkina, and D. R. Salahub, J. Am. Chem. Soc. **117**, 3294 (1995).

[187] A. Naito, S. Ganapathy, K. Akasaka, and C. A. McDowell, J. Chem. Phys. **74**, 3190 (1981).

[188] X. Chen and C.-G. Zhan, J. Mol. Struct. (THEOCHEM) **682**, 73 (2004).

[189] M. Dračínský, P. Unzueta, and G. J. O. Beran, Phys. Chem. Chem. Phys. **submitted**, (2019).

[190] M. Sherwood, D. Alderman, and D. Grant, J. Magn. Reson. A **104**, 132 (1993).

[191] M. H. Sherwood, J. C. Facelli, D. W. Alderman, and D. M. Grant, J. Am. Chem. Soc. **113**, 750 (1991).

[192] A. Naito and C. A. McDowell, J. Chem. Phys. **81**, 4795 (1984).

[193] J. C. Burley, M. J. Duer, R. S. Stein, and R. M. Vrcelj, Eur. J. Pharm. Sci. **31**, 271 (2007).

[194] C. Gervais, R. Dupree, K. J. Pike, C. Bonhomme, M. Profeta, C. J. Pickard, and F. Mauri, J. Phys. Chem. A **109**, 6960 (2005).

[195] J. Z. Hu, J. C. Facelli, D. W. Alderman, R. J. Pugmire, and D. M. Grant, J. Am. Chem. Soc. **120**, 9863 (1998).

[196] Y. Wei, A. C. de Dios, and A. E. McDermott, J. Am. Chem. Soc. **121**, 10389 (1999).

[197] S. Sharif, D. Schagen, M. D. Toney, and H.-H. Limbach, J. Am. Chem. Soc. **129**, 4440 (2007).

[198] T. Azaïs, G. Hartmeyer, S. Quignard, G. Laurent, C. Tourné-Péteilh, J.-M. Devoisselle, and F. Babonneau, Pure Appl. Chem. **81**, 1345 (2009).

[199] W. Limwikrant, A. Nagai, Y. Hagiwara, K. Higashi, K. Yamamoto, and K. Moribe, International Journal of Pharmaceutics **431**, 237 (2012).

[200] J. Czernek, Chem. Phys. Lett. **619**, 230 (2015).

[201] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni, J. Phys. Condens. Mat. **29**, 465901 (2017).

[202] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. De Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys. Condens. Mat. 395502 (2009).

[203] H. J. Monkhorst and J. D. Pack, Phys. Rev. B **13**, 5188 (1976).

[204] A. Togo and I. Tanaka, Scr. Mater. **108**, 1 (2015).

[205] J. Antony and S. Grimme, Phys. Chem. Chem. Phys. **8**, 5287 (2006).

[206] M. J. Frisch *et al.*, Gaussian Inc. (2013).

[207] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).

[208] R. Ditchfield, Mol. Phys. **27**, 789 (1974).

[209] D. B. Chesnut and K. D. Moore, J. Comp. Chem. 648 (1989).

[210] D. B. Chesnut and B. Rusiloski, J. Phys. Chem. **97**, 2839 (1993).

[211] R. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople, J. Chem. Phys. **72**, 650 (1980).

[212] A. D. McLean and G. S. Chandler, J. Chem. Phys. **72**, 5639 (1980).

[213] M. J. Frisch, J. A. Pople, and J. S. Binkley, J. Chem. Phys. **80**, 3265 (1984).

[214] T. Clark, J. Chandrasekhar, G. W. Spitznagel, and P. V. R. Schleyer, J. Comp. Chem. **4**, 294 (1983).

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

[216] P. C. Hariharan and J. A. Pople, Theor. Chim. Acta **28**, 213 (1973).

[217] F. Blanco, I. Alkorta, and J. Elguero, Magn. Reson. Chem. 797 (2007).

[218] S. T. Holmes, R. J. Iuliucci, K. T. Mueller, and C. Dybowski, J. Chem. Theory Comput. **11**, 5229 (2015).

[219] M. Addicoat, C. S. Adjiman, M. Arhangelskis, G. J. O. Beran, D. Bowskill, J. G. Brandenburg, D. E. Braun, V. Burger, J. Cole, A. J. Cruz-Cabeza, G. M. Day, V. L. Deringer, R. Guo, A. Hare, J. Helfferich, J. Hoja, L. Iuzzolino, S. Jobbins, N. Marom, D. McKay, J. B. O. Mitchell, S. Mohamed, M. Neumann, S. Nilsson Lill, J. Nyman, A. R. Oganov, P. Piaggi, S. L. Price, S. Reutzel-Edens, I. Rietveld, M. Ruggiero, M. R. Ryder, G. Sastre, J. C. Schön, C. Taylor, A. Tkatchenko, S. Tsuzuki, J. van den Ende, S. M. Woodley, G. Woollam, and Q. Zhu, Faraday Disc. **211**, 325 (2018).

[220] J. van de Streek and M. A. Neumann, Acta Cryst. B **70**, 1020 (2014).

[221] J. K. Harper, R. Iuliucci, M. Gruber, and K. Kalakewich, CrystEngComm **15**, 8693 (2013).

[222] M. Dračínský, P. Bouř, and P. Hodgkinson, J. Chem. Theory Comput. **12**, 968 (2016).

[223] L. M. LeBlanc, S. G. Dale, C. R. Taylor, A. D. Becke, G. M. Day, and E. R. Johnson, Angew. Chem. Int. Ed. **57**, 14906 (2018).

[224] S. L. Price, D. E. Braun, and S. M. Reutzel-Edens, Chem. Commun. **52**, 7065 (2016).

[225] S. L. Price and S. M. Reutzel-Edens, Drug Disc. Today **21**, 912 (2016).

[226] S. L. Price, Int. Rev. Phys. Chem. **27**, 541 (2008).

[227] A. Gavezzotti and G. Filippini, J. Am. Chem. Soc. **117**, 12299 (1995).

[228] G. Filippini, C. M. Gramaccioli, M. Simonetta, and G. B. Suffritti, Chem. Phys. Lett. **35**, 17 (1975).

[229] B. P. van Eijck, J. Comp. Chem. **22**, 816 (2001).

[230] G. M. Day and S. L. Price, J. Am. Chem. Soc. **125**, 16434 (2003).

[231] G. M. Day, S. L. Price, and M. Leslie, J. Phys. Chem. B **107**, 10919 (2003).

[232] E. Sanz, C. Vega, J. Abascal, and L. MacDowell, Phys. Rev. Lett. **92**, 255701 (2004).

[233] S. R. Whittleton, A. Otero-de-la Roza, and E. R. Johnson, J. Chem. Theory Comput. **13**, 5332 (2017).

[234] A. Erba, J. Maul, and B. Civalleri, Chem. Commun. **52**, 1820 (2016).

[235] J. G. Brandenburg, J. Potticary, H. A. Sparkes, S. L. Price, and S. R. Hall, J. Phys. Chem. Lett. **8**, 4319 (2017).

[236] Y. Heit and G. J. O. Beran, J. Comp. Chem. **35**, 2205 (2014).

[237] G. J. O. Beran, J. D. Hartman, and Y. N. Heit, Acc. Chem. Res. **49**, 2501 (2016).

[238] P. Ren, C. Wu, and J. W. Ponder, J. Chem. Theory Comput. **7**, 3143 (2011).

[239] W. H. Keesom and J. W. L. Kohler, Physica **1**, 655 (1934).

[240] I. Morrison, J.-C. Li, S. Jenkins, S. S. Xantheas, and M. C. Payne, J. Phys. Chem. B **101**, 6146 (1997).

[241] I. Nahringbauer, Acta Chem. Scand. **24**, 453 (1970).

[242] R. K. McMullan, J. Epstein, J. R. Ruble, and B. M. Craven, Acta Cryst. B **35**, 688 (1979).

[243] I. N. Krupskii, A. Prokhvatilov, A. I. Erenburg, and A. S. Barylnik, Fiz. Nizk. Temp. **8**, 533 (1982).

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

[245] P. Jönson, Acta Cryst. **B27**, 893 (1971).

[246] B. M. Craven, R. K. McMullan, J. D. Bell, and H. C. Freeman, Acta Cryst. B **B33**, 2585 (1977).

[247] G. Will, Z.Kristallogr.,Kristallgeom.,Kristallphys.,Kristallchem. **129**, 211 (1969).

[248] M. Azreg-Aïnou, Monatshefte fur Chemie **136**, 2017 (2005).

[249] W. F. Giauque and C. J. Egan, J. Chem. Phys. **5**, 45 (1937).

[250] G. Herzberg, *Electronic Spectra and Electronic Structure of Polyatomic Molecules* (D. Van Nostrand Company, Inc, ADDRESS, 1966).

[251] T. Shimanouchi, *Tables of Molecular Vibrational Frequencies*, Vol. 1 of *NSRDS NBS-39* (United States Department of Commerce, ADDRESS, 1972).

[252] R. Feistel and W. Wagner, Geochim. Cosmochim. Acta **71**, 36 (2007).

[253] NIST/TRC Web Thermo Tables (WTT), NIST Standard Reference Subscription Database 2—Lite Edition Version 2-2012-1-Lite.

[254] *Handbook of Chemistry and Physics*, 96 ed., edited by W. M. Haynes (CRC Press, ADDRESS, 2004).

[255] S. Verevkin, J. Chem. Eng. Data **45**, 953 (2000).

[256] J. Martin and R. L. Andon, J. Chem. Thermo. **14**, 679 (1982).

[257] P. Jiménez and C. T. M.V. Roux, J. Chem. Thermo. **19**, 985 (1987).

[258] H. G. M. De Wit, C. G. De Kruif, and J. C. Van Miltenburg, J. Chem. Thermo. **15**, 891 (1983).

[259] F. Billes, H. Endrédi, and G. Jalsovszky, J. Mol. Struct. (Theochem) **465**, 157 (1999).

[260] 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.

[261] A. Hesselmann, J. Chem. Phys. **128**, 144112 (2008).

[262] M. Pitonak and A. Hesselmann, J. Chem. Theory Comput. **6**, 168 (2010).

[263] L. A. Burns, M. S. Marshall, and C. D. Sherrill, J. Chem. Phys. **141**, 234111 (2014).

[264] Y. Huang, Y. Shao, and G. J. O. Beran, J. Chem. Phys. **138**, 224112 (2013).

[265] A. Karton and J. M. L. Martin, Theor. Chem. Acc. **115**, 330 (2006).

[266] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[267] J. W. Ponder, TINKER v7.1, 2015, `http://dasher.wustl.edu/tinker/`, accessed September 21, 2016.

[268] J. C. Wu, G. Chattree, and P. Ren, Theor. Chem. Acc. **131**, 1138 (2012).

[269] M. F. Peintinger, D. V. Oliveira, and T. Bredow, J. Comp. Chem. **34**, 451 (2013).

[270] R. Dovesi, R. Orlando, B. Civalleri, C. Roetti, V. R. Saunders, and C. M. Zicovich-Wilson, Z. Kristallogr. **220**, 571 (2005).

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

[272] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, J. Phys. Condens. Mat. **21**, 395502 (2009).

[273] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[274] A. Otero-de-la Roza and E. R. Johnson, J. Chem. Phys. **136**, 174109 (2012).

[275] A. Otero-De-La-Roza, B. H. Cao, I. K. Price, J. E. Hein, and E. R. Johnson, Angew. Chem. Int. Ed. **53**, 7879 (2014).

[276] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[277] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[278] S. Grimme, J. Comp. Chem. **16**, 1787 (2006).

[279] J. Li and D. K. Ross, Nature **365**, 327 (1993).

[280] J. Li, J. Chem. Phys. **105**, 6733 (1996).

[281] X. He, O. Sode, S. S. Xantheas, and S. Hirata, J. Chem. Phys. **137**, 204505 (2012).

[282] P. Zhang, L. Tian, Z. P. Zhang, G. Shao, and J. C. Li, J. Chem. Phys. **137**, 044504 (2012).

[283] K. E. Riley, M. Pitonak, P. Jurecka, and P. Hobza, Chemical Reviews **110**, 5023 (2010).

[284] M. A. Salim, S. Y. Willow, and S. Hirata, J. Chem. Phys. **144**, 204503 (2016).

[285] O. Demerdash, Y. Mao, T. Liu, M. Head-Gordon, and T. Head-Gordon, The Journal of Chemical Physics **147**, 161721 (2017).

[286] W. Acree and J. S. Chickos, J. Phys. Chem. Ref. Data **39**, 043101 (2010).

[287] G. L. Perlovich, T. V. Volkova, and A. Bauer-Brandl, J. Therm. Anal. Cal. **89**, 767 (2007).

[288] G. L. Perlovich, T. V. Volkova, and A. BauerBrandl, Journal of Pharmaceutical Sciences **95**, 2158 (2006).

[289] R. Picciochi, H. P. Diogo, and M. E. Minas da Piedade, J. Therm. Anal. Cal. **100**, 391 (2010).

[290] J. By, Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences **157**, 79 (1936).

[291] Y. Ebisuzaki, L. H. Askari, A. M. Bryan, and M. F. Nicol, The Journal of Chemical Physics **87**, 6659 (1987).

[292] S. Kichanov, D. Kozlenko, P. Bilski, J. Wąsicki, W. Nawrocik, A. Medek, B. Hancock, E. Lukin, C. Lathe, L. Dubrovinsky, and B. Savenko, Journal of Molecular Structure **1006**, 337 (2011).

[293] S. K. Deb, M. A. Rekha, A. P. Roy, V. Vijayakumar, S. Meenakshi, and B. K. Godwal, Physical Review B **47**, 11491 (1993).

[294] R. Rao, T. Sakuntala, and B. K. Godwal, Physical Review B **65**, 054108 (2002).

[295] K. Drużbicki, E. Mikuli, N. Pałka, S. Zalewski, and M. D. Ossowska-Chruściel, The Journal of Physical Chemistry B **119**, 1681 (2015).

[296] E. M. Gonçalves, F. Agapito, T. S. Almeida, and J. A. Martinho Simões, The Journal of Chemical Thermodynamics **73**, 90 (2014).

[297] S. P. Verevkin and S. A. Kozlova, Thermochimica Acta **471**, 33 (2008).

[298] Q. Zhu, A. G. Shtukenberg, D. J. Carter, T.-Q. Yu, J. Yang, M. Chen, P. Raiteri, A. R. Oganov, B. Pokroy, I. Polishchuk, P. J. Bygrave, G. M. Day, A. L. Rohl, M. E. Tuckerman, and B. Kahr, J. Am. Chem. Soc. **138**, 4881 (2016).

[299] J. Chatchawalsaisin, J. Kendrick, S. C. Tuble, and J. Anwar, CrystEngComm **10**, 437 (2008).

[300] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, and B. Kirtman, Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1360 (2018).

[301] Y. Huang and G. J. O. Beran, The Journal of Chemical Physics **143**, 044113 (2015).

[302] J. L. M. Chandler Greenwell and G. J. O. Beran, In Preparation (2019).

[303] R. Sedlak, K. E. Riley, J. ezáč, M. Pitoák, and P. Hobza, ChemPhysChem **14**, 698 (2013).

[304] J. ezáč, Y. Huang, P. Hobza, and G. J. O. Beran, Journal of Chemical Theory and Computation **11**, 3065 (2015).

[305] R. Podeszwa, B. M. Rice, and K. Szalewicz, Physical Review Letters **101**, 115503 (2008).

[306] J. Hoja and A. Tkatchenko, Faraday Discussions **211**, 253 (2018).

[307] B. Monserrat, N. D. Drummond, and R. J. Needs, Physical Review B **87**, 144302 (2013).

[308] N. D. Drummond, B. Monserrat, J. H. Lloyd-Williams, P. L. Ríos, C. J. Pickard, and R. J. Needs, Nature Communications **6**, 7794 (2015).

[309] E. A. Engel, B. Monserrat, and R. J. Needs, Physical Review X **5**, 021033 (2015).

[310] W. Sontising and G. J. O. Beran, Vis. Rev. Material Submitted .

[311] G. Kresse and J. Furthmüller, Physical Review B **54**, 11169 (1996).

# Appendix A

# Improving Predicted Nuclear Magnetic Resonance Chemical Shifts Using the Quasi-Harmonic Approximation

Section A.1 shows violin plots for the errors obtained by fully optimizing a larger $^{13}C$ and $^{15}N$ test set which had been used in previous studies[12]. Sections A.2.1 and A.2.2 provides sensitivity analysis on the linear regression fit parameters used in this study. Section A.3 provides the experimental and predicted chemical shift data that is analyzed in Chapter 3. Section A.4 displays the thermal expansion curves of every structure that was thermally expanded using the quasi-harmonic approximation (QHA). Section A.5 provides a visual analog to the tabular data displayed in Chapter 4.2. Section A.6 gives detailed his-

tograms which display how individual chemical shifts are improved by thermal expansion. Section 7A.7 provides the predicted $^{1}$H, $^{13}$C, and $^{15}$ N isotropic chemical shifts, and the thermal expansion curves related to the 5 theophylline structures that were taken from the Emsly's 2013 PCCP paper.[134] Section A.8 provides the Mercury RefCodes that are used to create Figure 3.1.

## A.1 Fully optimized crystals vs. Fixed cell optimizations

Previous studies have employed a larger test set than our current study (169 $^{13}$C isotropic chemical shifts and 52 $^{15}N$ isotropic chemical shifts compared to our 68 and 28 respectively).[104] While it could be possible to test the effects of employing the QHA on this entire set, the computational cost associated with doing so make this undesirable. Instead we model only a subset of the systems from the larger test set. The use of smaller benchmark sets might introduce some biases into the data. To get an idea of the difference in the errors statistics between the full set from the earlier work and the smaller subset used here, Figure A.1 compares the full optimization of the structures present in the $^{13}$C and $^{15}$N sets to their fixed cell optimizations. Like in Chapter 3 the chemical shifts for the full optimization are taken from a previous study.[12] The trends for the full set in Figure A.1 are consistent with the trends seen in the chosen smaller subset in Chapter 3.

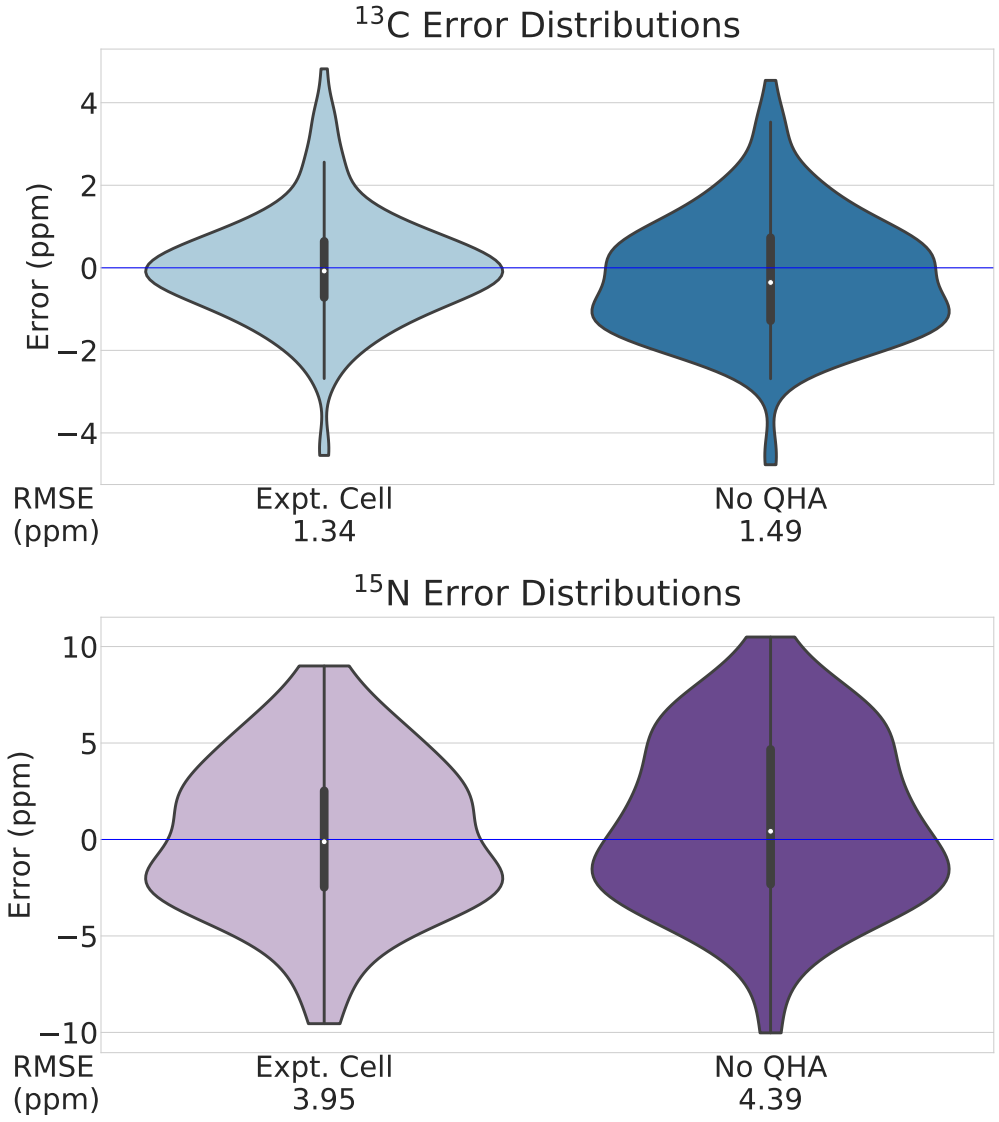



Figure A.1: Distribution of errors relative to experiment over a) 169 $^{13}C$ chemical shifts for 25 molecular crystals and b) 52 $^{15}N$ chemical shifts for 25 molecular crystals. Root-mean-squared errors relative to experiment are displayed below each distribution.

