Molecular Crystal Global Phase Diagrams III: Sufficient parameter space determination

J. Brandon Keith[a,b,c*] and Richard B. McClurg[a,d]

[a]*Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, Minnesota 55455, USA,* [b]*Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84606, USA,* [c]*California Institute of Technology, Division of Engineering and Applied Science, Mail Code 138-78, Pasadena, California 91125 USA, and* [d]*SSCI, a division of Aptuit, West Lafayette, Indiana 47906 USA. E-mail: jbrkeith@gmail.com*

(Received 0 XXXXXXX 0000; accepted 0 XXXXXXX 0000)

Abstract

In previous parts of this series (Mettes *et al.*, 2004; McClurg & Keith, 2009) we developed a method for constructing global phase diagrams (GPDs) for molecular crystals and applied the method to single-component ordered crystal structures of tetrahedral molecules. GPDs are useful for visualizing what types of crytal structures a given molecule may assume depending on molecular form / interaction. Their construction uses group theoretical methods which enumerate all possible symmetry-breakings during a statistical mechanical high-to-low temperature search. In this work we expand upon these results by outlining a method to determine a sufficiently rich parameter space to represent the experimentally observed crystal structures in a data set derived from the Cambridge Structural Database. This is significant because previous work (Mettes *et al.*, 2004) did not specify the number of parameters needed for GPDs. Although there are suggestions in the literature that thousands of parameters are required to adequately describe tetrahedral molecule intermolecular potentials, we find that 15 parameters are sufficient to represent the structures of our test data. We discuss the origin of this difference and its implications for determining GPD parameter values from a more detailed intermolecular potential and for interpreting GPD parameter values.

1. Introduction

Crystal engineering is the design and synthesis of solid-state structures with desired properties. For molecular crystals this necessitates a thorough understanding of intermolecular interactions. While the properties of isolated molecules are primarily attributable to strong covalent bonding between atoms, solid-state properties result from relatively weak interactions between molecules or low-dimensional aggregates of molecules called synthons. Two primary interactions holding together supramolecular synthons are hydrogen bonding and coordination complexation, though π-π, halogen-halogen, and ionic interactions have also been exploited (Thalladi *et al.*, 1996). These synthons then can be assembled into one-dimensional rods, two-dimensional sheets, and three-dimensional crystal structures. Since many of the bulk properties of molecular materials are dictated by the manner in which the molecules are ordered in the solid state, crystal engineers seek to control this ordering and thus a material's electrical, optical, thermal, and solubility properties (Desiraju, 1989; Braga *et al.*, 1999; Simon, 2000; Lommerse *et al.*, 2000; Holman *et al.*, 2001; Moulton & Zaworotko, 2001).

Thus supramolecular chemistry depends on subtle interactions and how to control them. Noncovalent bonds have low energies and often little or no activation energy for formation. This low bond energy results in structures stabilized by difficult-to-control entropic effects, low melting points, and frequent polymorphism. Likewise, as temperature changes, the balance of these effects changes, resulting in structural changes (Neumann *et al.*, 2003). Thus energetics and thermodynamics are both essential in designing, controlling, and studying molecular crystal chemistry.

Despite ongoing progress in understanding how molecular crystal structures form, there is still a need for tools to rationalize the crystal structures adopted by a molecule or collection of molecules as a function of temperature and intermolecular potential.

Tools for classification and rationalization of molecular packing tendencies under complex conditions aid crystal engineering to expand its material base and exploit new conditions to form novel structures. Phase diagrams are one such class of tools.

Global phase diagrams have been used to rationalize binary fluid mixture thermodynamics for many years. Van Konyenburg and Scott provided a general and systematic categorization of the different types of fluid-phase behavior in binary mixtures based upon the topology of the critical loci of a van der Waals equation of state with simple mixing rules (van Konyenburg & Scott, 1980). They devised a set of global phase diagrams (GPDs) to classify and rationalize liquid-vapor phase behavior in terms of general, non-specific molecular parameters. They showed that a simple model can predict qualitatively most of the known patterns of fluid-phase behavior, and can reveal the mechanisms of transition among the different types. Thus they describe five of the six main experimental classes of fluid behavior differentiated by the temperature-pressure projections of their critical loci. Although additional fluid-phase mixture behaviors have since been observed and a revised nomenclature is now available (Bolz *et al.*, 1998), the original van Konyenburg system is still pervasive in the literature (Aparicio-Martinez & Hall, 2007*a*; Aparicio-Martinez & Hall, 2007*b*; Cismondi & Michelsen, 2007). Many authors have repeated this type of study in which other equations of states based on simplified models, cubic equations of state, or intermolecular potential-based equations provide the parameters of GPDs (Polishuk *et al.*, 2000; Polishuk *et al.*, 2002; van Pelt *et al.*, 1995).

In previous work we have proposed and demonstrated the construction of molecular crystal GPDs which differ from fluid-phase GPDs in fundamental ways. Gasses and liquids are both isotropic fluids which lack long-range orientational and translational order. Therefore density as a function of temperature and pressure is a sufficient order parameter for fluid phase equilibrium. Describing equilibrium among crystalline

solids requires more complicated order parameters to account for the diversity of their long-range orientational and translational orderings. Also, most single component fluids share a set of common features in their phase behaviors including a vapor/liquid equilibrium locus terminating at a critical point with universal scaling properties, a supercritical fluid region at temperatures and pressures in excess of the critical point, and an asymptotic approach toward ideal gas properties at low pressures and/or high temperatures. The exceptions to these generalizations are ionic fluids that decompose before generating a significant vapor phase. Crystalline solids lack these thermodynamic generalizations which prevents the use of simple equation-of-state methods for generating GPDs for molecular solids.

The molecular crystal GPDs developed in this series are designed to elucidate the diverse and complex phase behavior of a set of arbitrarily-shaped molecular scale objects arranged in a long-range ordered packing rather than in disordered fluid phases. They do not use an equation of state and mixing rules but instead rely on fundamental postulates of equilibrium statistical mechanics, a set of basis functions over rotational space $SO(3)$ to represent the orientational intermolecular potential, and a set of translational packings over translational space T^3 to represent their translational ordering. The orientational interactions are the result of molecular orbitals fixed in a molecular frame of reference interacting with molecular orbitals of other molecules via the intervening orientational space S^2 so that the overall potential is a complete set over $SO(3) \times S^2 \times SO(3)$ with intermolecular potential coefficients indexed by three angular momenta. The mathematical details are explained in (Mettes *et al.*, 2004). Representing the orientational interactions in this way leads to a countably-infinite set of coefficients. A subset of these coefficients is used as the set of independent variables in GPDs. The space can then be projected into two- or three-dimensional spaces for visualization purposes. Crystalline phases occupy volumes in the GPDs and

equilibrium phase transitions occur at the boundaries between the phase volumes.

The proper thermodynamic condition for equilibrium of an isothermal system is minimization of the free energy. The free energy is calculated relative to a high-temperature reference lattice in which molecules remain translationally ordered but are rotationally disordered. Experimentally this is termed a plastically crystalline state (Sherwood, 1979) and is observable for many molecules, such as adamantane or methane. Other systems melt, sublime, or decompose before the plastically-crystalline state is obtained. Whether or not the reference state is observed, it provides a well-characterized thermodynamic reference state. This is similar to fluid phase systems which are routinely referenced to an ideal gas state, even if it has not been observed. As the temperature is lowered for a solid with fixed intermolecular potential, new phases arise through spontaneous symmetry breaking, leading to at least partial rotational ordering of the molecules.

An example of a GPD is shown in Figs. 1 and 2. In these diagrams an fcc lattice is used as the high-temperature reference state where the molecules are rotationally disordered and look like spheres. As the temperature is lowered, the molecules adopt a fixed orientation and the crystal undergoes a change a structural change as shown for the 10 new phases in these figures.

A series of additional phase transitions may arise as the temperature is lowered toward absolute zero. For molecules with non-trivial point-group symmetry, it is convenient to reduce the dimensionality of the phase diagrams through adaptation of the potential to account for the molecular symmetry. This approach of expanding the angle-dependent intermolecular interactions into symmetry-adapted rotation functions was originally developed for phase transitions in heavy methane by James and Keenan (James & Keenan, 1959) and has since been applied to crystals such as solid C_{60} (Michel *et al.*, 1992; Lamoen & Michel, 1999). Figures 1 and 2 were constructed

for molecules with tetrahedral point group symmetry.

Like van Konyenburg fluid-phase GPDs, molecular crystal GPDs are useful as a classification tool. They can be used in data mining intermolecular forces from structural repositories such as the Cambridge Structural Database (CSD) (Allen & Motherwell, 2002) or Crystallography Open Database (Grazulis *et al.*, 2009) portals such as the Virtual Neutron Facility (Lin & Keith, 2009). Also, by locating new structures that may have more desirable properties than an existing crystal structure, GPDs provide feedback for molecular synthesis for crystal engineering. Further, GPDs can provide rationalization and/or prediction of crystal polymorphs for a given molecule. Metastable polymorphs, for example, are likely to be adjacent phases on GPDs because their free energies are slightly higher than the global minimum.

In a companion paper (McClurg & Keith, 2009) we classify experimental crystal structures composed of tetrahedral point group molecules to determine what fraction of structures are amenable to inclusion in the GPDs and the number of reference lattices necessary to span the observed structures. We find that 60% of crystal structures composed of molecules with Td point group symmetry are amenable and that eight reference lattices are sufficient to span the observed structures. In the remaining 40% of the structures, the molecules form synthons that pack to generate the three-dimensional crystal structure. This kind of hierarchical structure is not incorporated in the GPDs as described in (Mettes *et al.*, 2004). For the majority of structures in the test set, eight reference lattices are sufficient to describe the lattice of molecular centers of mass. We consider a set of GPDs, one for each reference lattice.

This work discusses the issue of a minimal set of intermolecular potential coefficients needed to construct GPDs. For GPDs to be useful, the observed crystal structures for an ensemble of molecules with a given point group symmetry should occupy volumes of the GPD with a truncated set of intermolecular potential coefficients. If the full

infinite parameter set is needed, then the GPDs lose much of their utility. It has been suggested that too many parameters are needed to describe intermolecular potentials for practical use in GPDs (Briels, 1980). We find that a smaller number of parameters are sufficient. This finding greatly enhances the utility of GPDs for materials design. The smaller set of parameters allows construction of manageably small-dimensional diagrams. Still, 3D-projections are required to produce visualizations with which users may interact and explore crystal structure relationships and proximity in crystal phase space. Molecular crystal GPDs with a manageable number of parameters would be valuable to crystal engineers seeking to coerce an optically active molecule through synthetic perturbations to crystallize in a noncentrosymmetric space group, for example. It could also prove useful in polymorph screening to enumerate polymorphs of active pharmaceutical ingredients.

To explore the requisite number of parameters necessary to construct molecular crystal GPDs, we begin in Sec. 2 by recalling the rotational potential used previously in (Mettes *et al.*, 2004). We then outline the computational procedure to find potential parameters for each structure in the same experimental data set used in (McClurg & Keith, 2009). This is done using global optimization techniques among intermolecular potential parameter *and* translational symmetry space. In Sec. 3 we discuss the structures that are successfully located in potential parameter space using our candidate structure library and intermolecular potential. We also identify exceptions that cannot be located using the method utilized and make recommendations for further work to rectify these omissions.