## A.2 Sensitivity of fits analysis

Throughout Chapter 3 a linear regression scheme is used to map between the computed absolute chemical shieldings $\sigma_i$ and the experimentally observed chemical shifts $\delta_i$,

$$\delta_i = A\sigma_i + B \tag{A.1}$$

The regression parameters $A$ and $B$ were taken from Ref 12. The specific model parameters used to calculate the chemical shifts in that work are identical to the ones used here, but the geometry optimization protocols differed slightly. That work employed fixed experimental lattice parameter optimizations on a set that includes all twenty crystals in the benchmark here plus many additional structures. Those crystal structures were optimized in Quantum Espresso v6.0 with PBE-D2, an 80 Ry kinetic energy cutoff, ultrasoft pseudopotentials, and a 5x5x5 Monkhorst-Pack k-point grid.

In the present study, those regression parameters were employed for the No QHA, 0 K, and 300 K QHA structures. To ensure the regression parameters are transferable here we tested two additional linear regression fits (see Table A.1): (1) fitting directly on the 300 K quasi-harmonic (QHA) optimized structures and (2) fitting to structures optimized with B86bPBE-XDM and fixed cells. Since the previous fit will give optimal performance at the known experimental volumes, the first fit tests how close to the "optimal volumes" the 300 K QHA structures are. While there is a slight change in the slope and intercept when fitting against the 300K structures the effect on the test set remains minimal (See Figure A.2).

The second fit compares the impact of changing the density functional and asso-

ciated parameters (dispersion correction, basis set, k-point grid, etc) used to perform the fixed cell optimizations. Once again, the regression slopes and intercepts change only minimally. In other words, the regression models used are appropriate despite changes in the geometries and structure optimization protocols.

Table A.1: Comparison of regression models for three different fits of Eq A.1.[12]

| Nucleus | Geometries Used in Fit | Slope | Intercept |
|---|---|---|---|
| $^{13}$C | 300 K QHA | -0.9617 | 179.11 |
| | B86bPBE-XDM Fixed Cell | -0.9645 | 179.49 |
| | PBE-D2 Fixed Cell (Ref 12) | -0.9658 | 179.48 |
| $^{15}$N | 300 K QHA | -1.0044 | 195.07 |
| | PBE-D2 Fixed Cell (Ref 12) | -1.0106 | 197.46 |

### A.2.1 300 K fit

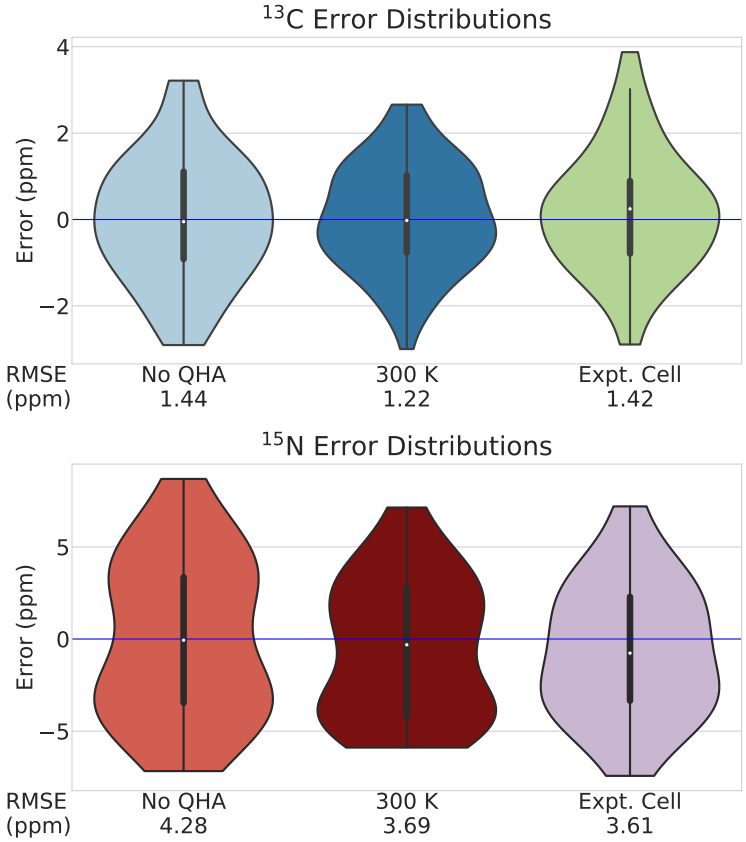



Figure A.2: Distribution of errors relative to experiment over a) 68 $^{13}C$ chemical shifts for 12 molecular crystals and b) 28 $^{15}N$ chemical shifts for 14 molecular crystals. Root-mean-squared errors relative to experiment are displayed below each distribution.

### A.2.2 B86bPBE-XDM vs. PBE-D2 Fixed cell opt.

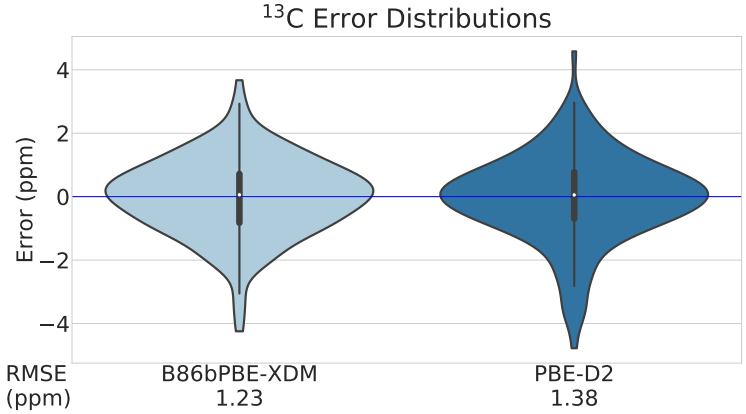



Figure A.3: Distribution of errors relative to experiment over a) 169 $^{13}C$ chemical shifts for 25 molecular crystals. Root-mean-squared errors relative to experiment are displayed below each distribution.

## A.3 Raw Chemical Shift Data

### A.3.1 All QHA structures

Table A.2: Comparison of the experimental and predicted **$^{13}$C isotropic chemical shifts** for the Fully optimized structure (No QHA), QHA-expanded to 0 K (0 K), QHA-expanded to 300 K (300 K), and the fixed cell optimization (Expt. Cell). Calculations were done according to the method in Sec. 3.3.3 in Chapter 3. The assigned functional group is displayed before the shifts.

| Crystal | Functional Group | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|---|---|
| HXACAN13 | Aromatic | 152.30 | 151.13 | 150.06 | 150.63 | 151.21 |
| | Aromatic | 116.40 | 115.25 | 115.41 | 115.06 | 115.05 |
| | Aromatic | 120.60 | 119.81 | 120.54 | 120.07 | 119.64 |
| | Aromatic | 133.10 | 131.57 | 131.38 | 131.59 | 131.42 |
| | Aromatic | 123.40 | 123.16 | 123.76 | 123.16 | 122.94 |
| | Aromatic | 115.70 | 115.04 | 115.10 | 115.09 | 114.76 |
| | Amide | 169.80 | 169.90 | 169.61 | 170.06 | 170.11 |
| | Alkane | 23.80 | 25.59 | 26.73 | 26.28 | 25.74 |
| GLYCIN03 | Carboxylate ion | 176.20 | 179.00 | 178.59 | 178.21 | 178.00 |
| | $\alpha$-C | 43.50 | 43.69 | 43.64 | 43.76 | 43.84 |

| Crystal | Functional Group | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|---|---|
| LALNIN12 | Carboxylate ion | 176.80 | 180.74 | 180.24 | 179.90 | 179.83 |
| | $\alpha$-C | 50.90 | 51.41 | 51.18 | 51.68 | 51.87 |
| | Alkane | 19.80 | 20.90 | 20.60 | 20.59 | 20.70 |
| LSERIN01 | Carboxylate ion | 175.10 | 177.84 | 177.09 | 176.77 | 176.65 |
| | $\alpha$-C | 55.60 | 55.57 | 55.41 | 55.69 | 55.88 |
| | Alcohol | 62.90 | 64.26 | 64.20 | 63.99 | 63.89 |
| LTYROS11 | Carboxylate ion | 175.42 | 178.80 | 178.95 | 178.53 | 178.24 |
| | Aromatic | 131.04 | 130.61 | 131.13 | 130.88 | 130.69 |
| | Aromatic | 116.42 | 116.90 | 117.72 | 117.00 | 116.56 |
| | $\alpha$-C | 56.41 | 56.82 | 56.68 | 57.16 | 57.39 |
| | Aromatic | 130.34 | 130.80 | 130.23 | 130.27 | 130.27 |
| | Aromatic | 118.02 | 118.73 | 119.55 | 118.96 | 118.59 |
| | Aromatic | 155.62 | 155.34 | 154.51 | 155.00 | 155.18 |
| | Alkane | 36.80 | 38.09 | 37.85 | 37.87 | 37.91 |
| | Aromatic | 123.64 | 122.01 | 122.42 | 122.62 | 122.65 |
| SUCROS04 | Ether | 93.30 | 92.75 | 91.67 | 92.28 | 92.54 |
| | Alcohol | 66.00 | 64.94 | 64.69 | 64.99 | 65.10 |
| | Alcohol | 73.70 | 73.15 | 72.74 | 72.98 | 73.07 |
| | Ether | 102.40 | 103.37 | 102.56 | 103.11 | 103.41 |
| | Alcohol | 72.80 | 72.08 | 72.01 | 72.34 | 72.45 |

| Crystal | Functional Group | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|---|---|
| | Alcohol | 82.90 | 81.93 | 81.13 | 81.63 | 81.89 |
| | Alcohol | 67.90 | 67.45 | 66.91 | 67.07 | 67.13 |
| | Alcohol | 71.80 | 70.63 | 70.04 | 70.38 | 70.57 |
| | Ether | 73.60 | 73.44 | 72.84 | 73.30 | 73.54 |
| | Ether | 81.80 | 80.39 | 79.11 | 79.58 | 79.80 |
| | Alcohol | 60.00 | 58.95 | 58.83 | 58.91 | 58.88 |
| | Alcohol | 61.00 | 61.27 | 60.97 | 61.12 | 61.19 |
| NAPHTA36 | Aromatic | 126.00 | 125.97 | 126.38 | 126.18 | 125.83 |
| | Aromatic | 129.30 | 130.03 | 130.58 | 130.18 | 129.68 |
| | Aromatic | 134.90 | 133.31 | 133.02 | 132.98 | 132.87 |
| | Aromatic | 129.90 | 130.42 | 131.08 | 130.70 | 130.18 |
| | Aromatic | 125.40 | 125.20 | 125.40 | 125.23 | 125.01 |
| ADENOS12 | Aromatic Purine | 154.09 | 154.18 | 154.69 | 154.50 | 154.12 |
| | Aromatic Purine | 147.94 | 146.06 | 145.71 | 145.67 | 146.11 |
| | Aromatic Purine | 119.30 | 119.31 | 120.26 | 120.69 | 120.82 |
| | Aromatic Purine | 154.69 | 152.05 | 152.29 | 152.87 | 152.96 |
| | Aromatic Purine | 137.38 | 138.05 | 139.58 | 139.46 | 139.26 |
| | Ether | 91.91 | 92.27 | 91.36 | 91.51 | 91.57 |
| | Alcohol | 74.61 | 75.37 | 74.95 | 75.12 | 75.18 |
| | Alcohol | 70.86 | 71.50 | 71.40 | 71.87 | 72.07 |

| Crystal | Functional Group | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|---|---|
| | Ether | 84.45 | 84.93 | 84.30 | 84.39 | 84.39 |
| | Alcohol | 62.45 | 62.62 | 61.70 | 62.10 | 62.19 |
| ASPARM03 | Carboxylate ion | 176.40 | 178.03 | 178.14 | 177.74 | 177.51 |
| | $\alpha$-C | 51.80 | 50.93 | 50.81 | 51.19 | 51.42 |
| | Alkane | 36.10 | 36.06 | 35.44 | 35.27 | 35.22 |
| | Amide | 177.10 | 174.55 | 174.54 | 174.49 | 174.45 |
| LTHREO01 | Carboxylate ion | 171.92 | 173.88 | 173.42 | 173.22 | 173.07 |
| | $\alpha$-C | 61.17 | 60.01 | 59.23 | 59.66 | 59.87 |
| | Alcohol | 66.82 | 67.16 | 66.55 | 66.90 | 67.15 |
| | Alkane | 20.40 | 21.67 | 20.72 | 20.94 | 21.09 |
| LCYSTN21 | Alkane | 28.77 | 32.39 | 30.01 | 29.99 | 29.91 |
| | $\alpha$-C | 56.67 | 54.80 | 55.63 | 56.07 | 56.25 |
| | Carboxylate ion | 174.00 | 175.23 | 175.49 | 174.97 | 174.84 |
| GLUTAM01 | Carboxylate ion | 172.97 | 175.91 | 175.34 | 175.14 | 175.00 |
| | $\alpha$-C | 53.30 | 54.31 | 54.03 | 54.49 | 54.74 |
| | Alkane | 25.50 | 26.79 | 25.98 | 26.00 | 26.07 |
| | Alkane | 28.52 | 30.75 | 30.89 | 30.67 | 30.60 |
| | Amide | 176.50 | 175.81 | 175.40 | 175.45 | 175.50 |

Table A.3: Comparison of the experimental and predicted **$^{15}$N isotropic chemical shifts** for the Fully optimized structure (No QHA), QHA-expanded to 0 K (0 K), QHA-expanded to 300 K (300 K), and the fixed cell optimization (Expt. Cell). Calculations were done according to the method in Sec. 3.3.3 in Chapter 3. The assigned functional group is displayed before the shifts.

| Crystal | Functional Group | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|---|---|
| HXACAN13 | Amide | 97.91 | 104.98 | 104.95 | 105.43 | 105.22 |
| LSERIN01 | Ammonium ion | -4.10 | -6.18 | -6.31 | -6.96 | -7.22 |
| BITZAF | Pyridyl | 249.50 | 244.76 | 245.02 | 245.70 | 246.30 |
| LHISTD02 | Pyrrole | 132.60 | 138.14 | 139.13 | 138.35 | 138.10 |
| | Pyrrole | 210.80 | 212.52 | 213.61 | 213.39 | 213.47 |
| GLYCIN03 | Ammonium ion | -6.50 | -7.99 | -8.71 | -9.57 | -10.06 |
| FUSVAQ | Nucleobase N | 183.20 | 183.15 | 179.78 | 180.52 | 181.31 |
| | Nucleobase N | 174.20 | 172.96 | 170.70 | 171.67 | 172.16 |
| | Nucleobase N | 192.20 | 193.12 | 191.54 | 191.61 | 191.99 |
| | Nucleobase N | 120.20 | 122.92 | 126.41 | 125.46 | 124.77 |
| | Amine | 50.20 | 46.75 | 50.14 | 48.98 | 48.46 |
| CYTSIN | Nucleobase N | 110.20 | 115.30 | 114.88 | 114.93 | 114.93 |

| Crystal | Functional Group | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|---|---|
| | Nucleobase N | 165.20 | 162.80 | 163.42 | 163.46 | 163.60 |
| | Amine | 54.20 | 56.18 | 58.64 | 58.51 | 58.57 |
| URACIL | Nucleobase N | 96.20 | 105.19 | 106.69 | 105.75 | 105.13 |
| | Nucleobase N | 120.20 | 122.17 | 122.04 | 121.85 | 121.78 |
| BAPLOT01 | Nucleobase N | 114.70 | 121.97 | 122.69 | 122.24 | 121.92 |
| | Nucleobase N | 72.70 | 77.70 | 79.27 | 78.84 | 78.65 |
| | Nucleobase N | 122.70 | 125.71 | 128.09 | 126.76 | 126.00 |
| | Nucleobase N | 178.70 | 178.66 | 178.81 | 179.55 | 180.57 |
| LCYSTN21 | Ammonium ion | -0.40 | -4.96 | -2.28 | -3.15 | -3.53 |
| GLUTAM01 | Ammonium ion | -1.30 | -3.04 | -3.64 | -3.70 | -3.66 |
| | Amide | 72.10 | 76.59 | 80.26 | 77.99 | 76.95 |
| ALUCAL04 | Ammonium ion | 3.00 | -0.11 | -0.47 | -0.73 | -0.77 |
| THYMIN01 | Nucleobase N | 90.20 | 94.68 | 95.22 | 94.72 | 94.09 |
| | Nucleobase N | 119.20 | 121.06 | 121.53 | 121.38 | 121.07 |
| ASPARM03 | Ammonium ion | 0.70 | -2.03 | -2.13 | -2.48 | -2.72 |
| | Amide | 74.90 | 78.41 | 79.57 | 78.04 | 77.22 |

### A.3.2 Pharmaceutical crystals

Table A.4: Comparison of the experimental and predicted **$^{13}$C isotropic chemical shifts** for the Fully optimized structure (No QHA), QHA-expanded to 0 K (0 K), QHA-expanded to 300 K (300 K), and the fixed cell optimization (Expt. Cell). Calculations were done according to the method in Sec. 3.3.3 in Chapter 3. The k-points used for each structure are displayed after the calculated chemical shifts.

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) | KPOINTS |
|---|---|---|---|---|---|---|
| HXACAN13 | 152.30 | 151.13 | 150.06 | 150.63 | 151.21 | 1 3 3 |
| | 116.40 | 115.25 | 115.41 | 115.06 | 115.05 | |
| | 120.60 | 119.81 | 120.54 | 120.07 | 119.64 | |
| | 133.10 | 131.57 | 131.38 | 131.59 | 131.42 | |
| | 123.40 | 123.16 | 123.76 | 123.16 | 122.94 | |
| | 115.70 | 115.04 | 115.10 | 115.09 | 114.76 | |
| | 169.80 | 169.90 | 169.61 | 170.06 | 170.11 | |
| | 23.80 | 25.59 | 26.73 | 26.28 | 25.74 | |
| BAPLOT01 | 150.80 | 149.81 | 150.05 | 149.95 | 149.98 | 5 5 5 |
| | 146.10 | 145.62 | 144.37 | 145.03 | 146.08 | |
| | 105.80 | 105.20 | 103.60 | 104.79 | 106.37 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) | KPOINTS |
|---|---|---|---|---|---|---|
| | 155.00 | 154.47 | 153.68 | 154.11 | 154.75 | |
| | 140.80 | 141.15 | 141.02 | 140.79 | 141.23 | |
| | 29.90 | 30.46 | 31.36 | 30.66 | 30.11 | |
| | 29.90 | 29.32 | 29.78 | 29.38 | 29.14 | |
| COYRUD12 | 157.80 | 157.27 | 156.72 | 156.78 | 156.83 | 3 3 1 |
| | 118.90 | 118.60 | 118.25 | 118.28 | 118.36 | |
| | 129.30 | 130.94 | 131.65 | 131.35 | 130.88 | |
| | 128.50 | 126.92 | 126.56 | 126.68 | 126.74 | |
| | 130.10 | 128.81 | 129.19 | 128.95 | 128.45 | |
| | 134.50 | 135.04 | 133.79 | 134.08 | 134.04 | |
| | 123.70 | 124.15 | 124.29 | 124.30 | 124.04 | |
| | 128.50 | 128.43 | 129.77 | 129.35 | 128.80 | |
| | 134.50 | 133.21 | 132.44 | 132.62 | 132.61 | |
| | 104.10 | 104.76 | 105.18 | 104.88 | 104.56 | |
| | 53.10 | 52.80 | 52.69 | 52.81 | 52.69 | |
| | 46.80 | 48.91 | 49.10 | 49.09 | 48.98 | |
| | 17.10 | 17.95 | 17.99 | 18.34 | 18.50 | |
| | 179.20 | 181.17 | 181.32 | 181.25 | 180.95 | |
| IBPRAC16 | 183.20 | 186.14 | 185.83 | 186.05 | 186.00 | 1 3 1 |
| | 44.20 | 45.29 | 44.42 | 44.92 | 45.10 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) | KPOINTS |
|---|---|---|---|---|---|---|
| | 15.40 | 16.42 | 17.23 | 16.99 | 16.71 | |
| | 137.20 | 136.02 | 135.31 | 136.17 | 135.45 | |
| | 132.30 | 132.70 | 133.63 | 133.04 | 133.33 | |
| | 126.70 | 126.47 | 127.58 | 126.83 | 126.76 | |
| | 129.00 | 129.34 | 129.62 | 129.58 | 129.35 | |
| | 130.80 | 131.33 | 131.82 | 131.54 | 131.26 | |
| | 142.00 | 141.69 | 141.56 | 141.91 | 141.97 | |
| | 46.00 | 45.74 | 45.47 | 45.67 | 45.84 | |
| | 32.60 | 35.11 | 34.46 | 34.67 | 34.94 | |
| | 22.00 | 22.27 | 23.17 | 22.64 | 22.29 | |
| | 25.10 | 25.17 | 25.95 | 25.51 | 25.24 | |
| CBMZPN23 | 137.30 | 135.21 | 134.84 | 135.68 | 136.47 | 3 3 1 |
| | 131.00 | 126.99 | 127.58 | 127.16 | 126.85 | |
| | 127.30 | 130.36 | 131.15 | 131.11 | 130.87 | |
| | 131.00 | 128.85 | 129.19 | 129.05 | 128.64 | |
| | 129.20 | 129.05 | 128.70 | 128.51 | 127.88 | |
| | 134.70 | 133.75 | 132.75 | 133.04 | 133.13 | |
| | 132.00 | 133.39 | 134.74 | 134.34 | 133.77 | |
| | 132.00 | 135.06 | 135.69 | 135.34 | 135.01 | |
| | 134.70 | 134.65 | 134.25 | 134.27 | 134.28 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) | KPOINTS |
|---|---|---|---|---|---|---|
| | 129.20 | 133.96 | 134.35 | 133.60 | 133.17 | |
| | 131.00 | 127.06 | 126.68 | 126.80 | 126.80 | |
| | 127.30 | 128.06 | 128.24 | 127.84 | 127.50 | |
| | 131.00 | 129.96 | 130.35 | 130.10 | 130.07 | |
| | 140.40 | 139.51 | 137.78 | 138.48 | 139.08 | |
| | 159.00 | 157.80 | 156.82 | 156.40 | 155.69 | |

### A.3.3 Fixed Cell v. Full Cell

Table A.5: Comparison of the experimental and predicted **$^{13}$C isotropic chemical shifts** for the fully optimized structure (Opt cell) and the fixed cell optimization (Expt. Cell). The k-points used for each fully optimized structure are displayed after the calculated chemical shifts.