2. Representative Potential Determination

Global phase diagrams require (1) reference lattices consistent with molecular centers of mass in experimental structures and (2) intermolecular potential (IP) parameters to

use as independent parameters. In (McClurg & Keith, 2009) we found that the majority of the experimental structures composed of tetrahedral molecules in the CSD can be classified using only four reference lattices. Since there are too few examples for other reference lattices, we focus on the set {bcc, fcc, hcp, sc} and use these as a test set to see how many parameters are required in molecular crystal GPDs. Each of these reference lattices has at least four crystal structures with monomer synthons in the data set. This means a given structure can be categorized as a subgroup of one of these reference lattices when single molecules are reduced to spheres (as opposed to dimers or chains–see (McClurg & Keith, 2009)). In this section we seek to determine a *sufficient* parameter space such that experimental structures are each stable in a finite volume in the IP parameter space. Thus we review our choice of IP and define the potential parameters. Then we outline a method for identifying representative potential parameters for each experimental structure. This is done using a low temperature structural limit which is convenient and consistent with the experimental database but is not strictly necessary. Then the IP is truncated to create a finite dimensional IP parameter space. A library of alternative crystal structures is constructed with which to compare experimental structures, and a figure of merit is specified when searching for the potential parameters. This procedure identifies a structure with similar cell shape, molecular center of mass, and molecular orientations as illustrated in Fig. 3.3.

2.1. Mesoscopic Hamiltonian

Previously (Mettes *et al.*, 2004) we discussed the construction of a nearest neighbor potential for molecules interacting primarily through van der Waals interactions. The potential V is a level of abstraction above an atom-atom or site-site potential but retains a firm basis in quantum mechanics (van der Avoird *et al.*, 1994). In (Mettes *et al.*, 2004) we discuss how the rotational contribution $V_{\rm or}$ of V consists of a two

center expansion constructed by coupling one center basis functions $U^{\ell_i}_{m_\tau n_\sigma}$ for pairs of molecules i and j through a coupling "matrix function" $J^{\ell_i\ell_j}_{m_\tau n_\sigma m_\rho n_\mu}$,

$$\begin{aligned} V_{\rm or} &= \frac{1}{2}\sum_{ij}\sum_{\ell_i\ell_j m_\tau m_\rho n_\sigma n_\mu} U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i) \\ &\quad \times J^{\ell_i\ell_j}_{m_\tau n_\sigma m_\rho n_\mu}(\boldsymbol{\nu},\boldsymbol{\Omega}_{ij}) U^{\ell_j}_{m_\rho n_\mu}(\boldsymbol{\omega}_i). \end{aligned} \tag{1}$$

The one half avoids overcounting pair-wise interactions, $\ell_i, \ell_j \in \mathbb{N}$, the natural numbers, $m_{\tau,\rho} \in \{\ell_{i,j}... - \ell_{i,j}\}$, and $n_{\sigma,\mu} \in \{\ell_{i,j}... - \ell_{i,j}\}$. A complete derivation of Eq. (1) is given in Appendix A, but here we explore its main components.

The functions $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ are functions of the orientation of the molecule through its Euler angles $\boldsymbol{\omega}_i$ using the passive convention (Varshalovich *et al.*, 1988). They are projected from SO(3) irreducible representations (IRs), also called Wigner functions, and contain both the point group symmetry of the molecule and that of the Wyckoff point in the crystal,

$$U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i) = \sum_{m_i n_i} S^{\ell_i *}_{m_i m_\tau} D^{\ell_i}_{m_i n_i}(\boldsymbol{\omega}_i) S^{\ell_i}_{n_i n_\sigma}. \tag{2}$$

The sparse unitary matrix $\boldsymbol{S}$ provides the linear combinations of Wigner functions which give a particular point group IR symmetry (Bradley & Cracknell, 1972). Specifically, the left multiplication by $\boldsymbol{S}^{-1}$ in Eq. (2) gives basis functions transforming like Wyckoff point group IRs and the right multiplication by $\boldsymbol{S}$ gives basis functions of the molecular point group IRs. Subscript τ is a compound index referring to multiple copies of the Wyckoff point group IR subduced in ℓ_i and m_τ spans the dimensions of each IR. Subscript σ is a compound index referring to multiple copies of the molecular point group unit IR subduced in the ℓ_i-th manifold of $SO(3)$ and n_σ is its dimension. Point group IR subduction frequencies in spherical harmonics are discussed elsewhere (Bradley & Cracknell, 1972) and in Sec. 2.2. The symmetry adaption leaves relatively few basis functions for molecules with non-trivial point group symmetry since only matrix elements where σ is the unit IR (*i.e.* A_1) are kept. This gives

basis functions in the molecular frame with the full molecular symmetry. Thus the set $\{U^{\ell_i}_{m_\tau n_\sigma}\}$ is a complete set of basis functions taking full advantage of molecular symmetry.

The coupling matrix $J^{\ell_i\ell_j}_{m_\tau n_\sigma m_\rho n_\mu}$ specifies the angular dependence with respect to molecular centers,

$$\begin{aligned} J^{\ell_i\ell_j}_{m_\tau n_\sigma m_\rho n_\mu}(\boldsymbol{\nu}, \boldsymbol{\Omega}_{ij}) &= \textstyle\sum_{\ell m_i m m_j} \nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}(r_{ij}) \begin{pmatrix} \ell_i & \ell & \ell_j \\ m_i & m & m_j \end{pmatrix} \\ &\quad \times S^{\ell_i}_{m_\tau m_i} C^{\ell}_{m}(\boldsymbol{\Omega}_{ij}) S^{\ell_j}_{m_\rho m_j}, \end{aligned} \tag{3}$$

where $\ell \in \{\ell_i+\ell_j \ldots |\ell_i-\ell_j|\}$, $m_{i,j} \in \{\ell_{i,j} \ldots -\ell_{i,j}\}$, and $m \in \{\ell \ldots -\ell\}$. The potential coefficients $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}(r_{ij})$ are a function of the distance r_{ij} between molecular centers. The neighbor distance r_{ij} and orientation Ω_{ij} of the intermolecular vector are determined by the reference lattice. Since we consider pairwise interactions only among the equidistant nearest neighbors of the reference lattices, the functions $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}(r_{ij})$ are only evaluated at the nearest neighbor distance and are treated as scalars. Thus the coupling matrix $J^{\ell_i\ell_j}_{m_\tau n_\sigma m_\rho n_\mu}$ contains reference lattice information through Ω_{ij} and r_{ij} as well as pairwise intermolecular potential information through the coefficients $\boldsymbol{\nu}$. These coefficients have inversion symmetries that reduce the number of independent values, even for asymmetric molecules. In particular $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}$ is zero if $\ell_i+\ell+\ell_j$ is odd and $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j} = (-1)^{\ell_i+\ell_j}\nu^{n_\mu,n_\sigma}_{\ell_j,\ell,\ell_i}$ for single component crystals (van der Avoird *et al.*, 1980).

As discussed previously (Mettes *et al.*, 2004), this pairwise-additive potential implicitly contains multimolecular many-body effects in the values of $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}(r_{ij})$ as a mean-field contribution. Also, because Eq. (1) is currently limited to nearest-neighbors, it is most appropriate for "van der Waals molecules", or molecular crystals held together primarily by dispersion forces rather than ionic or strongly dipolar molecular crystals which require Ewald sums over distant neighbors. Van der Waals molecules seem to be the majority of molecular crystals found in the CSD, so that this pairwise additive

potential over molecules with an implicit multibody contribution is a fair parametrization and useful as we seek to map CSD molecules to a specific number of $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}(r_{ij})$. The full set of $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}(r_{ij})$, denoted $\boldsymbol{\nu}$, are the axes of GPDs and constitute the search space for constructing molecular crystal GPDs.

2.2. Computational Method

Having established translational and rotational components of the intermolecular potential as the parameter space for GPDs, we seek to determine potential parameters $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}$ sufficient to produce each experimental crystal structure. Our method has five steps. (1) Define a figure of merit by which to order the crystal structures. (2) Ensure the potential, Eq. (1), is truncated such that it geometrically is capable of producing a symmetry change from the assigned reference lattice to the experimental phase. (3) Develop a library of structural types against which the energy of the phase of interest may be compared. (4) Determine the energy of an experimental phase as a function of $\boldsymbol{\nu}$. (5) Search potential parameter space until a $\boldsymbol{\nu}$ vector is found which makes the experimental phase energetically minimal relative to the alternatives in the library.

2.2.1. Figure of Merit In our previous work (Keith *et al.*, 2004; Mettes *et al.*, 2004) linear response theory was used to seek phase transitions as a model crystal is cooled from a disordered plastic crystalline reference state at high temperature. These transitions are bifurcation points of the free energy surface. Following each transition, the phase was identified by the presence of nonzero thermal averages of rotator functions $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ which serve as order parameters. This work examines the low temperature limit of the previous model where low is relative to the temperature T_{pt} at which the molecules undergo the first phase transition from the reference phase. In this regime the function $U^{\ell_i}_{m_\tau n_\sigma}$ itself is an adequate approximation to the thermal averages of the

rotator functions,

$$\lim_{T/T_{\rm pt}\rightarrow 0}\langle U^{\ell_i}_{m_\tau n_\sigma}\rangle = U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_0). \tag{4}$$

Also, the free energy $A(T)$ is equal to the potential V in that limit,

$$\lim_{T/T_{\rm pt}\rightarrow 0} A(T) = V. \tag{5}$$

Equations (4) and (5) are common approximations used in many crystal structure prediction codes (Verwer & Leusen, 1998) and are further justified by the exclusion of disordered structures (mesophases) in *CSDSymmetry* from which our data set is derived.

2.2.2. Potential Truncation The potential in Eq. (1) is a doubly infinite sum over manifolds ℓ_i and ℓ_j that must be truncated for practical applications. To determine an adequate truncation of the potential, we make use of space group irreducible representations (IRs). Symmetry-breaking mechanisms are classified by space group IRs and an order parameter direction such as (a, b) (Stokes *et al.*, 2007). The temperature-dependent values a and b are given in our model by components of space group IR-adapted basis functions $\boldsymbol{q}$ which are linear combinations of $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ (Mettes *et al.*, 2004). Space group IR distortions of the crystal can be decomposed into point group IR distortions of the distinguishable molecules in the crystal (Stokes *et al.*, 1991). To determine which IRs τ_w of point group w are in a symmetry-breaking space group IR τ_{SG}, we calculate their subduction frequencies n_{SG}

$$n_{SG} = \frac{1}{|w|}\sum_{g\in w}\chi^{\tau_{SG}}(g)^*\chi^{\tau_w}(g) \quad \forall\ \tau_{SG} \tag{6}$$

where $|w|$ is the number of elements g in w. $\chi(g)^{\tau_{SG}}$ and $\chi(g)^{\tau_w}$ are the traces of the matrix representation of the space group IR and point group IR, respectively. The calculations are easily performed using the ISOTROPY software package (Stokes

et al., 2007) using the "show frequency" command. Space group IR and point group IR characters may also be generated.

Using these point groups one may also calculate the number of times each point group IR τ_w appears for each manifold of SO(3), or ℓ_i value,

$$n_{SO(3)} = \frac{1}{|w|} \sum_{g \in w} \chi^{\ell_i}(g)^* \chi^{\tau_w}(g) \quad \forall\ \ell_i \tag{7}$$

where χ^{ℓ_i} is the trace of the matrix representation of an IR of SO(3) also called a Wigner rotation matrix. Such subduction frequencies may be easily calculated but are also tabulated in standard references (Bradley & Cracknell, 1972) and are shown for O_h and D_{3h} in Table 1. As there is no molecular unit IR in the first, second, or fifth manifold, there are no Wyckoff IR rows of $U^{\ell_i}_{m_\tau n_\sigma}$ so these manifolds are not shown.