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| GLYCIN03 | 176.20 | 179.00 | 178.59 | 5 1 5 |
| | 43.50 | 43.69 | 43.64 | |
| LALNIN12 | 176.80 | 180.74 | 180.24 | 5 1 5 |
| | 50.90 | 51.41 | 51.18 | |
| | 19.80 | 20.90 | 20.60 | |
| LSERIN01 | 175.10 | 176.65 | 177.09 | 7 3 3 |
| | 55.60 | 55.88 | 55.41 | |
| | 62.90 | 63.89 | 64.20 | |
| LTYROS11 | 175.42 | 178.80 | 178.95 | 5 1 5 |
| | 131.04 | 130.61 | 131.13 | |
| | 116.42 | 116.90 | 117.72 | |
| | 56.41 | 56.82 | 56.68 | |
| | 130.34 | 130.80 | 130.23 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 118.02 | 118.73 | 119.55 | |
| | 155.62 | 155.34 | 154.51 | |
| | 36.80 | 38.09 | 37.85 | |
| | 123.64 | 122.01 | 122.42 | |
| LCYSTN21 | 28.77 | 32.39 | 30.01 | 5 5 5 |
| | 56.67 | 54.80 | 55.63 | |
| | 174.00 | 175.23 | 175.49 | |
| MGLUCP11 | 101.00 | 100.28 | 99.42 | 5 5 5 |
| | 72.30 | 70.83 | 70.17 | |
| | 74.60 | 73.82 | 73.07 | |
| | 72.50 | 71.50 | 70.55 | |
| | 75.30 | 73.63 | 73.67 | |
| | 63.80 | 63.49 | 62.69 | |
| | 56.50 | 57.40 | 56.93 | |
| MBDGAL02 | 105.70 | 105.02 | 103.88 | 5 5 5 |
| | 71.20 | 70.73 | 70.67 | |
| | 72.10 | 72.18 | 70.81 | |
| | 69.30 | 70.34 | 69.40 | |
| | 75.60 | 74.50 | 73.59 | |
| | 62.80 | 62.43 | 61.56 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 57.60 | 59.04 | 58.61 | |
| MEMANP11 | 99.60 | 99.61 | 98.52 | 5 5 5 |
| | 71.30 | 71.12 | 70.06 | |
| | 71.70 | 71.99 | 70.83 | |
| | 64.80 | 64.55 | 64.09 | |
| | 71.90 | 72.37 | 71.25 | |
| | 58.90 | 58.56 | 57.94 | |
| | 54.90 | 56.44 | 55.88 | |
| MGALPY01 | 100.40 | 100.09 | 99.12 | 5 5 5 |
| | 67.60 | 67.06 | 66.92 | |
| | 72.60 | 72.86 | 72.52 | |
| | 70.00 | 69.57 | 68.59 | |
| | 72.90 | 72.47 | 71.33 | |
| | 61.40 | 61.27 | 60.26 | |
| | 55.20 | 56.15 | 55.51 | |
| XYLOBM01 | 104.20 | 103.37 | 102.76 | 5 5 5 |
| | 72.20 | 71.41 | 70.38 | |
| | 78.20 | 77.20 | 76.81 | |
| | 69.50 | 69.19 | 68.70 | |
| | 66.90 | 66.77 | 66.54 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 57.30 | 57.60 | 58.05 | |
| FRUCTO02 | 65.40 | 65.32 | 64.47 | 5 5 5 |
| | 99.70 | 101.44 | 100.46 | |
| | 67.20 | 67.45 | 67.72 | |
| | 69.00 | 67.74 | 67.50 | |
| | 71.40 | 72.18 | 71.13 | |
| | 64.90 | 65.71 | 65.08 | |
| RHAMAH12 | 94.50 | 94.49 | 93.32 | 5 5 5 |
| | 72.20 | 71.89 | 70.65 | |
| | 71.00 | 69.77 | 69.05 | |
| | 72.50 | 72.57 | 71.22 | |
| | 69.80 | 70.44 | 70.19 | |
| | 17.80 | 18.29 | 18.56 | |
| SUCROS04 | 93.30 | 92.75 | 91.67 | 3 3 3 |
| | 66.00 | 64.94 | 64.69 | |
| | 73.70 | 73.15 | 72.74 | |
| | 102.40 | 103.37 | 102.56 | |
| | 72.80 | 72.08 | 72.01 | |
| | 82.90 | 81.93 | 81.13 | |
| | 67.90 | 67.45 | 66.91 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 71.80 | 70.63 | 70.04 | |
| | 73.60 | 73.44 | 72.84 | |
| | 81.80 | 80.39 | 79.11 | |
| | 60.00 | 58.95 | 58.83 | |
| | 61.00 | 61.27 | 60.97 | |
| GLUTAM01 | 172.97 | 175.91 | 175.34 | 5 5 5 |
| | 53.30 | 54.31 | 54.03 | |
| | 25.50 | 26.79 | 25.98 | |
| | 28.52 | 30.75 | 30.89 | |
| | 176.50 | 175.81 | 175.40 | |
| ASPARM03 | 176.40 | 178.03 | 178.14 | 5 5 5 |
| | 51.80 | 50.93 | 50.81 | |
| | 36.10 | 36.06 | 35.44 | |
| | 177.10 | 174.55 | 174.54 | |
| LSERMH10 | 175.60 | 178.03 | 177.78 | 5 5 5 |
| | 58.30 | 58.95 | 58.54 | |
| | 61.80 | 61.46 | 61.23 | |
| LTHREO01 | 171.92 | 173.88 | 173.42 | 5 5 5 |
| | 61.17 | 60.01 | 59.23 | |
| | 66.82 | 67.16 | 66.55 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 20.40 | 21.67 | 20.72 | |
| NAPHTA36 | 126.00 | 125.97 | 126.38 | 3 3 3 |
| | 129.30 | 130.03 | 130.58 | |
| | 134.90 | 133.31 | 133.02 | |
| | 129.90 | 130.42 | 131.08 | |
| | 125.40 | 125.20 | 125.40 | |
| ACENAP03 | 148.10 | 149.06 | 149.83 | 5 5 5 |
| | 120.30 | 120.18 | 120.77 | |
| | 129.40 | 130.03 | 131.34 | |
| | 122.30 | 122.21 | 122.25 | |
| | 131.90 | 130.03 | 130.43 | |
| | 139.90 | 137.95 | 138.66 | |
| | 29.50 | 30.84 | 31.21 | |
| TRIPHE11 | 126.40 | 126.36 | 127.26 | 5 5 5 |
| | 129.50 | 128.68 | 129.00 | |
| | 124.50 | 125.01 | 124.90 | |
| | 125.90 | 125.59 | 126.41 | |
| | 127.50 | 126.94 | 127.44 | |
| | 122.30 | 122.11 | 122.24 | |
| | 130.20 | 129.35 | 129.34 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 129.50 | 128.20 | 127.00 | |
| | 120.90 | 120.37 | 120.50 | |
| | 125.90 | 125.78 | 125.43 | |
| | 121.70 | 121.82 | 122.33 | |
| | 129.50 | 128.68 | 127.79 | |
| | 129.50 | 128.29 | 127.90 | |
| | 122.30 | 122.79 | 123.02 | |
| | 126.90 | 126.46 | 126.16 | |
| | 126.90 | 128.29 | 128.06 | |
| | 123.80 | 124.04 | 124.24 | |
| | 129.80 | 129.16 | 129.40 | |
| HXACAN13 | 152.30 | 151.13 | 150.06 | 1 3 3 |
| | 116.40 | 115.25 | 115.41 | |
| | 120.60 | 119.81 | 120.54 | |
| | 133.10 | 131.57 | 131.38 | |
| | 123.40 | 123.16 | 123.76 | |
| | 115.70 | 115.04 | 115.10 | |
| | 169.80 | 169.90 | 169.61 | |
| | 23.80 | 25.59 | 26.73 | |
| INDMET | 167.70 | 169.73 | 169.31 | 5 5 5 |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 136.70 | 132.16 | 132.35 | |
| | 134.40 | 132.83 | 132.63 | |
| | 129.20 | 130.22 | 130.29 | |
| | 140.30 | 143.94 | 144.09 | |
| | 127.00 | 126.65 | 126.43 | |
| | 132.00 | 132.83 | 132.80 | |
| | 13.70 | 16.26 | 16.84 | |
| | 28.20 | 30.46 | 30.34 | |
| | 179.20 | 184.02 | 183.74 | |
| | 138.00 | 137.66 | 136.58 | |
| | 55.20 | 55.47 | 55.21 | |
| | 112.60 | 113.13 | 114.20 | |
| | 131.10 | 129.93 | 129.17 | |
| | 97.90 | 97.58 | 98.88 | |
| | 156.70 | 155.62 | 155.48 | |
| | 112.60 | 111.68 | 111.12 | |
| | 115.70 | 115.83 | 116.81 | |
| | 131.10 | 129.16 | 129.26 | |
| SULAMD06 | 127.10 | 122.87 | 122.33 | 5 5 5 |
| | 129.50 | 129.62 | 129.82 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 117.10 | 115.56 | 116.38 | |
| | 153.40 | 151.33 | 151.38 | |
| | 112.30 | 110.19 | 110.31 | |
| | 129.50 | 129.71 | 130.30 | |
| ADENOS12 | 154.09 | 154.18 | 154.69 | 5 5 5 |
| | 147.94 | 146.06 | 145.71 | |
| | 119.30 | 119.31 | 120.26 | |
| | 119.30 | 119.31 | 120.26 | |
| | 137.38 | 138.05 | 139.58 | |
| | 91.91 | 92.27 | 91.36 | |
| | 74.61 | 75.37 | 74.95 | |
| | 70.86 | 71.50 | 71.40 | |
| | 84.45 | 84.93 | 84.30 | |
| | 62.45 | 62.62 | 61.70 | |
| PERYTO10 | 50.20 | 50.64 | 50.62 | 5 5 5 |
| | 58.40 | 58.37 | 58.02 | |

Table A.6: Comparison of the experimental and predicted **$^{15}$N isotropic chemical shifts** for the fully optimized structure (Opt cell) and the fixed cell optimization (Expt. Cell). The k-points used for each fully optimized structure are displayed after the calculated chemical shifts.

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| LSERIN01 | -4.10 | -6.18 | -6.31 | 7 3 3 |
| BITZAF | 249.50 | 244.76 | 245.02 | 5 5 5 |
| LHISTD02 | 132.60 | 138.14 | 139.13 | 5 5 5 |
| | 210.80 | 212.52 | 213.61 | |
| LHISTD13 | 132.40 | 138.44 | 138.34 | 5 5 5 |
| | 210.60 | 212.62 | 212.76 | |
| TEJWAG | 143.90 | 146.22 | 146.27 | 5 5 5 |
| FUSVAQ01 | 183.20 | 180.99 | 179.78 | 5 5 5 |
| | 174.20 | 171.29 | 170.70 | |
| | 192.20 | 192.00 | 191.54 | |
| | 120.20 | 123.48 | 126.41 | |
| | 50.20 | 47.69 | 50.14 | |
| CYTSIN | 110.20 | 115.30 | 114.88 | 5 5 5 |
| | 165.20 | 162.80 | 163.42 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| | 54.20 | 56.18 | 58.64 | |
| URACIL | 96.20 | 105.19 | 106.69 | 5 5 5 |
| | 120.20 | 122.17 | 122.04 | |
| BAPLOT01 | 114.70 | 121.97 | 122.69 | 5 5 5 |
| | 72.70 | 77.70 | 79.27 | |
| | 122.70 | 125.71 | 128.09 | |
| | 178.70 | 178.66 | 178.81 | |
| LCYSTN21 | -0.40 | -4.96 | -2.28 | 5 5 5 |
| GLUTAM01 | -1.30 | -3.04 | -3.64 | 5 5 5 |
| | 72.10 | 76.59 | 80.26 | |
| HXACAN13 | 97.91 | 91.51 | 91.53 | 1 3 3 |
| LGLUAC11 | -0.40 | -2.54 | -2.77 | 5 5 5 |
| CYSCLM11 | 1.50 | 0.80 | 0.85 | 5 5 5 |
| ALUCAL04 | 3.00 | -0.11 | -0.47 | 5 5 5 |
| GLYHCL01 | -1.70 | -3.95 | -3.49 | 5 5 5 |
| THYMIN01 | 90.20 | 94.68 | 95.22 | 5 5 5 |
| | 119.20 | 121.06 | 121.53 | |
| LTYRHC10 | 8.00 | 6.25 | 7.12 | 5 5 5 |
| ASPARM03 | 0.70 | -2.03 | -2.13 | 5 5 5 |
| | 74.90 | 78.41 | 79.57 | |

| Crystal | Expt. Shifts Isotropic (ppm) | Expt. Cell (ppm) | Opt. Cell (ppm) | KPOINTS |
|---|---|---|---|---|
| GEHHAD | 261.80 | 264.16 | 263.04 | 5 5 5 |
| | 253.60 | 248.60 | 247.30 | |
| GEHHEH | 187.40 | 177.85 | 177.38 | 5 5 5 |
| | 261.00 | 253.55 | 256.86 | |
| GEHHIL | 268.50 | 270.63 | 274.29 | 5 5 5 |
| | 261.20 | 259.31 | 262.38 | |
| CIMETD | 130.50 | 135.51 | 137.33 | 5 5 5 |
| | 213.10 | 213.83 | 214.78 | |
| | 56.60 | 58.50 | 59.67 | |
| | 43.50 | 44.66 | 46.24 | |
| | 45.80 | 39.10 | 40.48 | |
| | 149.90 | 147.44 | 150.64 | |
| Structure 1 | 223.40 | 219.79 | 220.47 | 3 3 1 |
| | 202.50 | 199.48 | 200.02 | |
| | 76.40 | 74.07 | 75.47 | |
| | 285.10 | 285.38 | 284.44 | |
| | 81.50 | 78.51 | 79.19 | |

## A.4 Thermal expansion curves

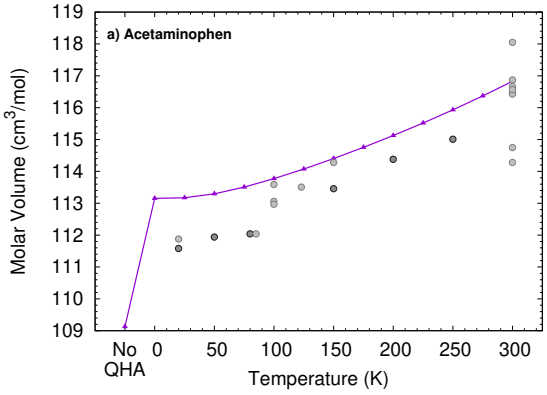
The following plots show the predicted QHA thermal expansion for each crystal and compare them against the experimental data. Experimental volumes were obtained from the following structures from the Cambridge Structure Database (indicated by their RefCodes):

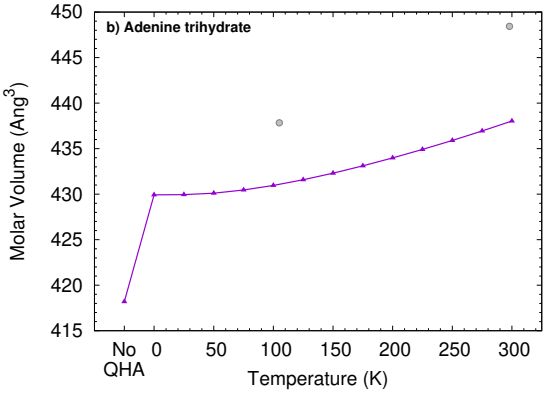
- **Acetaminophen:** HXACAN01, HXACAN02, HXACAN03, HXACAN04, HXACAN05, HXACAN06, HXACAN07, HXACAN13, HXACAN14, HXACAN15, HXACAN16, HXACAN17, HXACAN18, HXACAN19, HXACAN20, HXACAN26, HXACAN27, HXACAN28, HXACAN30, HXACAN34, HXACAN35, HXACAN36
- **Adenosine:** ADENOS01, ADENOS10, ADENOS11, ADENOS12

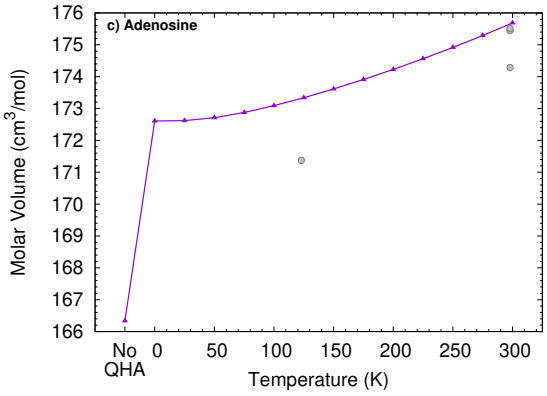
  $\alpha$-Glycine: GLYCIN98, GLYCIN96, GLYCIN97, GLYCIN95, GLYCIN94, GLYCIN93, GLYCIN92, GLYCIN91, GLYCIN89, GLYCIN90, GLYCIN87, GLYCIN88, GLYCIN69, GLYCIN85, GLYCIN86, GLYCIN09, GLYCIN17, GLYCIN30, GLYCIN02, GLYCIN03, GLYCIN04, GLYCIN05, GLYCIN08, GLYCIN19, GLYCIN20, GLYCIN28, GLYCIN29, GLYCIN80, GLYCIN99, GLYCIN21, GLYCIN22, GLYCIN23, GLYCIN07, GLYCIN06, GLYCIN24
- **Carbamazepine:** CBMZPN18, CBMZPN21, CBMZPN22, CBMZPN23, CBMZPN, CBMZPN02, CBMZPN10, CBMZPN14, CBMZPN17, CBMZPN19, and additional structures taken from Ref 235
- **Cytosine:** CYTSIN, CYTSIN01
- **D-Alanine:** ALUCAL05, ALUCAL02, ALUCAL01, ALUCAL04