If the potential is truncated before the manifold at which a Wyckoff IR forming a symmetry-breaking space group IR is present (*i.e.* if $n_{SG} = 0$ in Eq. (6)), the desired phase transition cannot occur with that potential. Therefore we calculate minimal SO(3) manifolds necessary to achieve a transition from the assigned reference lattice to an experimental structure. This is shown for the fcc reference lattice in Table 2, where we use the convention of appending the occupied Wyckoff positions of a crystal structure to its space group (thus a structure with space group 64 in which molecular centers of mass occupy Wyckoff points d and f would be denoted 64d,f). Point group IRs subduced in the space group IR and the first manifold or ℓ_i value of SO(3) at which a given Wyckoff IR appears are also shown. The pathways shown are calculated in ISOTROPY by finding a group-subgroup transition from the high-temperature reference phase, such as fcc, to the observed structure.

For some pathways, such as that of carbon tetraiodide, 225a $\rightarrow$ 121a in Table 2, there is a single space group IR Γ_5^- subducing a single point group IR T_{2u}. Table 1 shows that this point group IR first occurs in the third manifold of SO(3) when split by an O_h point group crystal field. Thus $\ell_i^{\max} \geq 3$ may give the 121a structure but $\ell_i^{\max} < 3$ is

insufficient. For other pathways, such as that of tetrakis(M_3-t-Butylimido)-tetraiodo-tetra-indium, 225a $\rightarrow$ 12i, there is more than one Wyckoff IR subduced in a space group IR. E_u, T_{1u}, and T_{2u} can lead to this transition. As two of them are present at the third manifold (Table 1), a potential with $\ell_i^{\max} \geq 3$ may also produce this transition. For still others, such as that of adamantane, 225a $\rightarrow$ 114a, the pathway is coupled, or composed of the direct sum of two space group IRs. Three different coupled space group IRs decompose into three different point group IRs. One of these pathways uses IRs on the third manifold of SO(3) ($\ell_i^{\max} \geq 3$) while the other two require fourth manifold basis functions. A similar analysis for the bcc, hcp, and sc reference lattices given in the appendix shows symmetry-breaking Wyckoff IRs appear by at least the third or fourth manifold in nearly all cases.

In real phase transitions, IRs on the first allowed non-trivial manifold typically induce symmetry breaking. Second manifold induced symmetry breaking is uncommon and higher manifold induced phase transitions are not generally observed (Lynden-Bell & Michel, 1994). The reasons for this are different for small and large molecules. For small molecules with only a few atoms (*i.e.* CH_4) molecular orbitals tend to form low-energy, slowly varying topologies to lessen kinetic energy contributions in Schrödinger's equation. The result is an IP well represented by smooth, slowly varying basis functions. For large molecules, low manifold contributions to the potential are sufficient to locate attractive/repulsive regions of the potential and the relative magnitudes of attractive configurations, even though these basis functions are not sufficient to represent the fine structure in the potential due to individual atoms (Hloucha *et al.*, 2001). In either case, the most important aspects of the IP are given by slowly varying functions while more rapidly varying functions produce finer details. Thus as a minimal basis set we truncate Eq. (1) at $\ell_{\max} = 3$ or 4.

Truncating at $\ell_{\max} = 4$ leaves fifteen coefficients in the potential, $\nu_{0,3,3}$, $\nu_{0,4,4}$, $\nu_{3,0,3}$,

$\nu_{3,2,3}$, $\nu_{3,4,3}$, $\nu_{3,6,3}$, $\nu_{3,1,4}$, $\nu_{3,3,4}$, $\nu_{3,5,4}$, $\nu_{3,7,4}$, $\nu_{4,0,4}$, $\nu_{4,2,4}$, $\nu_{4,4,4}$, $\nu_{4,6,4}$, and $\nu_{4,8,4}$. The coefficient $\nu_{0,0,0}$ is negligible since it only affects the trivial basis function $U^0_{0,0} = 1$ which is isotropic and therefore unimportant in rotational ordering. As the unit IR for T_d appears *once* in the zeroth, third and fourth manifolds, n_σ and n_μ are always 1_{A_1} and have been dropped in the notation for the coefficients in Eq. (3). For space groups whose occupied Wyckoff point group in the reference lattice has the inversion as one of its symmetry operators, $\nu_{0,x,x}$ cancels in the crystal field for odd values of x. This is the case with the ν_{033} coefficient in the bcc, fcc, and sc reference lattices with 14 coefficients but not for hcp with 15.

2.2.3. Candidate Lattice Library We seek a set of potential parameters $\boldsymbol{\nu}$ such that the energy of an observed structure is lower than other alternatives, whether the alternative structures are experimentally observed or not. In our previous work (Mettes *et al.*, 2004), high symmetry point IR distortions (Stokes & Hatch, 1988; Stokes *et al.*, 2007) from a reference lattice led to structures classified using their isotropy subgroups (ISs). These subgroups represent the most common types of phase transitions and are minimal in the sense that only one domain of each structure is tested. This compares favorably to many crystal structure prediction algorithms which generate thousands of multi-domain duplicate structures (Gavezzotti, 2007). Therefore they are convenient for constructing a library of candidate lattices. They are discussed further in Sec. 3. As most molecular crystals have one molecule or less per asymmetric part of the unit cell ($Z' \leq 1$), we have discarded ISs that imply more than one occupied Wyckoff point (Padmaja *et al.*, 1990). ISs whose *primitive* unit cell contains more than eight molecules are also discarded (Gdanitz, 1997). We note these constraints are made based on what is commonly observed in the Cambridge Structural Database and can be incrementally extended in order to achieve a larger candidate library.