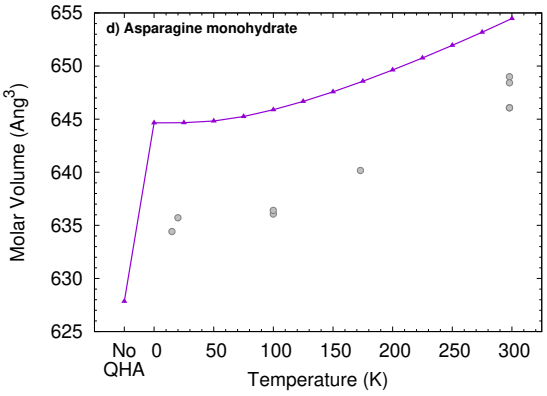
- **Ibuprofen:** IBPRAC, IBPRAC03, IBPRAC16, IBPRAC17, IBPRAC18, IBPRAC19, IBPRAC06
- **L-Adenine trihydrate:** FUSVAQ, FUSVAQ01
- **L-Alanine:** LALNIN03, LALNIN24, LALNIN23, LALNIN52, LALNIN55, LALNIN15, LALNIN14, LALNIN, LALNIN01, LALNIN12, LALNIN22, LALNIN53, LALNIN54, LALNIN56, LALNIN57, LALNIN58
- **L-Asparagine monohydrate:** ASPARM06, ASPARM08, ASPARM05, ASPARM09, ASPARM07, ASPARM03, ASPARM, ASPARM02, ASPARM10
- **L-Cysteine:** LCYSTN22, LCYSTN21, LCYSTN12, LCYSTN23, LCYSTN28, LCYSTN29, LCYSTN30, LCYSTN31, LCYSTN32, LCYSTN33, LCYSTN34, LCYSTN35, LCYSTN36
- **L-Glutamine:** GLUTAM02, GLUTAM, GLUTAM03, GLUTAM01
- **L-Histidine monoclinic polymorph:** LHISTD02
- **L-Serine:** LSERIN20, LSERIN19, LSERIN01, LSERIN10, LSERIN18
- **L-Threonine:** LTHREO03, LTHREO, LTHREO01
- **L-Tyrosine:** LTYROS10, LTYROS11
- **Naphthalene:** NAPHTA31, NAPHTA23, NAPHTA24, NAPHTA32, NAPHTA25, NAPHTA33, NAPHTA26, NAPHTA27, NAPHTA28, NAPHTA34, NAPHTA29, NAPHTA30, NAPHTA35, NAPHTA36, NAPHTA06, NAPHTA15, NAPHTA07, NAPHTA04, NAPHTA16, NAPHTA08, NAPHTA17, NAPHTA09, NAPHTA18 ,NAPHTA10, NAPHTA11, NAPHTA, NAPHTA37

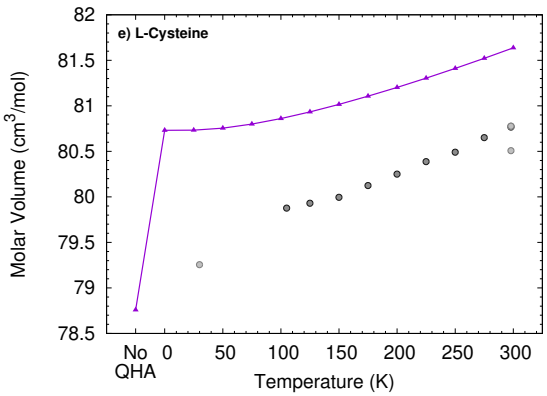
- **Naproxen:** COYRUD12, COYRUD, COYRUD01, COYRUD11, COYRUD13
- **Pyridoxine:** BITZAF
- **Sucrose:** SUCROS14, SUCROS41, SUCROS13, SUCROS01, SUCROS03, SUCROS04, SUCROS08, SUCROS11, SUCROS12, SUCROS15, SUCROS16, SUCROS42
- **Theophylline:** BAPLOT01, BAPLOT06
- **Thymine:** THYMIN, THYMIN01
- **Uracil:** URACIL

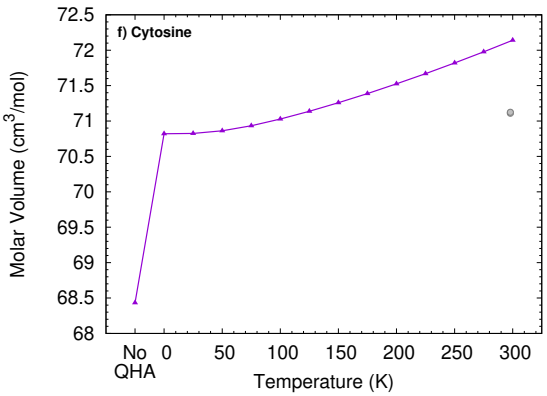

a) Acetaminophen
Molar Volume (cm3/mol)
Temperature (K)
No QHA
0
50
100
150
200
250
300
109
110
111
112
113
114
115
116
117
118
119

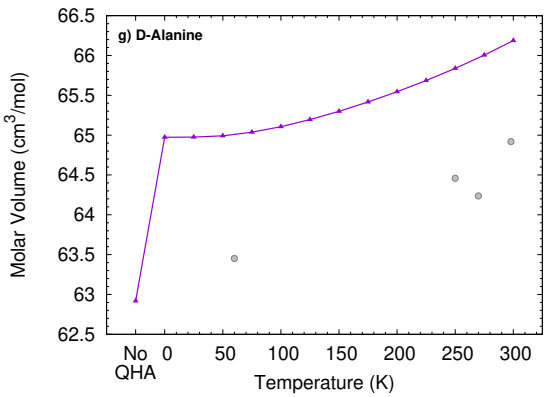

b) Adenine trihydrate
Molar Volume (Ang3)
Temperature (K)
No QHA
0
50
100
150
200
250
300
415
420
425
430
435
440
445
450

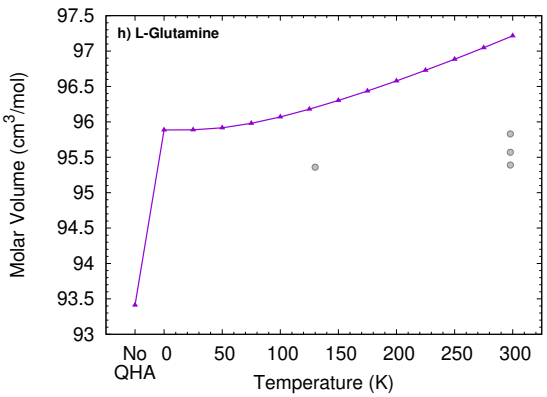

c) Adenosine
Molar Volume (cm3/mol)
Temperature (K)
No QHA
0
50
100
150
200
250
300
166
167
168
169
170
171
172
173
174
175
176

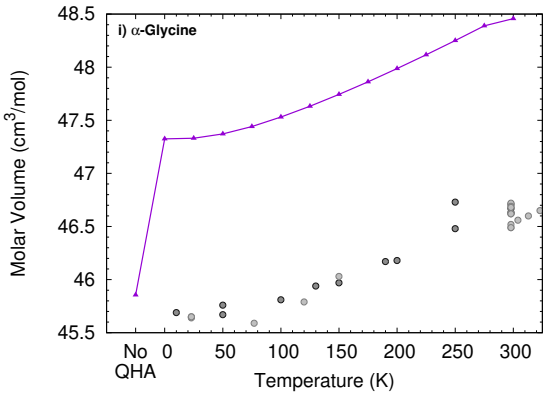

d) Asparagine monohydrate
Molar Volume (Ang3)
Temperature (K)
No QHA
0
50
100
150
200
250
300
625
630
635
640
645
650
655

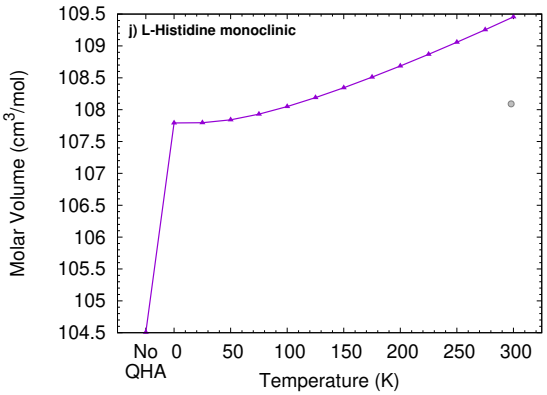

e) L-Cysteine
Molar Volume (cm3/mol)
Temperature (K)
No QHA
0
50
100
150
200
250
300
78.5
79
79.5
80
80.5
81
81.5
82

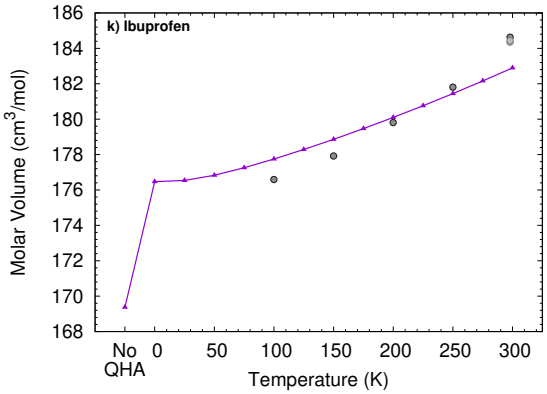

f) Cytosine
Molar Volume (cm3/mol)
Temperature (K)
No QHA
0
50
100
150
200
250
300
68
68.5
69
69.5
70
70.5
71
71.5
72
72.5

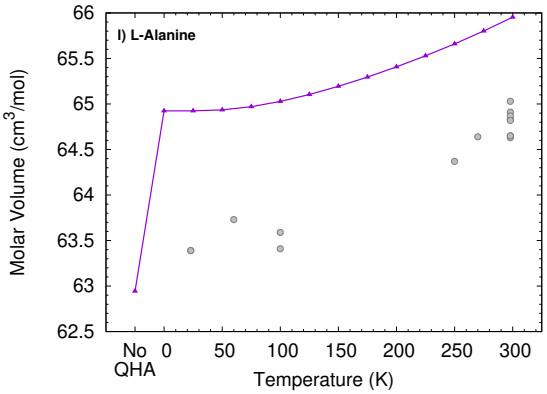

g) D-Alanine
Molar Volume (cm3/mol)
Temperature (K)
No QHA
0
50
100
150
200
250
300
62.5
63
63.5
64
64.5
65
65.5
66
66.5

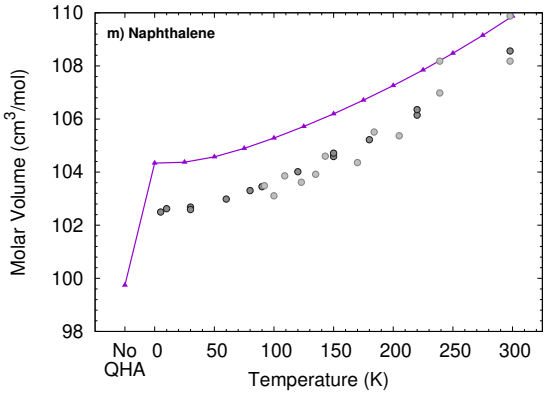

h) L-Glutamine
Molar Volume (cm3/mol)
Temperature (K)
No QHA
0
50
100
150
200
250
300
93
93.5
94
94.5
95
95.5
96
96.5
97
97.5

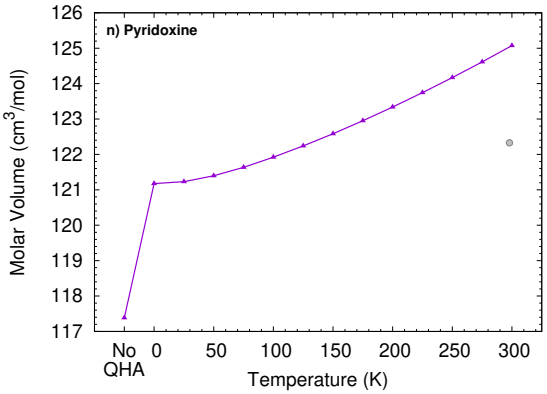

i) α-Glycine
Molar Volume (cm3/mol)
48.5
48
47.5
47
46.5
46
45.5
No QHA
0
50
100
150
200
250
300
Temperature (K)

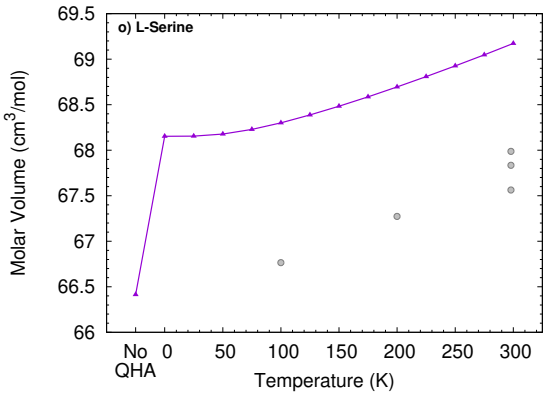

j) L-Histidine monoclinic
Molar Volume (cm3/mol)
109.5
109
108.5
108
107.5
107
106.5
106
105.5
105
104.5
No QHA
0
50
100
150
200
250
300
Temperature (K)

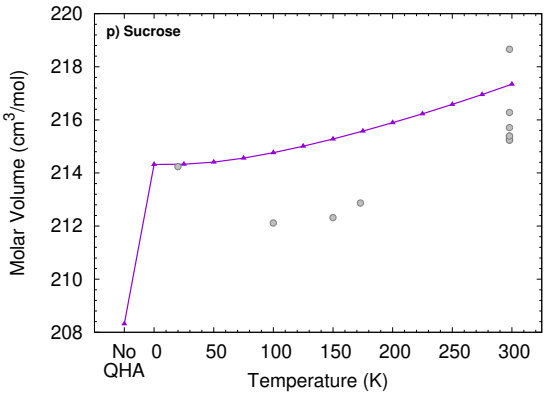

k) Ibuprofen
Molar Volume (cm3/mol)
186
184
182
180
178
176
174
172
170
168
No QHA
0
50
100
150
200
250
300
Temperature (K)


l) L-Alanine
Molar Volume (cm3/mol)
66
65.5
65
64.5
64
63.5
63
62.5
No QHA
0
50
100
150
200
250
300
Temperature (K)


m) Naphthalene
Molar Volume (cm3/mol)
110
108
106
104
102
100
98
No QHA
0
50
100
150
200
250
300
Temperature (K)


n) Pyridoxine
Molar Volume (cm3/mol)
126
125
124
123
122
121
120
119
118
117
No QHA
0
50
100
150
200
250
300
Temperature (K)


o) L-Serine
Molar Volume (cm3/mol)
69.5
69
68.5
68
67.5
67
66.5
66
No QHA
0
50
100
150
200
250
300
Temperature (K)


p) Sucrose
Molar Volume (cm3/mol)
220
218
216
214
212
210
208
No QHA
0
50
100
150
200
250
300
Temperature (K)

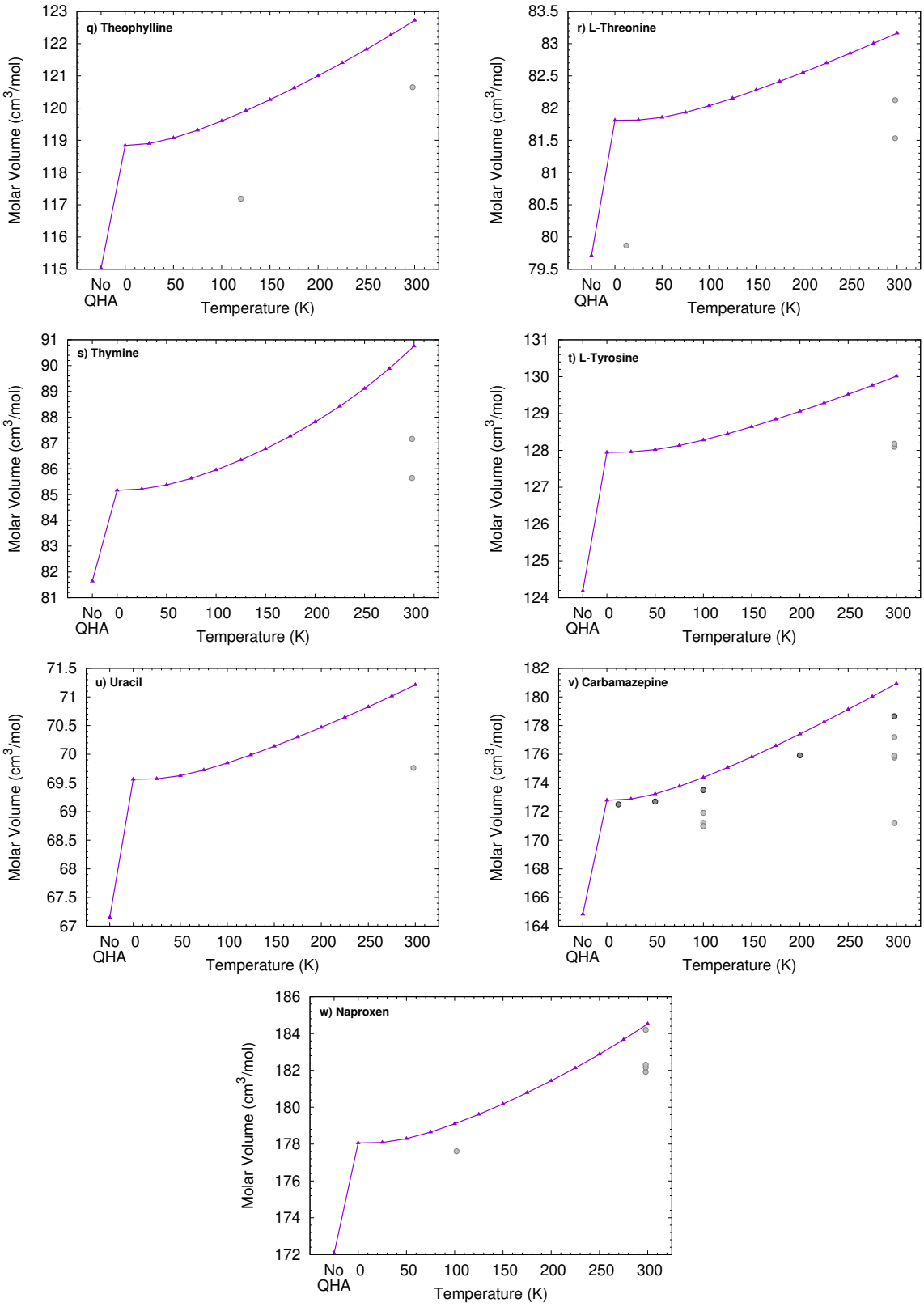

q) Theophylline
r) L-Threonine
s) Thymine
t) L-Tyrosine
u) Uracil
v) Carbamazepine
w) Naproxen
Molar Volume (cm3/mol)
Temperature (K)
No QHA
0
50
100
150
200
250
300



Figure A.4:

## A.5 Correlating Volumes, RMSD, and RMSE

In this section we provide a visual analog of the data that is displayed in Tables 3.1 and 3.2. What this data attempts to do is relate the overall effect the QHA has on a molecular crystal to simple gross descriptors of the crystal and its predicted chemical shifts (See Figures A.5 and A.6). The left plot shows the % change in volume relative to the chosen experimental volumes at room temperatures. The center plot presents the change in root-mean-squared deviation relative to a 15-molecule cluster of the experimental crystal is displayed (rmsd15). Finally, the right figure plots the change in root-mean-squared error (rms error) of predicted chemical shifts relative to known experimental chemical shift data.

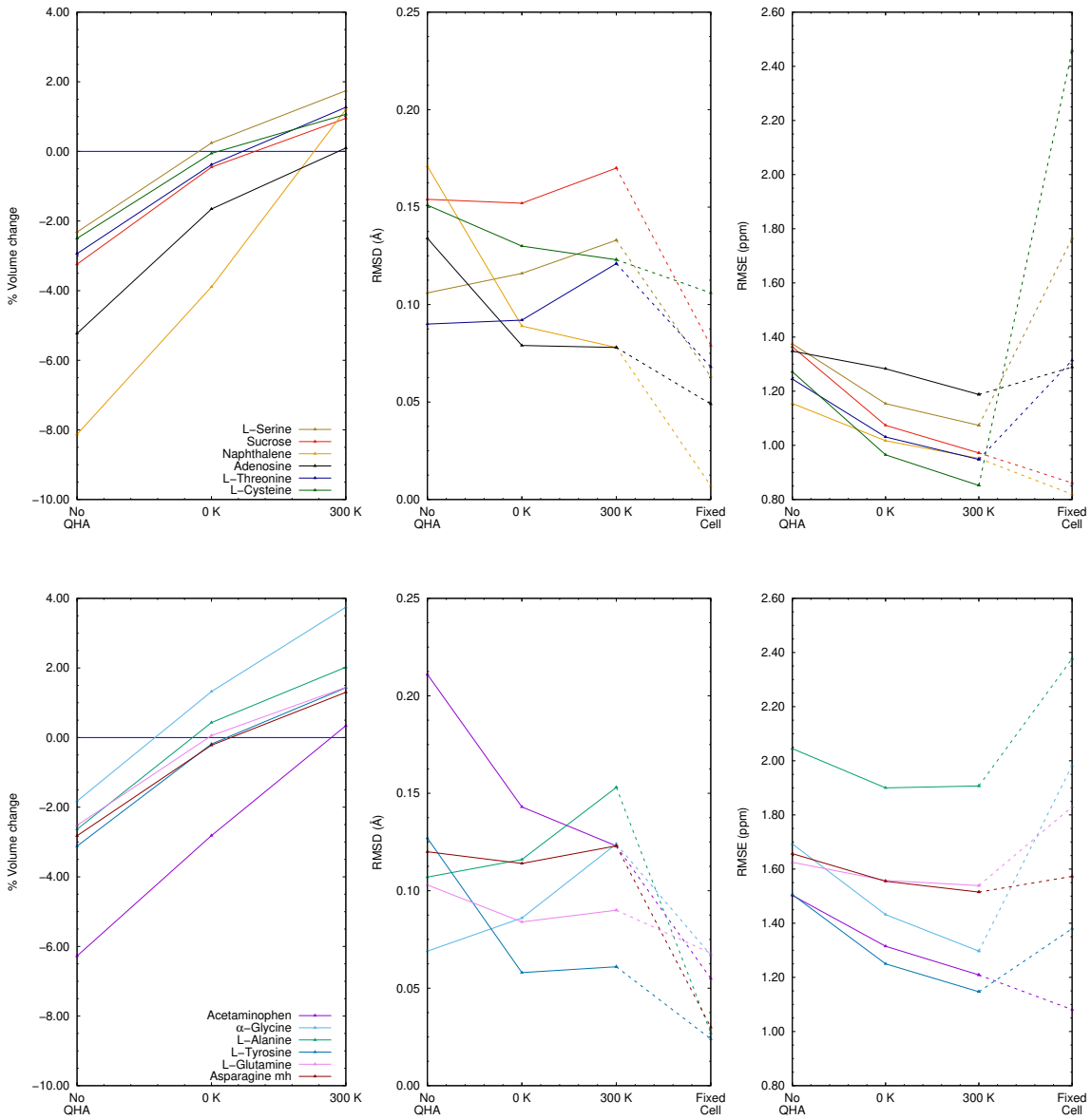



Figure A.5: Showing the effect QHA has on the 12 structures that contributed $^{13}$C chemical shifts.