To minimize the energy of these candidate lattices, basis functions $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ are placed at one molecular center of the molecular positions of the lattice of each IS and space group operations of the IS are applied to generate basis functions $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ for all other molecules in the Wyckoff orbit. Although these candidate lattices are generated from ISs, after Euler angle minimization they are free to assume whatever structure the minimization algorithm can find provided the reference orientation of the basis functions $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ are related by IS space group operations. Thus a candidate lattice has fixed (super) cell parameters and Wyckoff points but variable Euler angles and space group. This procedure gives a candidate lattice library that, at a minimum, includes all high symmetry k point isotropy subgroups but which can contain other lower symmetry structures as well (due to additional symmetry breaking while rotating the basis functions $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$) and provided translational symmetry requirements are met. As indicated above, this candidate library may be systematically improved by explicitly including new isotropy subgroup supercells.

Applying this procedure yielded 75 candidate lattices for bcc, 67 for fcc, 19 for hcp, and 105 for sc. Minimizing the energy of all candidate lattices gives a lowest energy structure with which to compare the energy of the experimental structure while searching parameter space. Although these candidate lists are remarkably small compared to the thousands of structures generated in many crystal structure prediction programs, they are sufficient to place most of the observed structures as discussed in Sec. 3.

2.2.4. Experimental Structure Energies With a set of candidate lattices with which to compare a given observed structure in $\boldsymbol{\nu}$-space, we now need to find the energy of each observed structure, leaving $\boldsymbol{\nu}$ unspecified. This can be done by translating the atomic positions of the observed structure into the needed Euler angles ω and intermolecular angles Ω of Eq. (1). The intermolecular angles Ω are easy to obtain because they

are determined by the reference lattice, but the Euler angles are dependent upon the orientation of the molecules in the reference lattice. This is complicated by the fact that the experimental atomic positions in each molecule are slightly distorted from perfect molecular Td point group symmetry and are slightly shifted from the ideal Wyckoff points of their reference lattice. This is due to small strains resulting from embedding the experimental lattice in the almost perfect but nonetheless approximate symmetry of its reference lattice (see (McClurg & Keith, 2009)). To obviate these difficulties, the reference lattice molecular centers of mass are first slightly shifted to coincide with the centers of mass of the molecules in their native space group. Then the Euler angles are calculated by rotating a rigid tetrahedral molecule to minimize its atomic distances compared to the coordinates of the experimental molecule. Thus the Euler angles of a perfect tetrahedron are obtained by a "best fit" to molecular coordinates of a slightly distorted tetrahedron. We note the reference orientation of the molecule is with three-fold axes pointing in the [111], [$\bar{1}\bar{1}$1], [$\bar{1}$1$\bar{1}$], and [1$\bar{1}\bar{1}$] directions.

Next the rotator functions $U^{\ell_i}_{m_\tau n_\sigma}$ may be calculated directly using Eq. (2). $U^{\ell_i}_{m_\tau n_\sigma}$ is fully symmetric with respect to molecular point group symmetry and is not influenced by non-unique Euler angles. Substituting the $U^{\ell_i}_{m_\tau n_\sigma}$ into Eq. (1) yields the interaction potential as a function of the potential parameters $\boldsymbol{\nu}$. This gives the energy $E = V(\boldsymbol{\nu})$ as a linear polynomial in $\boldsymbol{\nu}$. For instance,

$$\begin{aligned} E \;=\; & 0.2901\,\nu_{0,4,4} - 0.0028\,\nu_{3,0,3} + 0.0036\,\nu_{3,1,4} \\ & - 0.1020\,\nu_{3,2,3} - 0.0425\,\nu_{3,3,4} - 0.0127\,\nu_{3,4,3} \\ & + 0.0053\,\nu_{3,5,4} - 0.0006\,\nu_{3,6,3} + 0.0071\,\nu_{3,7,4} \\ & - 0.0055\,\nu_{4,6,4} - 0.0183\,\nu_{4,8,4} + 1.2640\,\nu_{4,0,4} \\ & + 0.0109\,\nu_{4,2,4} + 0.0316\,\nu_{4,4,4}, \end{aligned} \tag{8}$$

for tetrakis(trimethystannyl)silane [CSD structure MEZDIE01], which crystallizes in

space group 2.

2.2.5. Minimum Energy Gap Now that we have a potential that is capable of producing all the experimental structures from (McClurg & Keith, 2009) for the {fcc, sc, hcp, bcc} reference lattices, and we have input each structure's molecular Euler angles to reduce it to a form such as Eq. (8), we need to develop an algorithm that will isolate this structure in $\boldsymbol{\nu}$-space as the most thermodynamically stable structure. To this end we propose a general minimum energy difference algorithm that makes use of the candidate lattice libraries derived in Sec. 2.2.3.

This algorithm proceeds as follows. First, normalize the $\boldsymbol{\nu}$ vector by projecting it onto the unit hypersphere since only relative $\boldsymbol{\nu}$ magnitudes affect the phase as $T \rightarrow 0$. Then, taking the library of structures discussed previously in this section, minimize the difference of the energy of the target structure E_{target} and the minimal energy structure in the collection $\{E_{lib}(\boldsymbol{\nu}, \boldsymbol{\omega})\}$. Thus a vector ν_{RP} is sought that minimizes

$$\Delta E = E_{target}(\boldsymbol{\nu}) - \min\{E_{lib}(\boldsymbol{\nu}, \boldsymbol{\omega})\}. \tag{9}$$

We note that for a finite library, this gives the point in $\boldsymbol{\nu}$ space of the *largest* energy gap between E_{target} and any other structures in the library and is representative of the family of intermolecular potentials consistent with the target structure. For an infinite library, ΔE is bounded below by zero since $\{E_{lib}(\boldsymbol{\nu}, \boldsymbol{\omega})\}$ contains the target structure and is minimized with respect to ω while the target structure is held at fixed molecular orientation. We call it the representative parameter vector $\boldsymbol{\nu}_{RP}$.