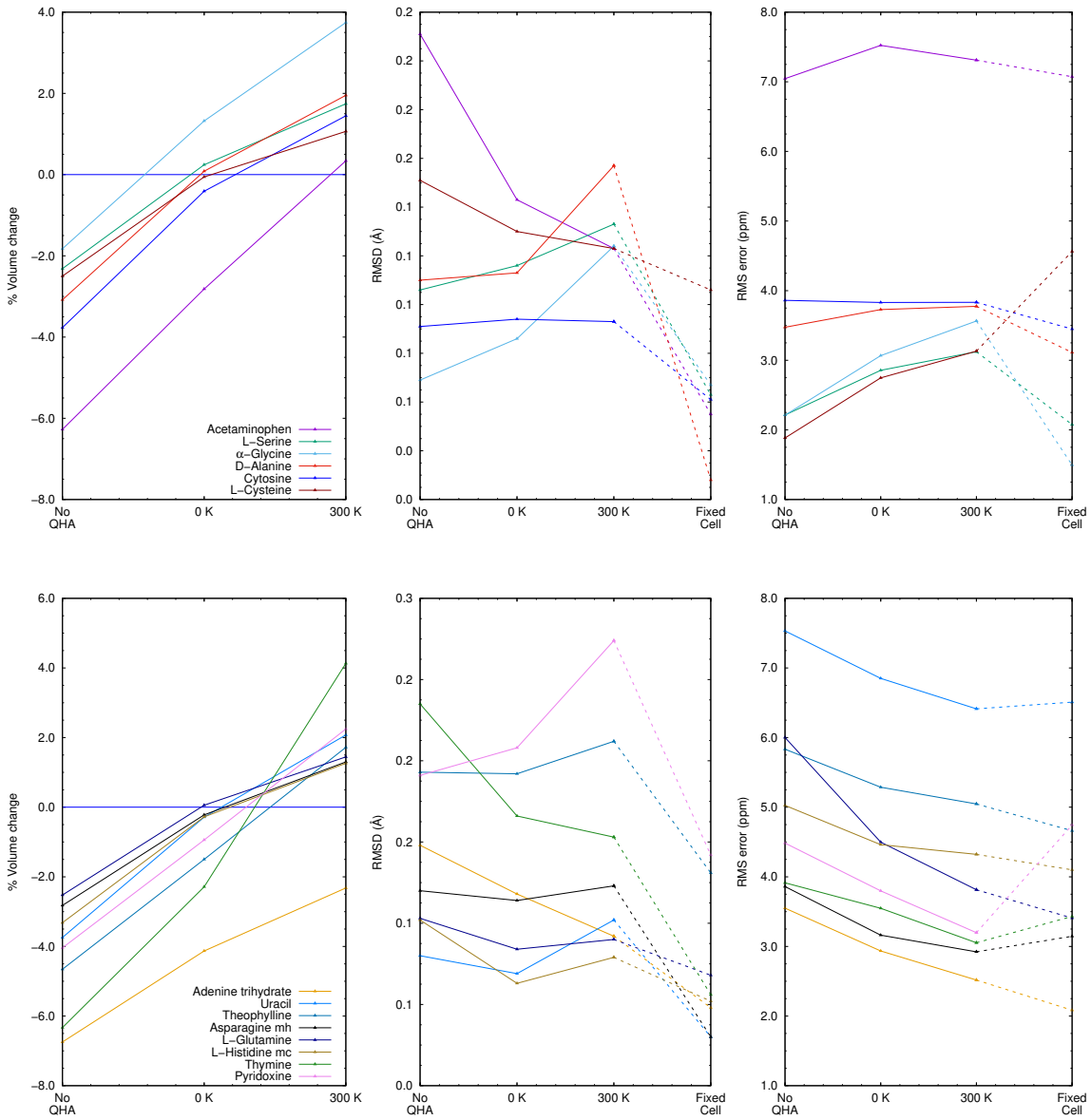



Figure A.6: Showing the effect QHA has on the 14 structures that contributed $^{15}$N chemical shifts.

## A.6 How QHA affects chemical shifts

Below are histograms displaying how the predicted isotropic chemical shifts change from the electronic minimum to 300 K for the species studied in Chapter 3.

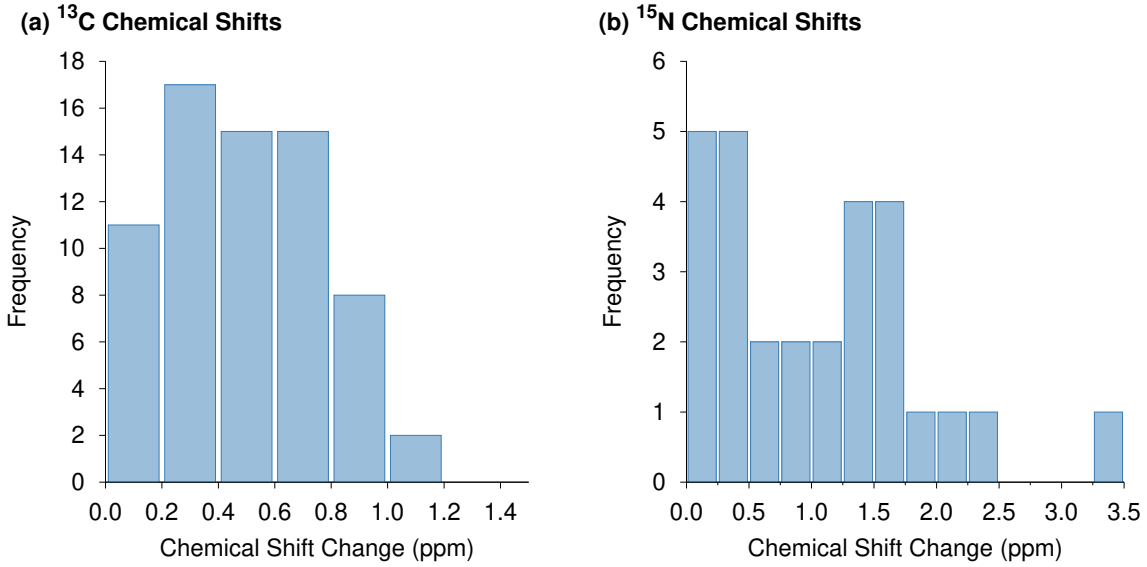



Figure A.7: Chemical shift changes from the electronic minima to 300 K for a) 68 $^{13}$C chemical shifts for 12 molecular crystals and b) 28 $^{15}$N chemical shifts for 14 molecular crystals.

## A.7 Theophylline CSP analysis

### A.7.1 Predicted Isotropic chemical shifts

Table A.7: Comparison of the experimental and predicted **$^1$H isotropic chemical shifts** for the Fully optimized structure (No QHA), QHA-expanded to 0 K (0 K), and QHA-expanded to 300 K (300 K). Calculations were done according to the method in Sec. 3.3.3 in Chapter 3.

| Crystal | Expt. Shifts Isotropic (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|
| Structure 1 | 14.60 | 14.64 | 14.17 | 14.13 |
| | 7.70 | 7.60 | 7.50 | 7.60 |
| | 3.40 | 3.47 | 3.33 | 3.35 |
| | 3.40 | 3.27 | 3.19 | 3.31 |
| Structure 2 | 14.60 | 14.61 | 14.32 | 14.29 |
| | 7.70 | 7.49 | 7.45 | 7.58 |
| | 3.40 | 3.42 | 3.29 | 3.30 |
| | 3.40 | 3.18 | 3.07 | 3.18 |
| Structure 3 | 14.60 | 15.49 | 15.32 | 15.07 |
| | 7.70 | 8.77 | 8.86 | 8.70 |
| | 3.40 | 2.46 | 2.56 | 2.50 |

| Crystal | Expt. Shifts Isotropic (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|
| | 3.40 | 3.37 | 3.45 | 3.37 |
| Structure 4 | 14.60 | 14.75 | 14.72 | 14.52 |
| | 7.70 | 7.60 | 7.87 | 7.77 |
| | 3.40 | 3.06 | 3.19 | 3.15 |
| | 3.40 | 3.38 | 3.47 | 3.44 |
| Structure 5 | 14.60 | 14.70 | 14.42 | 14.33 |
| | 7.70 | 8.61 | 8.43 | 8.40 |
| | 3.40 | 3.36 | 3.37 | 3.36 |
| | 3.40 | 3.44 | 3.37 | 3.44 |

Table A.8: Comparison of the experimental and predicted **$^{13}$C isotropic chemical shifts** for the Fully optimized structure (No QHA), QHA-expanded to 0 K (0 K), QHA-expanded to 300 K (300 K), and the fixed cell optimization (Expt. Cell). Calculations were done according to the method in Sec. 3.3.3 in Chapter 3. The k-points used for each structure are displayed after the calculated chemical shifts.

| Crystal | Expt. Shifts Isotropic (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) | KPOINTS |
|---|---|---|---|---|---|
| Structure 1 | 150.80 | 150.15 | 150.10 | 150.30 | 1 7 5 |
| | 146.10 | 144.99 | 145.72 | 146.36 | |
| | 105.80 | 105.26 | 105.79 | 106.41 | |
| | 155.00 | 154.32 | 154.50 | 154.85 | |
| | 140.80 | 141.12 | 141.22 | 141.51 | |
| | 29.90 | 31.87 | 31.16 | 30.70 | |
| | 29.90 | 29.98 | 29.50 | 29.33 | |
| Structure 2 | 150.80 | 150.03 | 149.90 | 150.06 | 7 1 5 |
| | 146.10 | 145.01 | 145.47 | 146.07 | |
| | 105.80 | 104.07 | 105.82 | 106.81 | |
| | 155.00 | 153.86 | 154.52 | 154.90 | |
| | 140.80 | 141.33 | 140.81 | 141.04 | |

| Crystal | Expt. Shifts Isotropic (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) | KPOINTS |
|---|---|---|---|---|---|
| | 29.90 | 31.41 | 30.41 | 30.01 | |
| | 29.90 | 29.39 | 29.13 | 29.05 | |
| Structure 3 | 150.80 | 149.44 | 149.98 | 149.94 | 3 3 3 |
| | 146.10 | 145.41 | 145.90 | 145.99 | |
| | 105.80 | 108.09 | 108.40 | 108.31 | |
| | 155.00 | 154.26 | 154.53 | 154.49 | |
| | 140.80 | 141.63 | 141.72 | 141.61 | |
| | 29.90 | 30.05 | 29.61 | 28.95 | |
| | 29.90 | 30.75 | 30.35 | 29.57 | |
| Structure 4 | 150.80 | 149.88 | 150.21 | 150.24 | 1 7 3 |
| | 146.10 | 144.69 | 145.67 | 146.22 | |
| | 105.80 | 106.90 | 107.74 | 108.06 | |
| | 155.00 | 153.87 | 154.08 | 153.99 | |
| | 140.80 | 142.88 | 143.42 | 143.15 | |
| | 29.90 | 30.72 | 30.18 | 29.49 | |
| | 29.90 | 30.98 | 30.44 | 29.81 | |
| Structure 5 | 150.80 | 149.40 | 149.46 | 149.59 | 1 3 7 |
| | 146.10 | 144.58 | 144.93 | 145.45 | |
| | 105.80 | 105.04 | 105.37 | 106.00 | |
| | 155.00 | 153.22 | 153.48 | 153.77 | |

| Crystal | Expt. Shifts Isotropic (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) | KPOINTS |
|---|---|---|---|---|---|
| | 140.80 | 142.00 | 141.54 | 141.39 | |
| | 29.90 | 30.74 | 30.08 | 29.54 | |
| | 29.90 | 29.80 | 29.39 | 29.21 | |

Table A.9: Comparison of the experimental and predicted **$^{15}$N isotropic chemical shifts** for the Fully optimized structure (No QHA), QHA-expanded to 0 K (0 K), QHA-expanded to 300 K (300 K), and the fixed cell optimization (Expt. Cell). Calculations were done according to the method in Sec. 3.3.3 in Chapter 3.

| Crystal | Expt. Shifts Isotropic (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|
| Structure 1 | 114.70 | 122.51 | 122.14 | 122.08 |
| | 72.70 | 79.18 | 78.71 | 78.53 |
| | 122.70 | 127.80 | 126.35 | 125.71 |
| | 178.70 | 179.68 | 180.26 | 180.93 |
| Structure 2 | 114.70 | 121.71 | 121.34 | 121.13 |
| | 72.70 | 79.30 | 78.15 | 77.68 |

| Crystal | Expt. Shifts Isotropic (ppm) | No QHA (ppm) | 0 K (ppm) | 300 K (ppm) |
|---|---|---|---|---|
| | 122.70 | 128.31 | 127.27 | 126.36 |
| | 178.70 | 178.91 | 179.36 | 180.17 |
| Structure 3 | 114.70 | 118.11 | 118.16 | 118.18 |
| | 72.70 | 79.84 | 79.56 | 78.95 |
| | 122.70 | 130.02 | 128.49 | 127.19 |
| | 178.70 | 184.54 | 185.33 | 185.27 |
| Structure 4 | 114.70 | 120.03 | 120.08 | 119.96 |
| | 72.70 | 79.70 | 79.58 | 79.38 |
| | 122.70 | 126.19 | 125.42 | 124.97 |
| | 178.70 | 178.73 | 179.78 | 180.58 |
| Structure 5 | 114.70 | 120.63 | 120.62 | 120.45 |
| | 72.70 | 80.14 | 79.59 | 79.27 |
| | 122.70 | 127.21 | 126.53 | 126.11 |
| | 178.70 | 184.18 | 184.66 | 184.85 |

### A.7.2 Thermal expansion curves

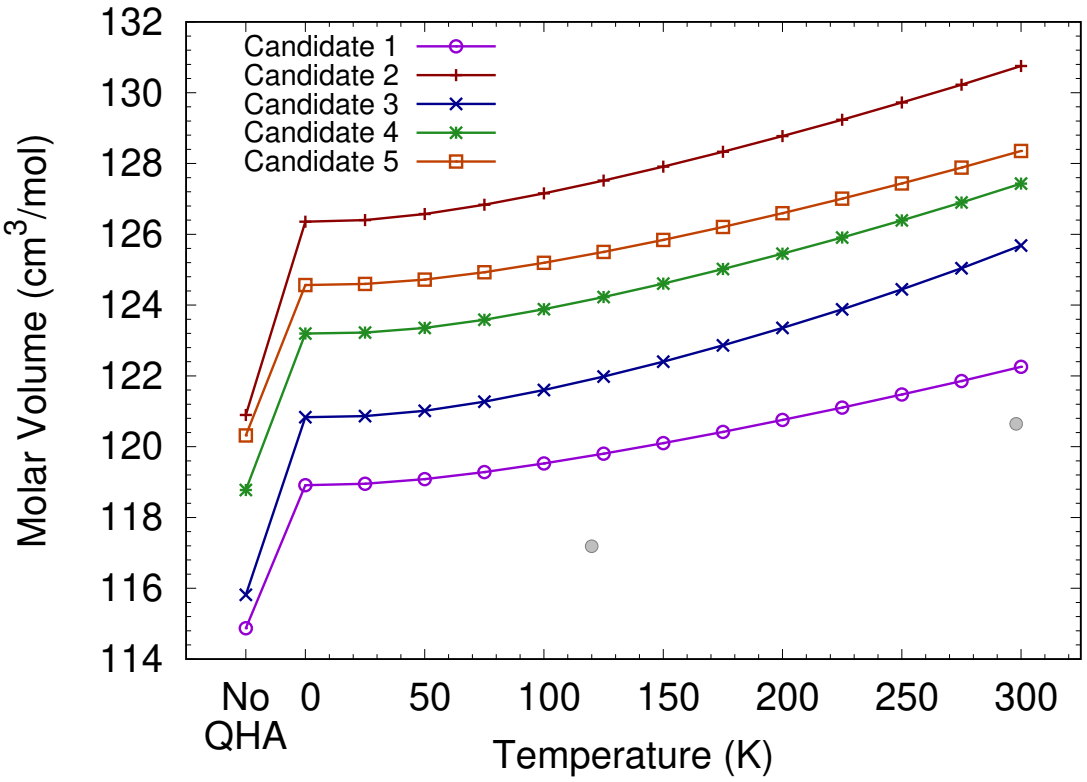



Figure A.8: Displaying the thermal expansion curves for all 5 candidate structures of Theophylline.

## A.8 Volume expansion of 44 molecular crystals

Listed below are the structures whose volume change is displayed in Figure 3.1:

GLYCIN03, LALNIN12, LSERIN01, LTYROS11, LCYSTN21, MGLUCP11, MBDGAL02, MEMANP11, MGALPY01, XYLOBM01, FRUCTO02, RHAMAH12, SUCROS04, GLUTAM01, ASPARM03, LSERMH10, LTHREO01, NAPHTA36, ACENAP03, TRIPHE11, HXACAN26 (with fractional coordinates from HXACAN13), INDMET, SULAMD06, ADENOS12, PERYTO10, BITZAF, LHISTD02, LHISTD13, TEJWAG, FUSVAQ01, CYTSIN, URACIL, BAPLOT01, LGLUAC11, CYSCLM11, ALUCAL04, GLYHCL01, THYMIN01, LTYRHC10, GEHHAD, GEHHEH, GEHHIL, CIMETD, Structure 1

# Appendix B

# Identifying pragmatic quasi-harmonic electronic structure approaches for modeling molecular crystal thermal expansion

## B.1 Construction of crystalline Gibbs free energies

The crystalline Gibbs free energies at various temperatures are constructed by combining the temperature-independent electronic energy curves (e.g. Figure B.1a,d) with the Helmholtz vibrational free energy $F_{vib}$ (e.g. Figure B.1b,e) to give the free energy curves shown in Figure B.1c,f. The optimal volume occurs at the minimum of the free energy curve. For carbon dioxide, the Tier 1–3 energy-volume curves are nearly identical, but the

differences in the phonons lead to modest shifts in the position of the minima. At 0 K, Tiers 1–3 still exhibit almost the same minimum, but by 190 K, the optimal volumes differ somewhat.

Analogous plots are shown for ice at 0 K and 200 K in Figure B.2. Ice has the unusual behavior that Tier 3 exhibits volume contraction upon adding the zero-point vibrational contribution. The reasons for this behavior can be seen in the plots. The optimal volume for the electronic energy curves of ice varies, with Tier 4 < Tier 1/2 < Tier 3. The MP2 Helmholtz vibrational free energy contribution decreases monotonically with volume, while the $F_{vib}$ derived from B86bPBE-XDM phonons using either MP2 or DFT geometries is quite flat and exhibits a shallow minimum. In the Tier 2 and Tier 4 cases, the minimum of $F_{vib}$ occurs at a volume greater than the $E(V)$ minimum, so adding $F_{vib}$ to $E(V)$ to form the Gibbs free energy shifts the system toward larger volumes. In Tier 3, on the other hand, the minimum in $E(V)$ occurs at a larger volume than that of the B86bPBE-XDM $F_{vib}$, so the zero-point vibrational energy contribution shifts it toward smaller volumes, as seen in Figure 4.2 in Chapter 4. The large difference in slopes of $F_{vib}$ largely accounts for the differences in thermal expansion between Tier 1 and the other Tiers. As the crystal is heated, the minimum in $F_{vib}$ shifts toward larger volumes (e.g. see 200 K plots). Therefore, thermal expansion occurs as usual at higher temperatures.