While E_{target} is based on observed Euler angles $\boldsymbol{\omega}$ and is only a function of $\boldsymbol{\nu}$ (*i.e.* Eq. (8)), E_{lib} contains a library of structures whose energies are functions of $\boldsymbol{\nu}$ and $\boldsymbol{\omega}$. Since minimizing with respect to $\boldsymbol{\omega}$ is computationally demanding, we solve Eq. (9) in an iterative manner. First Eq. (9) is minimized for each experimental lattice with $\{E_{lib}\}$ based only on the energies of the remaining experimental lattices. For each trial $\boldsymbol{\nu}_{RP}$

found from this relative minimization, all additional candidate lattices are minimized with respect to their Euler angles at fixed $\boldsymbol{\nu}$ and the lowest energy solution is appended to the library. Again Eq. (9) is minimized for the energy of each experimental structure with respect to the augmented set to find a new trial RP in $\boldsymbol{\nu}$-space and the process iterates until the RPs converge.

Minimization methods using genetic algorithms (Holland, 1992) (differential evolution) and simulated annealing (Kirkpatrick *et al.*, 1983) are used on the Euler angle minimization and energy difference minimization, respectively. This is particularly important since the topologies of the energy and energy difference are both nonlinear with many local minima, the latter also having discontinuous derivatives. We reiterate that this is a relative energy minimization for ΔE in parameter space $\boldsymbol{\nu}$ and not a simulated annealing with fixed potential and varying temperature as is more commonly done.

Since the goal of our work is to determine the minimum truncation point that provides a sufficiently rich potential to span the observed crystal structures, we adopt an iterative procedure that begins with lower order manifolds and proceeds to higher orders if necessary. First, the energy difference is minimized using $\ell_{\max} = 3$ for all experimental structures. Only if a minimum is not found for a given structure are fourth manifold basis functions appended to the parameter space and the process repeated.

3. Discussion and Conclusion

3.1. Representative Potentials

Calculating potential parameters according to the procedure outlined in Sec. 2 gives the results in Table 3 where the potential coefficients are given for experimental structures in the four most prevalent reference lattices. Calculations for other reference

lattices such as the distorted diamond cubic reference lattices identified in part II of this series (McClurg & Keith, 2009) are not reported here since they represent a smaller percentage of distinct structures. The difference in energy between the experimental structure and the next lowest energy structure among the candidate lattices is shown in the second column.

For 8 of the 35 structures, $\Delta E < 0$. This indicates the target structure is not directly in the library but has been bounded in potential parameter space. Thus $\Delta E = 0$ on the borders, but $\Delta E < 0$ in the interior. $\boldsymbol{\nu}_{RP}$ gives a set of potential parameters where the target structure is relatively stable and is expected to be near the center of the region.

For 20 of the 35 structures, $\Delta E = 0$. This indicates the target structure is in the library and $\boldsymbol{\omega}_{\text{exp}} \approx \boldsymbol{\omega}_{\text{min}}$. Since $\Delta E = 0$ throughout the region where the target is stable, $\boldsymbol{\nu}_{RP}$ is an arbitrary point in the region and is not necessarily centrally located.

For 5 of the 35 structures, $0 < \Delta E < 0.01$ which is close enough to zero to conclude that a representative potential for these structures has been found. This can be asserted because the potential is composed of products of $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$, Clebsch-Gordan coefficients, spherical harmonics, and unit sphere normalized potential coefficients $\boldsymbol{\nu}$ that are all unitary. This makes the energy and therefore energy difference to be $\mathcal{O}(1)$. An energy difference less than this tolerance likely results from the method employed to determine orientational Euler angles for the experimental structures, $\boldsymbol{\omega}_{\text{exp}}$, and/or incomplete minimization in the determination of Euler angles at the global minimum, $\boldsymbol{\omega}_{\text{min}}$. Whereas $\boldsymbol{\omega}_{\text{exp}}$ is determined by analysis of the experimental structures, which is subject to small errors, the minimizer is free to vary $\boldsymbol{\omega}_{\text{min}}$ to attain the lowest possible energy. Thus the small positive energy differences in these cases are attributed to inaccuracies in the parameters used to describe the experimental structures and $\boldsymbol{\nu}_{RP}$ is therefore likely to be in the correct region of parameter space. A close approximation

to the experimental structure is contained in the comparison library, and therefore the RP may not be centrally located in its domain for these cases for the same reasons as discussed above for the $\Delta E = 0$ set.

For 2 of 35 structures, $\Delta E > 0.01$. This indicates the absence of potential parameters consistent with the experimental crystal structure relative to the other structures using the truncation specified. The $\boldsymbol{\nu}_{RP}$ listed identifies a point in parameter space where the target structure is least metastable compared to the global minimum in the library. The target structure is also usually quite similar to the global minimal structure. This is the case with YEMRIR, for example. It has a positive ΔE but is similar to the global minimum structure, space group 132 with molecules at Wyckoff point a. Both have two molecules per unit cell and one of the molecules, that at (0,0,0) in both structures, has the same orientation. The difference is the second molecule which is rotated $\sim 90^\circ$ between the two structures. Reasons for this difference may be related to truncation of the potential after the fourth manifold, lack of molecular displacements and lattice strains in the model, lack of explicit second nearest-neighbor interactions, the low-temperature limit discussed in Sec. 2.2.1, or missing quantum effects. Another reason may be that the reported experimental phase is in fact metastable in the limit $T \to 0$. Ostwald's step rule supports this possibility. The rule is based on the observation that during crystal nucleation and growth a metastable phase frequently crystallizes first which is closest in structure to the liquid or solution from which it is grown and which has the lowest free energy barrier. The crystallite then undergoes a series of phase changes until it reaches the thermodynamically stable structure (Schmalzried, 2003). The experimentally observed structure may have formed during an early stage of nucleation as one of these intermediate structures. No attempts have been made to attribute either of these two outlying cases to one or more of the above plausible explanations.