Figures B.3 and B.4 present similar plots for acetic acid and imidazole. For acetic acid, the energy-volume and Helmholtz vibrational free energies are quite similar across the different tiers. For imidazole, the DFT results differ appreciably from MP2 due to the

problems MP2 has describing the van der Waals dispersion in that crystal.

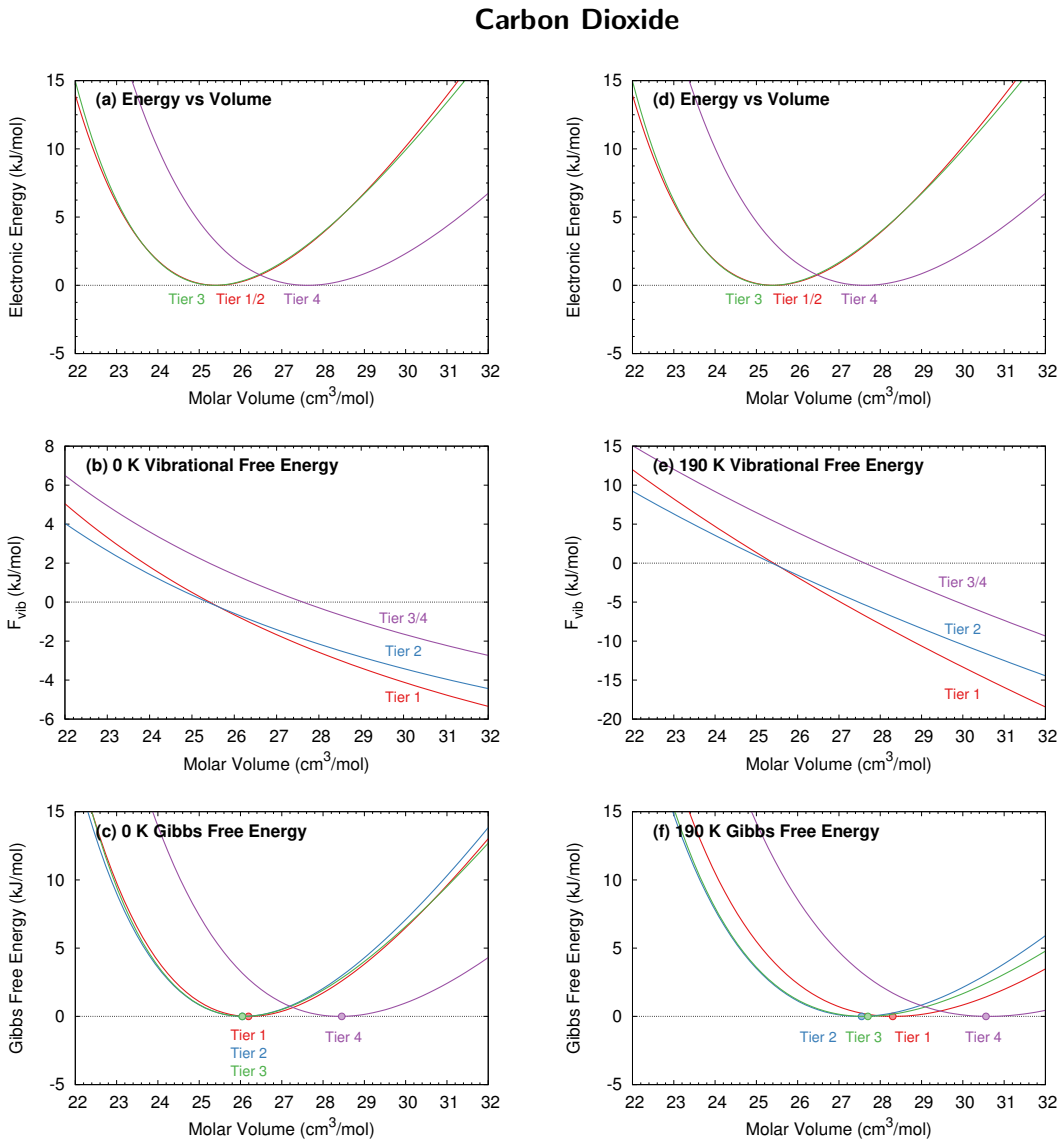



Figure B.1: (a) Electronic energy versus volume curves, (b) 0 K vibrational free energies, and (c) total Gibbs free energies per unit cell with different mixtures of MP2/aug-cc-pVTZ + AMOEBA and B86bPBE-XDM. Plots (d)–(f) are the same but at 190 K. Energies correspond to the full unit cell, and arbitrary vertical offsets were employed to each curve for ease of viewing. Points indicate minima.

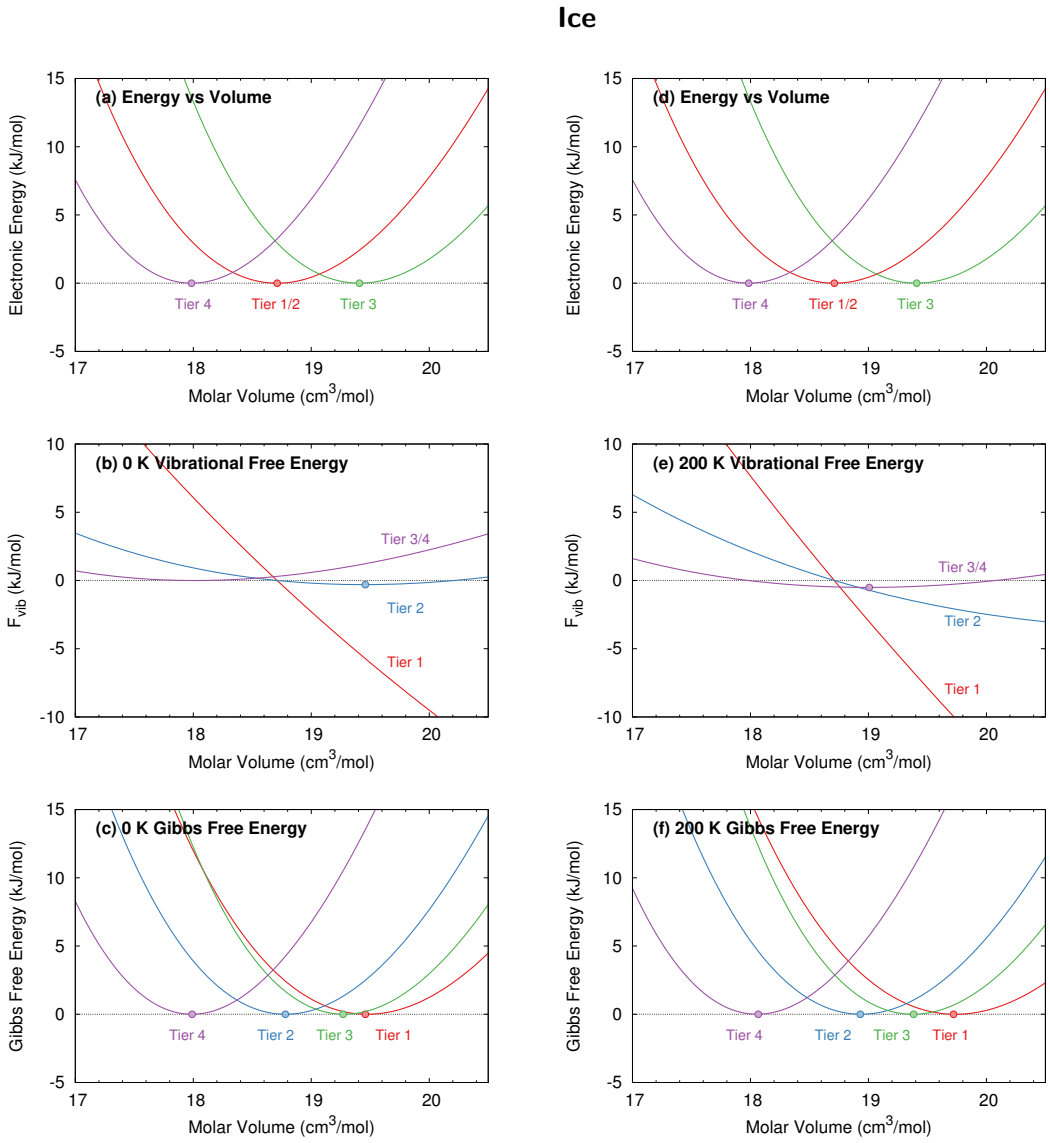



Figure B.2: (a) Electronic energy versus volume curves, (b) 0 K vibrational free energies, and (c) total Gibbs free energies per unit cell with different mixtures of MP2/aug-cc-pVTZ + AMOEBA and B86bPBE-XDM. Plots (d)–(f) are the same but at 200 K. Energies correspond to the full unit cell, and arbitrary vertical offsets were employed to each curve for ease of viewing. Points indicate minima.

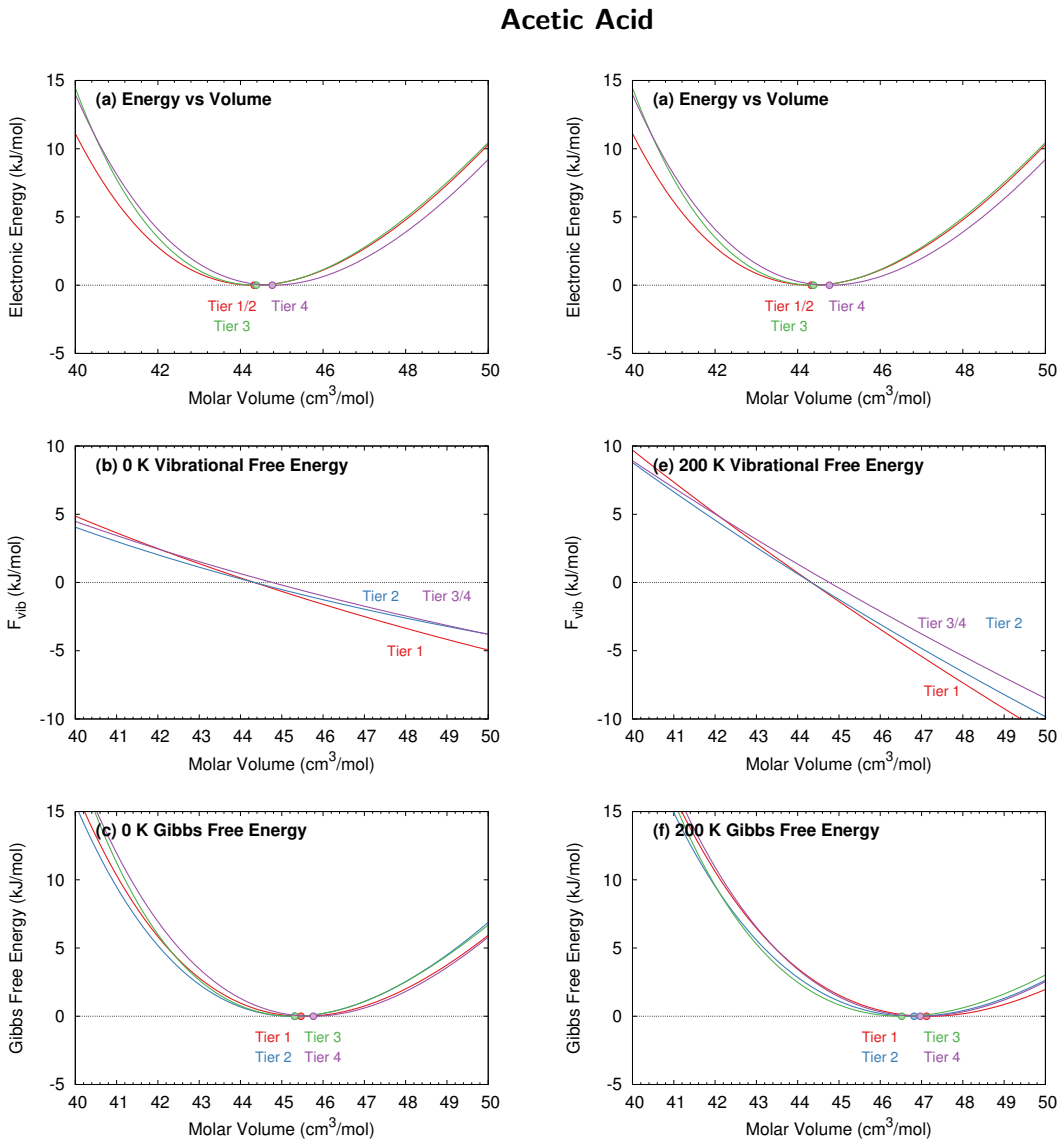



Figure B.3: (a) Electronic energy versus volume curves, (b) 0 K vibrational free energies, and (c) total Gibbs free energies per unit cell with different mixtures of MP2/aug-cc-pVTZ + AMOEBA and B86bPBE-XDM. Plots (d)–(f) are the same but at 200 K. Energies correspond to the full unit cell, and arbitrary vertical offsets were employed to each curve for ease of viewing. Points indicate minima.

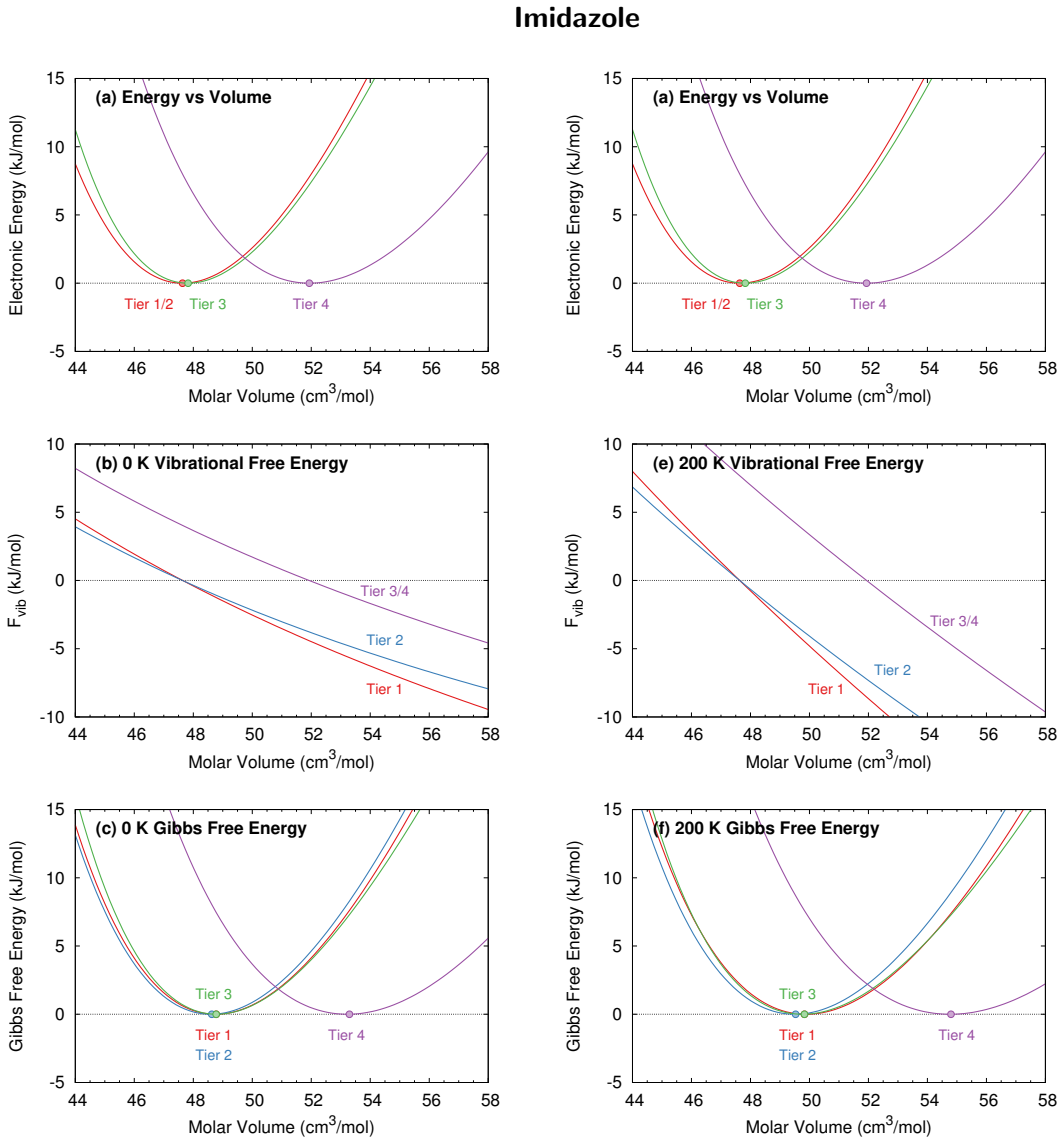



Figure B.4: (a) Electronic energy versus volume curves, (b) 0 K vibrational free energies, and (c) total Gibbs free energies per unit cell with different mixtures of MP2/aug-cc-pVTZ + AMOEBA and B86bPBE-XDM. Plots (d)–(f) are the same but at 200 K. Energies correspond to the full unit cell, and arbitrary vertical offsets were employed to each curve for ease of viewing. Points indicate minima.

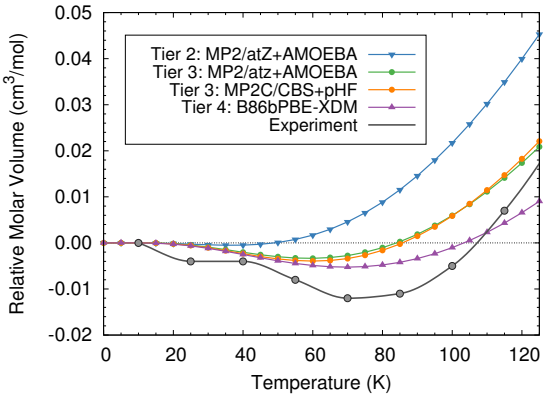



Figure B.5: Comparison of predicted and experimental molar volumes for ice showing the degree to which various models predict the negative thermal expansion at low temperatures.

## B.2 Negative thermal expansion in ice Ih

Figure B.5 focuses on the temperature dependence of the molar volume of ice in the low-temperature region, highlighting the negative thermal expansion. For ease of comparison, the volumes are plotted relative to the lowest-temperature volume for each case.

Experimentally, ice contracts by 0.01 $cm^3$/mol (0.06%) upon heating to around 70 K. As shown here, all the models do predict some contraction, though it is barely noticeable for Tier 2. On the other hand, Tier 3 and Tier 4 predict the temperature of maximum contraction fairly well, though they underestimate the amount of contraction by a factor of 2–3.

Table B.1: Comparison of experimental and predicted % contraction of the unit cell from 10 K - 70 K

| Method | Source | % Contraction |
|---|---|---|
| Experiment | Ref 244 | 0.06 |
| B86bPBE-XDM | Tier 4 | 0.029 |
| MP2/aTZ+AMOEBA | Tier 3 | 0.017 |
| MP2/CBS+pHF | Tier 3 | 0.023 |
| MP2C/CBS+pHF | Tier 3 | 0.021 |
| MP2/aTZ+AMOEBA | Tier 2 | 0.003 |
| MP2/aTZ+AMOEBA | Tier 1 | 0.010 |

## B.3 Sensitivity to choice of density functional

To complement the carbon dioxide results shown in Figure 4.4 in Chapter 4, Figure B.6 compares the energy-volume curves and Helmholtz vibrational free energies of crystalline carbon dioxide at 150 K across several different density functionals. While the DFT $E(V)$ curves vary considerably with the choice of functional, the $F_{vib}$ curves exhibit generally similar slopes. This helps explain why the Tier 2 and 3 molar volumes less sensitive to the choice of density functional used to generate the phonons.

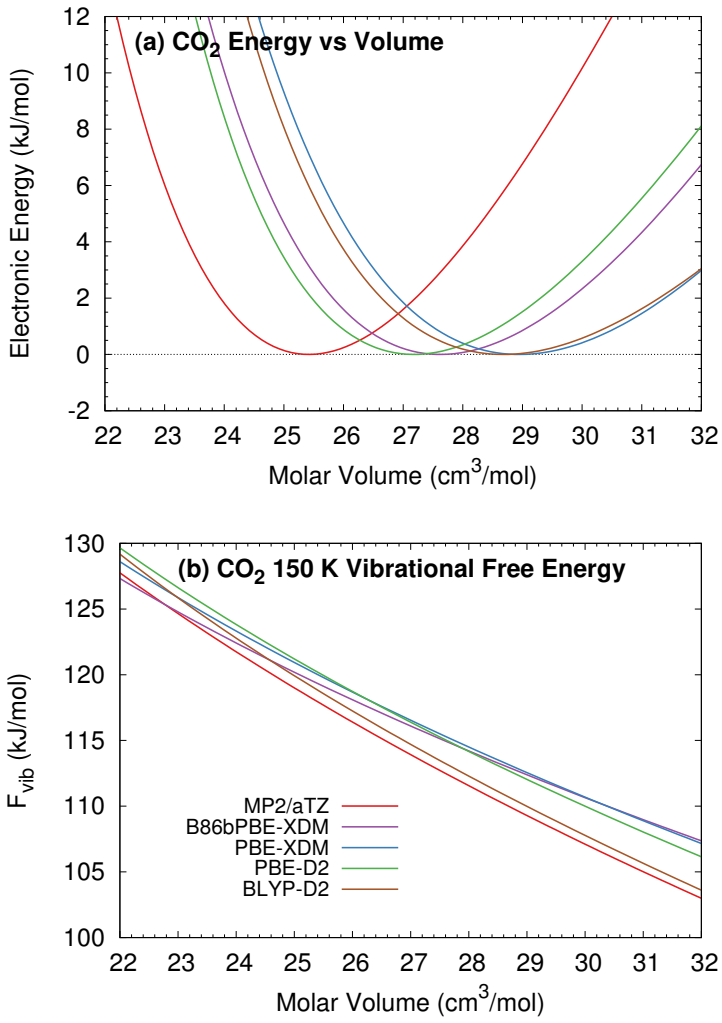



Figure B.6: Variations in (a) the energy-volume curves and (b) 150 K Helmholtz vibrational free energies of carbon dioxide with several different density functionals. For ease of comparison, the energy-volume curves in (a) were shifted relative to their individual minimum energies.

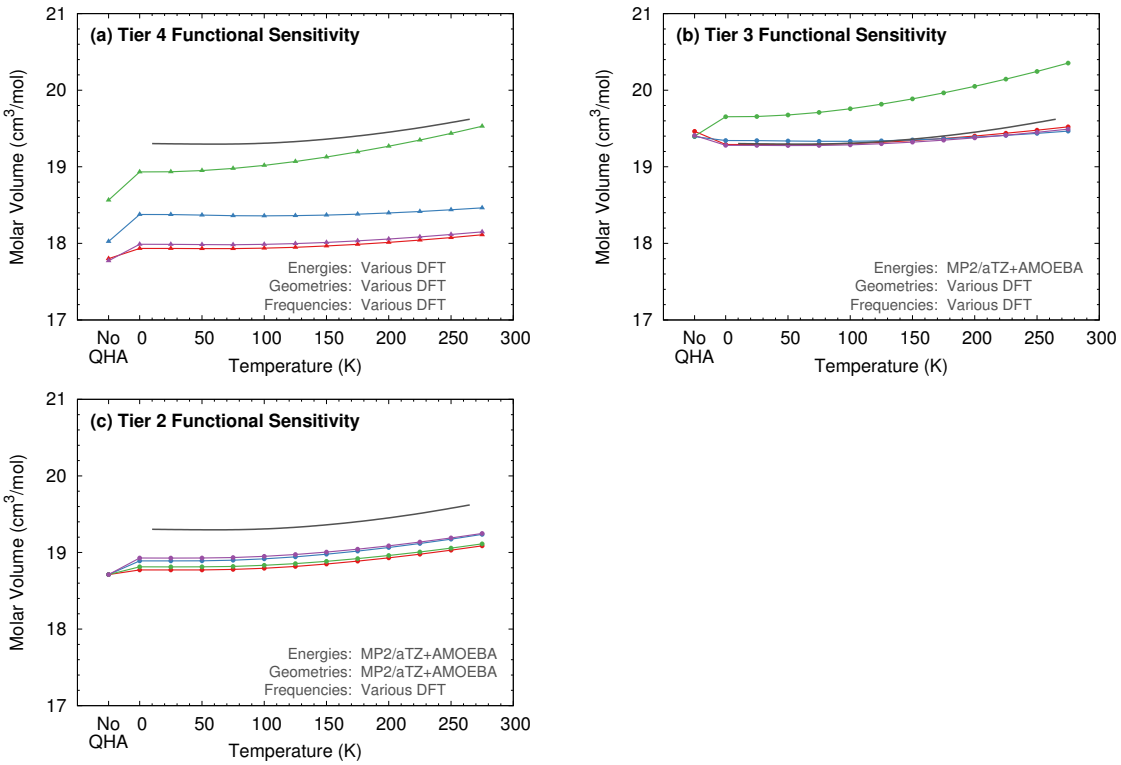



Figure B.7: Sensitivity of the predicted ice molar volume to the density functional used for the low level in Tiers 2–4. Density functionals: PBE-XDM (red), BLYP-XDM (blue), BLYP-D2 (green), and B86bPBE-XDM (purple).

To complement the carbon dioxide data presented in Figure 4.4 in Chapter 4, Figure B.7 plots how ice molar volumes predicted with Tiers 2-4 vary with the choice of density functional. As for carbon dioxide, Tiers 2 and 3 are much less sensitive to the specific density functional used than Tier 4.

## B.4 Tier 3 energy refinement

As discussed in Chapter 4, different Tier 3 single-point energy refinements lead to molar volume curves that are largely parallel. This indicates that while the refinement does shift the energy and volume at which the minimum occurs, it does not significantly alter the well curvature.

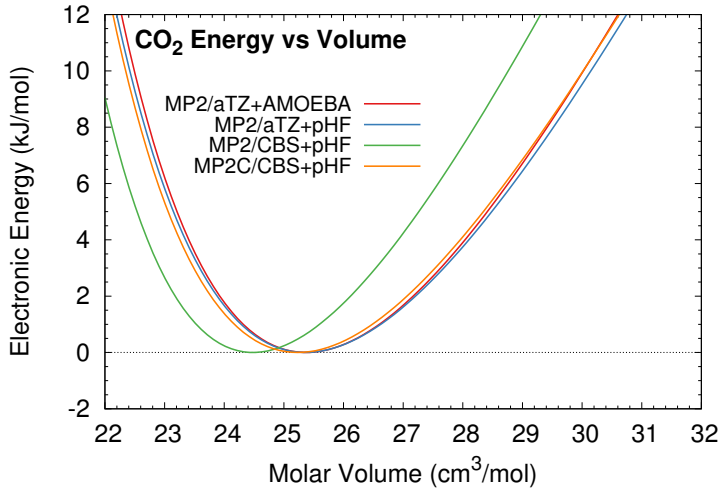



Figure B.8: The Tier 3 single-point energy refinements for carbon dioxide do not significantly alter the curvature of the electronic energy well. For ease of comparison, the volumes are plotted relative to the lowest-temperature volume for each case.

# Appendix C

# Investigating the phase transition of $\alpha-$ and $\beta-$ Resorcinol

## C.1 Electronic energy curve

Figure C.1 plots the electronic energy curves for both alpha and beta resorcinol. For the alpha form the expansion branch for the B86bPBE-XDM is similar to the MP2C/CBS+pHF potential energy surface while the compression branch is similar to the MP2C/CBS+AMOEBA. The slightly raised expansion branch will cause the predicted volumes for the MP2C/CBS+AMOEBA structures to expand more slowly while the more aggressive raising of the MP2C/CBS+pHF compression branch will cause the predicted volumes on this structure to compress more slowly.

In comparison for the beta polymorph the B86bPBE-XDM and the MP2C/CBS+pHF energy surfaces are practically identical on the compression branch while on the expan-

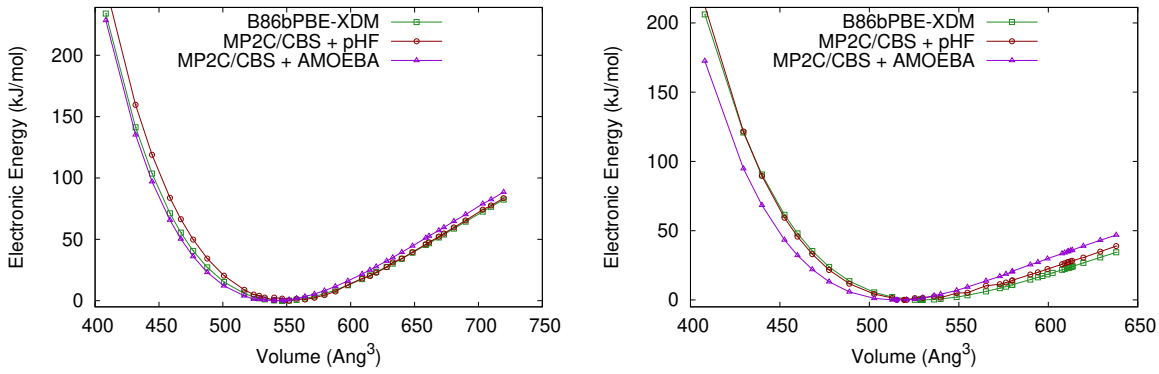



Figure C.1: Electronic energy surfaces for $\alpha$-resorcinol (left) and $\beta$-resorcinol (right). The point which minimizes each structure is shaded in the figure.

sion branch the MP2C/CBS+pHF expands slightly more aggressively. Switching out the many-body correction for AMOEBA causes a shift in both the expansion and compression branches indicating that the beta polymorph should be slightly smaller at all volumes than the B86bPBE-XDM and MP2C/CBS+pHF predicted volumes.

## C.2 Lattice parameter prediction

Once the Gibbs free energy equation of state $G(V)$ has been obtained at a chosen temperature, the free energy can be minimized to find the optimal molar volume. The lattice constants and atomic positions at this optimal volume are interpolated based on the explicitly optimized structures that were obtaining in generating electronic energy surface. Figure C.2 reports the lattice parameters at each level of theory for both $\alpha$- and $\beta-$resorcinol.

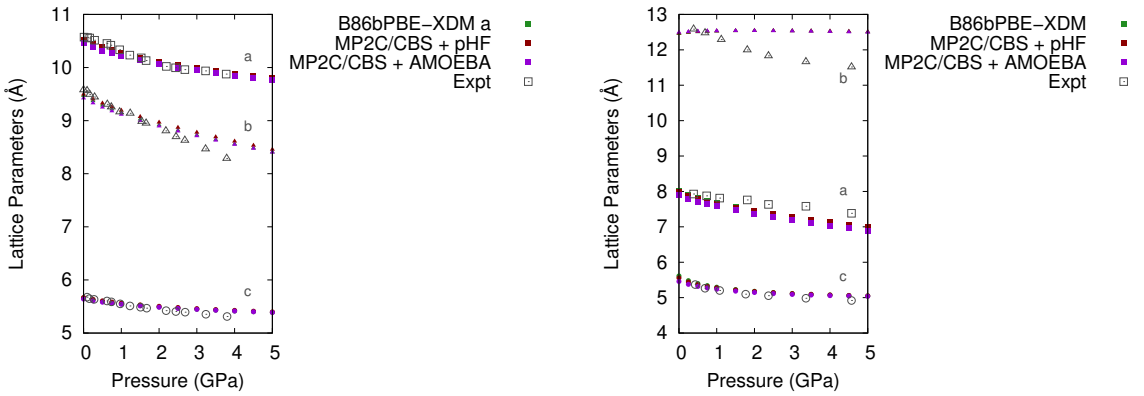



Figure C.2: Predicted lattice parameters at multiple levels of theory for $\alpha$-resorcinol (on the left) and $\beta$-resorcinol (on the right).

## C.3 How pressure and temperature affects the minima on the Gibbs surface

For the pressure curve the minima is well-resolved up to 4.0 GPa with clear indications that the structure could be pushed to higher pressures if need be. At 0.0 GPa the gibbs minima moves $\sim$ 25 Å$^3$ from 0 K to 500 K. While it appears the volumes are oversampled on the expansion branch molecular crystals can quite often experience aggressive expansion between the electronic minima and higher temperature so it was necessary to pad the expansion branch to ensure a minima always existed. The same behavior can be seen for the beta polymorph.

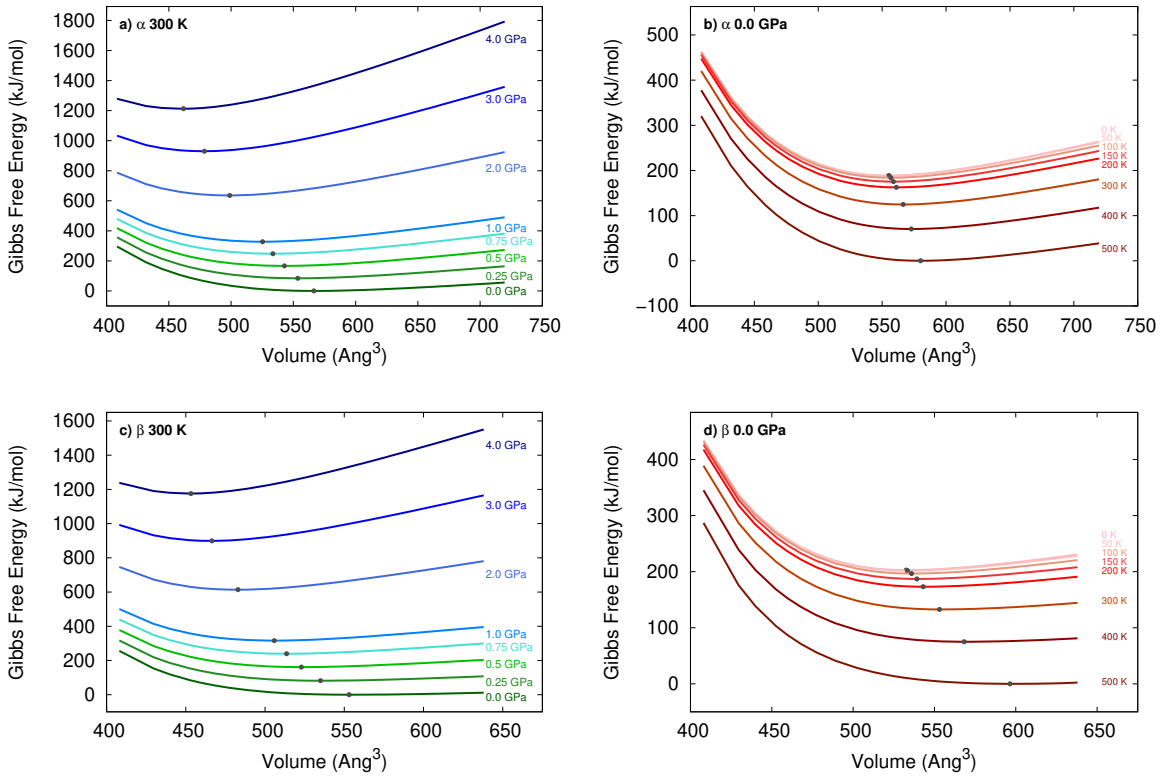



Figure C.3: How the minimum gibbs volume for $\alpha-$ and $\beta-$resorcinol moves upon a,c) applying pressure at a fixed temperature of 300 K and b,d) increasing temperature at a fixed pressure of 0.0 GPa. The minimum volume is marked in gray on each surface and the relevant temperatures and pressures are labeled in the plot.

# Appendix D

# Pedagogical Quasi-Harmonic Approximation

In this section, I will outline some practical considerations that need to be taken into account when attempting to use the Tiered Quasi-Harmonic Approximation (QHA) method that is introduced in Chapter 4. This method was originally developed in an attempt to address some of the problems associated with using the classical QHA (most importantly the computational cost). Previously, the full QHA (optimizations, frequencies, and thermal expansion) would be calculated at the desired level of theory, such as CCSD(T)/CBS, which quickly became computationally prohibitive. Chapter 4 demonstrates that it is often acceptable to perform structure optimizations and frequency calculations using a cheaper theory (e.g. DFT) and single-point energy correct up to the desired level of theory.

Ultimately this method works by breaking up the contributions of the Gibbs free energy into different levels of theory. From statistical thermodynamics, the Gibbs free

energy combines the electronic internal energy $U_{el}$, the Helmholtz vibrational free energy $F_{vib}$, and a pressure-volume ($PV$) contribution.

$$G(T,P) = U_{el} + F_{vib}(T) + PV \tag{D.1}$$

In crystals at ambient pressure, the $PV$ term contributes negligibly and hence is often neglected. In these cases, the Gibbs free energy would simplify to the Helmholtz free energy. The Helmholtz vibrational free energy is computed from standard harmonic oscillator vibrational partition functions as,

$$F_{vib}(T) = N_a \sum_i \left( \frac{\hbar\omega_i}{2} + k_b T \ln\left[1 - \exp\left(-\frac{\hbar\omega_i}{k_b T}\right)\right]\right) \tag{D.2}$$

where $N_a$ is Avogadro's number, $\hbar$ is Plank's constant, $k_b$ is the Boltzmann constant, and $\omega_i$ is the vibrational frequency of mode $i$. The first term corresponds to the zero-point vibrational contribution, while the second gives the thermal vibrational contribution.

In principle, one ought to re-evaluate the phonons at every volume, but that is computationally impractical given the cost of typical electronic structure theory calculations. Instead we employ mode-specific Grüneisen parameters $\gamma_i$ to estimate how individual phonon modes vary with unit cell volume,

$$\gamma_i = -\left(\frac{\partial \ln \omega_i}{\partial \ln V}\right) \tag{D.3}$$

Integrating Eq D.3 yields,

$$\omega_i = \omega_i^{ref} \left(\frac{V}{V^{ref}}\right)^{-\gamma_i} \tag{D.4}$$

which allows the reference phonons $\omega_i^{ref}$ to be scaled to any given new volume $V$.

As the electronic internal energy has the largest contribution to the Gibbs free energy, this will be calculated at the desired level of theory. Typically, this is done using the fragment-based hybrid many-body interaction model (HMBI):

$$U_{el}^{HMBI} = E_{1-body}^{QM} + E_{SR\ 2-body}^{QM} + E_{LR\ 2-body}^{MM} + E_{many\ body}^{MM} \tag{D.5}$$

HMBI treats individual molecules in the unit cell and their short-range pairwise interactions quantum mechanically (QM), while the longer-range and many-body interactions are typically approximated using a classical molecular mechanics (MM) polarizable force field. In some cases, the MM terms will be computed from periodic Hartree-Fock (HF) instead, in which case HMBI is equivalent to the method of increments.

The following sections will outline how to go about performing these calculations, problems to look out for, and ultimately what you should expect to see as a result of your efforts. Section D.1 outlines how to generate an electronic energy curve at the desired level of theory. Section D.2 explains which structures to calculate harmonic frequencies on in order to generate the Grüneisen parameters. Section D.3 outlines the various fits and calculations that are carried out in MATLAB. Finally, Section D.4.1 will outline how to obtain the volumes which minimized the Gibbs free energy surface at any given temperature and pressure. While this method currently requires a lot of human supervision we hope in the future to make this more user-friendly. For a more practical outline of how to run these calculations using HMBI see the HMBI manual.

## D.1 Calculating the Electronic Internal Energy Curve ($U_{el}$)

### D.1.1 Obtaining the Reference geometry

To begin, you must optimize the lattice parameters and atomic positions of your structure. This can be done with any method, but we recommend optimizing using Quantum Espresso and the B86bPBE pseudopotential with the XDM dispersion correction.[2–4,201] Prepare the input file and optimize the geometry. This optimized structure will be referred to as the reference structure from now on. It is recommended that harmonic phonons are calculated on this structure before generating the electronic potential energy surface to ensure the reference structure is fully optimized. These reference phonons will also be used later along with the Grüneisen parameters to generate the Helmholtz vibrational free energy.

### D.1.2 Generating the E(V) curve

Once the reference structure is obtained, an electronic potential energy surface must be mapped out. There are two ways to accomplish this 1) Isotropically expand and contract the reference unit cell volume and perform a fixed-cell optimization on each structure (Isotropic E(V) curve) or 2) Apply external pressure to the reference structure and allow the structure to optimize (Anisotropic E(V) curve). We recommend using the latter method as this will allow the lattice parameters to relax naturally. If you are using Quantum Espresso for the quantum calculation then for both methods it is recommended that the optimization be carried out using the Quantum Espresso optimizer instead of HMBI's.

When performing these calculations it is ideal to have a minimum of 10-15 volumes

for E(V) curves. The basin of the electronic energy surface needs to be well sampled as there can be significant curvature in this region. Additionally, in order to obtain a good fit of the Gibbs free energy surface it is best to sample volumes that are at least 20-30 Å$^3$ away from your reference volume in both directions. If temperatures are expected to exceed 300 K, then sample at least 40-50 Å$^3$ away from your reference volume in the large volume-regime. Do the same in the small-volume regime for pressures greater than 2 GPa. For the calculation of the $F_{vib}(T)$ you will need two structures about $\pm$10 Å$^3$ away from your reference volume. It is recommended that you continue generating points until you achieve structures within 2-3 Å$^3$ of this target (See Figure D.1). Future work implementing constant-volume optimization routines such as the ones that exist in VASP and CRYSTAL17 would greatly speed up the generation of this potential energy surface.[300,311]

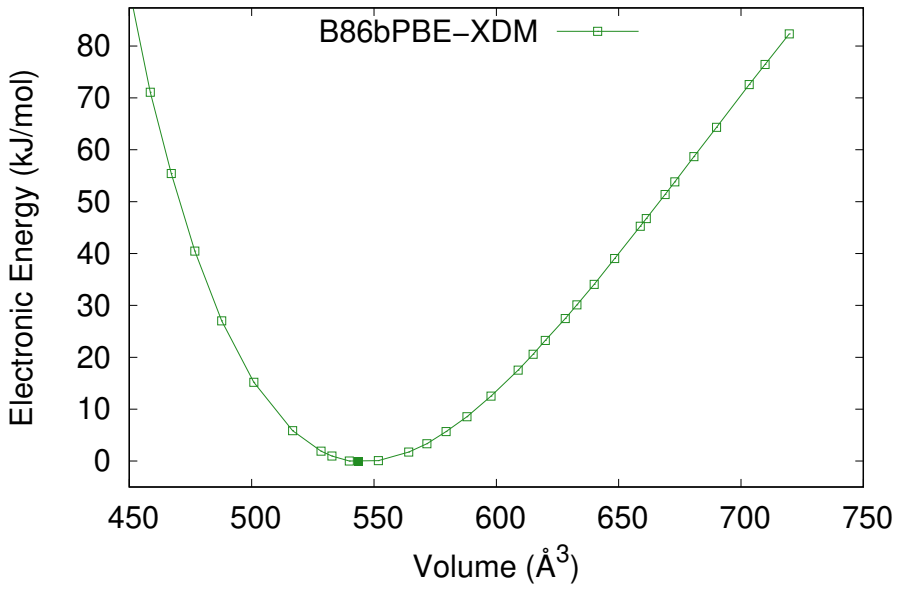



Figure D.1: An example of an electronic energy curve generated with DFT. The reference structure is shown with the solid square.

The anisotropic E(V) curve approach can often require a few iterations to achieve sufficient coverage of the electronic potential energy surface. While it is relatively straightforward to do this with isotropic expansion and contraction of the lattice volume, crystals containing larger molecules or more layered systems typically expand in a more anisotropic manner. Oftentimes the final energy reported will include the external pressure that was applied to the reference volume. To correctly obtain the electronic potential energy surface this PV term will need to be subtracted from the reported enthalpy. These optimizations are fairly repetitive so scripting is recommended.

### D.1.3 Optional: Perform Single-point Energy Corrections

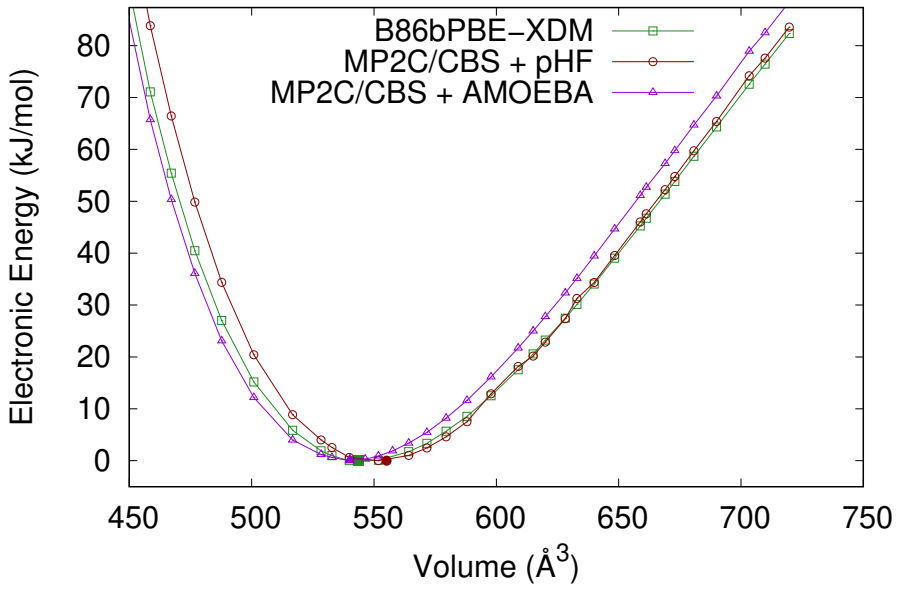



Figure D.2: An example of a single-point energy corrected electronic energy curve. The minima on each potential energy surface are displayed with solid points.

After obtaining an E(V) plot at the cheaper level of theory it is possible to single-

point energy correct up to the desired level of theory. This involves taking the geometries at each volume along the E(V) curve, setting up a new HMBI job at the desired theory, and running the job. While this is simple in concept, in practice this can often take a long time to accomplish. Oftentimes it is possible to set up the monomer and dimer jobs and submit these separately rather than attempting to run the entire job on one node. Always ensure that the jobs are only calculating energies and not gradients as these are not needed and can take significantly longer to calculate.

Once the single point energy calculations are complete it is necessary to visualize the change (see Figure D.2). Oftentimes, the energy well will shift since the potential energy surface has changed. In order to ensure that the bottom of the new energy well is sufficiently sampled it is recommended that the SPE corrections are performed immediately after each geometry finishes optimizing. If the original set of geometries does not sufficiently describe the new energy well then it may be necessary to continue generating structures at the cheaper level of theory until the new E(V) curve has sufficient coverage.

## D.2 Calculating the vibrational frequencies ($\omega_i$)

Once the potential energy surface is generated the Grüneisen parameters must now be calculated. To do this, two structures that were generated on the electronic potential energy surface are chosen to calculate additional harmonic phonons. Note that this is only done at the cheaper level of theory and does not need to be re-done even if the energy well has shifted (it is expected that the Grüneisen parameters will correctly adapt the phonons to account for the volume shift). At the moment structures are typically chosen to be

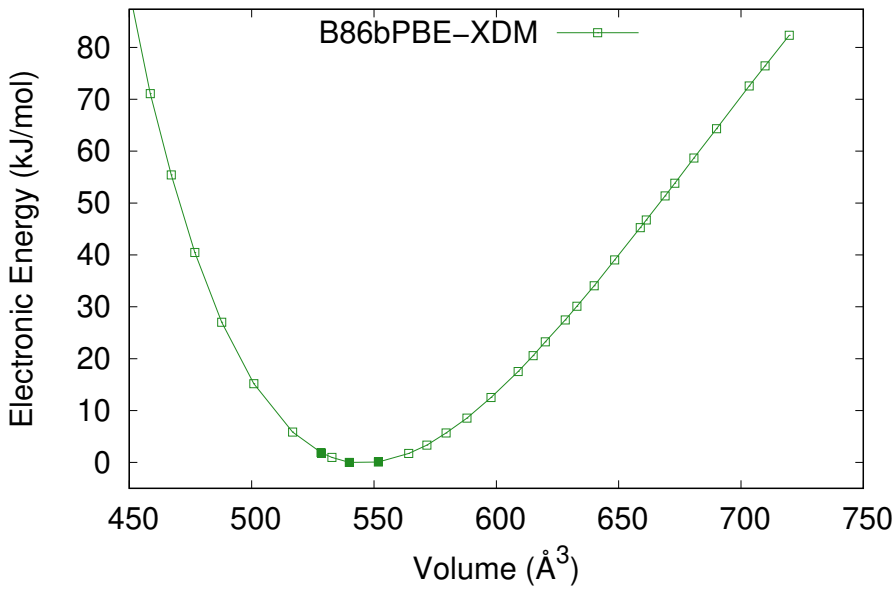



Figure D.3: An example of the structures chosen to calculate harmonic frequencies. The chosen structures are displayed with solid points.

$\sim$10 Å$^3$ away from the reference volume (see Figure D.3). While the choice of this volume is based on a previous study, future work investigating a % volume change rather than an absolute volume threshold would be useful.[105] Once the phonons are calculated, HMBI is used to mode-match these new phonons to the reference phonons. This generates a .freq file that is used by the MATLAB script.

## D.3 Calculations in MATLAB

We have created a MATLAB script to automatically calculate the volume which minimizes the Gibbs free energy (see https://github.com/jmcki003/ for the MATLAB script). Once you have obtained the electronic energy surface (E(V)) at the desired energy level you will need to place the data in a file format that MATLAB can readily read in. A .dat or

a .txt file works well and will typically contain a commented out header line with useful reminders of what data the file contains, and the volumes (in $\text{Å}^3$) and electronic energies (kJ/mol). Note that MATLAB is sensitive to tabs, so take care that only one tab or space exists between the volume and electronic energy data points. It might also be useful to name the file in a way that you and (hopefully) others can understand 3 months from now.

The MATLAB script also requires the .freq file that was created in Section D.2. Finally, the desired temperature and pressure range that the Gibbs free energy surface should be calculated at must be set in the script.

### D.3.1 Fitting the E(V) curve

The $E(V)$ curve will be read by the MATLAB script and fitted to a Murnaghan equation of state (EOS),

$$E(V) = E_0 + \frac{B_0 V}{B_0'}\left[\frac{(V_0/V)^{B_0'}}{B_0' - 1} + 1\right] - \frac{B_0 V_0}{B_0' - 1} \tag{D.6}$$

where $E_0$, $V_0$, $B_0$, and $B_0'$ are the fit parameters. $E_0$ gives the electronic energy at the minimum, $V_0$ is the molar volume at the minimum energy, $B_0$ is the bulk modulus, and $B_0'$ is the first derivative of the bulk modulus with respect to pressure. This is done to report the correct minimum on the electronic energy surface, which is primarily used to calculate the thermal expansion that is due to the zero-point vibrational energy contribution (ZPVE).

The procedure described in Chapter 4 calls for the fitted electronic energy curve to be used to calculate the Gibbs free energy surface at any given temperature and pressure. Previously, this was used to get away with a sparser sampling of the electronic energy

surface as the MATLAB script can be used to generate a finer grid of volumes to both interpolate and extrapolate over. While this strategy works well for volumes in the basin of the electronic energy surface, it was found that the murnaghan EOS fit can introduce 2-4 kJ/mol error in the fitted electronic energy surface at extreme volumes (large/small volumes). Especially at high pressures (greater than 2 GPa) or high temperatures (greater than 500 K) this error becomes non-negligible. Subsequent tests with additional EOS fits (Birch-Murnaghan, Poirier-Tarantola logarithmic, and Vinet EOS fits) showed that this error was reproducible. Instead, we now recommend explicitly calculating more raw data points on the electronic energy surface. For all future calculations, the raw data points are used instead of the fit to ensure no unnecessary error is introduced in the Gibbs free energy surface.

### D.3.2 Calculating the Helmholtz Free Energy Curve ($F_{vib}(T)$)

Next, the MATLAB script reads in the frequencies that are provided in the .freq file that HMBI generated. The Grüneisen parameters are automatically calculated and these are used to adapt the reference phonons to each volume that exists on the electronic energy curve. The Helmholtz vibrational free energy is then calculated using Equation D.2 at every temperature that was set in the MATLAB script. This generates the Helmholtz vibrational free energy curve as a function of volume and temperature (see D.4).

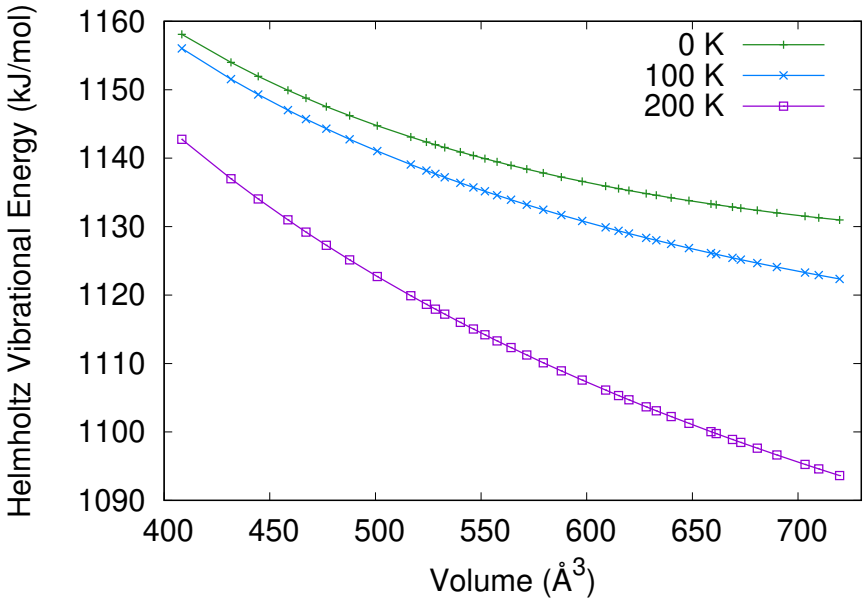



Figure D.4: An example of the Helmholtz vibrational free energy contribution generated at each temperature.

### D.3.3 Calculating the Gibbs Free Energy Curve ($G(T, P)$)

Once the electronic energy surface (E(V)) at the desired level of theory is obtained, and the Helmholtz vibrational free energy surface is generated, these curves are combined with the PV contribution at the desired pressures to obtain the Gibbs free energy surface (see Figure D.5). Since the particular volume which minimizes the free energy for a given temperature is unlikely to correspond to one of the sampled volumes, each free energy curve is fitted to a weighted double-Murnaghan equation of state. The Murnaghan equation of state is given by,

$$G(V) = G_0 + \frac{B_0 V}{B_0'}\left[\frac{(V_0/V)^{B_0'}}{B_0' - 1} + 1\right] - \frac{B_0 V_0}{B_0' - 1} \tag{D.7}$$

where $G_0$, $V_0$, $B_0$, and $B_0'$ are the fit parameters. $G_0$ gives the Gibbs free energy at the minimum, $V_0$ is the molar volume at the minimum energy, $B_0$ is the bulk modulus, and $B_0'$

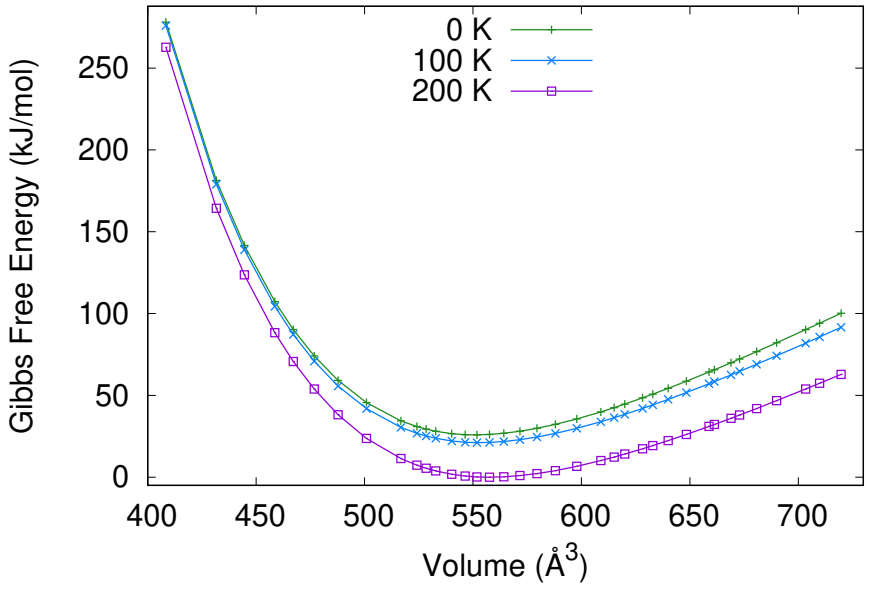



Figure D.5: An example of the Gibbs free energy surface generated at 0, 100, and 200 K. The pressure is set to 0 GPa.

is the first derivative of the bulk modulus with respect to pressure. This method identifies the free-energy minimum effectively while avoiding artifacts that can be caused by simpler equation of state fits or splines. The partitioning of the Gibbs free energy surface and fitting to the weighted double-Murnaghan equation of state is handled automatically by the MATLAB script (see Figure D.6).

## D.4 Processing the results

A number of useful files are generated by the MATLAB script. Briefly, the most useful files that are generated are the Helmholtz vibrational energy vs. volume, the Gibbs free energy vs. volume, and a summary file which contains a breakdown of the Gibbs free energy, Entropy, Enthalpy, and Electronic energy contributions at the minimum volume for

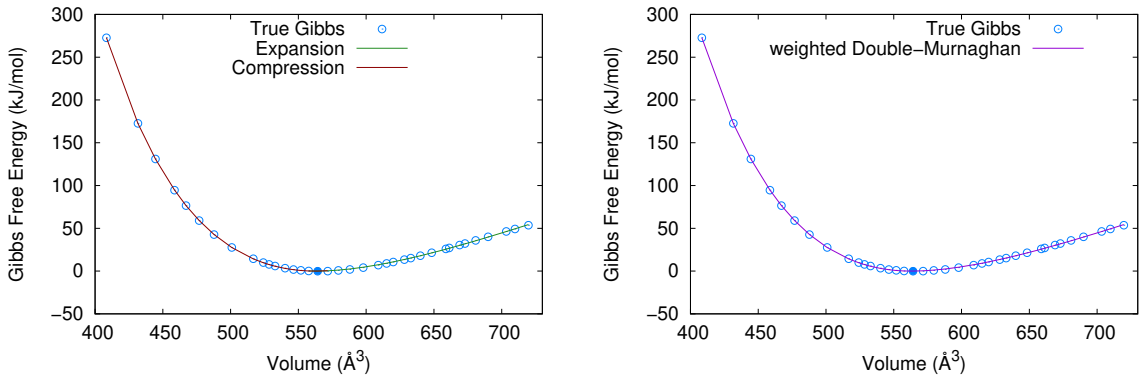



Figure D.6: An example of a weighted double-Murnaghan equation of state fit applied to the Gibbs free energy surfacee. On the left is the partitioned free energy surface which are separately fitted to a Murnaghan EOS fit. On the right is the weighted double-Murnaghan fit.

each temperature and pressure. The energy breakdown can be extremely useful especially when determining the relative stabilities of different polymorphs. While heat capacities are currently not reported, future work implementing this into the MATLAB script could prove useful.

### D.4.1 Obtaining the geometry at a given temperature and pressure

Once the volumes which minimize the Gibbs free energy surface at each temperature and pressure are known, it is now possible to obtain the structure. To do this, the lattice parameters for the volumes that were used to generate the electronic energy surface are fitted to a high-order polynomial (see Figure D.7). Unfortunately, not every fit will be smooth as the lattice angles especially tend to be difficult to fit. It is recommended to use at least a third-order fit in order to correctly obtain a measure of curvature in the fits. Using the volume that minimized the Gibbs free energy surface, the lattice parameters can now

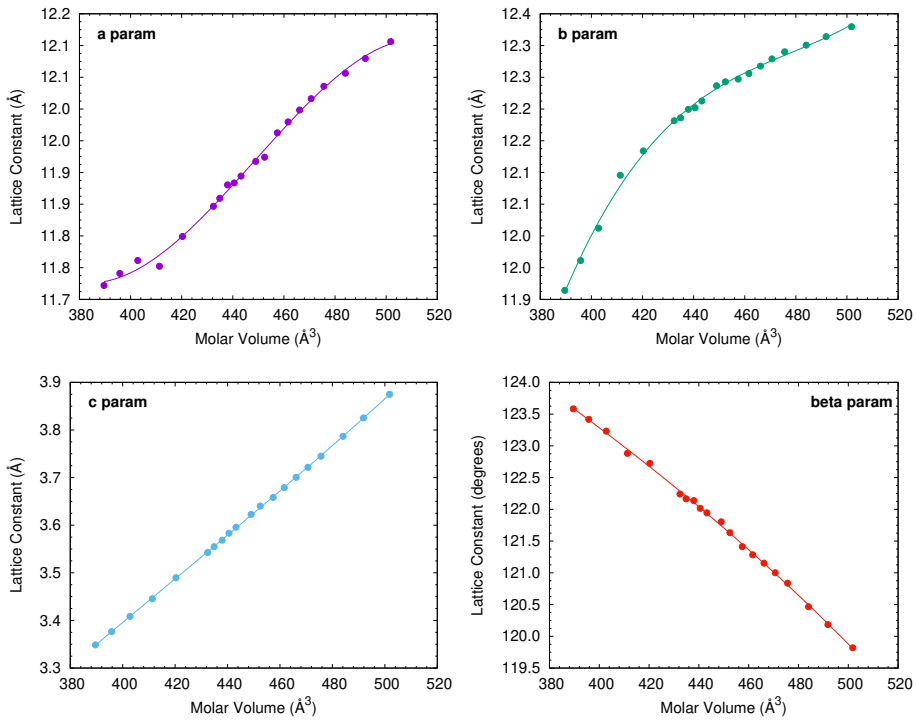



Figure D.7: An example of the fitted lattice parameters fitted to a third-order polynomial.

be obtained. The fractional coordinates of the reference structure are then used to generate the atomic positions at each volume. Finally, the volumes are frozen and the atomic positions are optimized using the lower level of theory. There is a question as to whether, for the structures whose potential energy surface were single-point energy corrected up to the desired level of theory, if the atomic positions should be optimized at this level of theory. Future work investigating this problem would be welcome. Implementing constant volume optimization routines would also be useful as the fitting of the lattice parameters could be avoided. Once the structure is obtained this can then be used for the evaluation of secondary properties such as NMR and Raman spectroscopy.