22 of the 35 distinct structures can be located in potential parameter space using only third manifold basis functions while 13 require fourth manifold functions. This is contrary to previous suggestions (Briels, 1980) that many manifolds (i.e. ℓ_i, ℓ_j go up to 10 or more) are needed to accurately describe tetrahedral molecules such as methane and adamantane. It is true that an unwieldly large number of parameters are needed if the parameters are determined using the orthogonality of the basis functions to project out each parameter through integration over its domain. This is the standard method for evaluating intermolecular potential parameters from an arbitrary intermolecular potential. The drawback to this method is that the integrations are dominated by contributions near singularities in the potential. Therefore the potential parameter values evaluated in this manner are strongly influenced by high energy interactions between molecules that are insignificant in determining the free energy of the crystal. The low energy interactions that dominate the free energy calculation are insignificant in evaluating the IP parameters evaluated in this way. The result is a slowly convergent potential with many additive contributions in the vicinity of singularities and offsetting contributions elsewhere. What is needed for useful GPDs is a method for evaluating IP parameters from an arbitrary intermolecular potential that stresses the important low energy interactions and results in large, but not necessarily infinite energies in the vicinity of singularities (Missaghi *et al.*, 2009). Current work is a proof by demonstration.

Truncation of the intermolecular potential after the fourth manifold leaves up to 15 potential parameters which is far fewer than the previous assertions, but still too many to easily visualize. Software such as GGOBI facilitates investigations of high dimensional spaces in an interactive way (Lang, 2003). Any number of axes may be displayed and spatial relationships between the RPs analyzed using the software.

Interpreting the potential coefficients in Table 3 is analogous to describing multi-

pole electrostatic interactions. These potential coefficients include contributions from all interaction modes and are not limited to electrostatic interactions, however. In the case of tetrahedral molecules the octopole ($\ell_i = 3$) is the first nonzero multipole. Coefficients such as $\nu_{0,3,3}$ and $\nu_{0,4,4}$ where ℓ_i or ℓ_j is zero represent an octopole or hexadecapole of one molecule interacting with the zeroth pole of its neighbors. They are the crystal field coefficients. As the basis functions may also be used as quantum basis sets, the $\boldsymbol{\nu}$ may be calculated *ab initio* and have been given physical interpretations via symmetry-adpated perturbation theory (van der Avoird *et al.*, 1994). For instance, although dispersion and induction forces can be found in all components of $\boldsymbol{\nu}$, electrostatic forces contribute only to $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell_i+\ell_j,\ell_j}$ (Stone & Tough, 1984) such as $\nu_{3,6,3}$, $\nu_{3,7,4}$, and $\nu_{4,8,4}$.

One drawback of using the algorithm in Sec. 2.2 is that, by finding the maximum energy difference between a phase and all others, the RPs of neighboring phases tend to be spread apart. This is because an RP is a vector representing a region in a space, not a unique set of IP parameters. Homologous series of molecules (*i.e.* CF_4, CCl_4, CBr_4, CI_4) are expected to show trends in $\boldsymbol{\nu}$ space. If members of homologous series have different crystal structures, then they have different RPs. However, different RPs are widely spread by our algorithm. Therefore RPs of homologous series are often widely spaced, even if the molecules are expected to have similar intermolecular potentials. This hides the expected trends among homologues. Other methods are required to further constrain the RPs assigned to individual molecules before trends among homologous series can be investigated using GPDs.

Correlations are evident in reference lattice assignments. An example of this is a series of molecules with molecular formula $C_{16}H_{36}X_4In_4N_4$ where X is Cl, Br, or I. The first two structures, 14e MECKUA and 12i MECKOU, pack in an fcc reference lattice while the last, 11e MECKIO, packs in a tetragonal reference lattice intermediate

between fcc and bcc which is slightly closer to the bcc reference lattice. This indicates at least the translational part of homologue potentials is similar and that small changes in the atomic constituency can slightly alter the intermolecular potential (IP), but significantly alter crystal structure. This is consistent with experience that similar molecules often have very different crystal structures despite similar intermolecular potentials. GPDs acknowledge this by placing the seemingly disparate structures close to one another in IP parameter space.

In (McClurg & Keith, 2009) we noted that structures 195a and 197a were very similar to higher symmetry structures (215a and 217a). If the molecules remained tetrahedral then the crystal would retain higher symmetry but the reported atomic coordinates indicate a minor molecular distortion reducing the space group symmetry. Assuming that the reported space group is correct, these structures require a symmetry-breaking pathway from sc and bcc with a minimal truncation manifold of $\ell_i^{\max} = 6$ or 9 (depending on the potential-dependent transition pathway taken). In contrast to these high manifold requirements most pathways require a third or fourth manifold basis set. In view of the success of finding RPs using only the two first manifolds ($\ell_i^{\max} = 3$ or 4) for all other molecules in the data set, it seems unlikely such a large number of rapidly oscillating basis functions would be required to properly describe the intermolecular potential for these structures. It seems more likely that, barring a Jahn-Teller crystal distortion, the crystal symmetry may have been underspecified when reported to the CSD as suggested in (McClurg & Keith, 2009) on different grounds.

3.2. Extensions and Features of the Methodology

We now discuss some of the ways to improve the model and algorithms. An issue affecting the numerical accuracy of the RP is how large a library of alternate crys-

tal structures is needed to localize the RP in $\boldsymbol{\nu}$ space. In our previous work (Mettes *et al.*, 2004) we considered just the high-symmetry point isotropy subgroups and in this work have followed suit since these are the most common (Stokes & Hatch, 1988). Recall that space group IRs are indexed by reciprocal space vectors (Kovalev *et al.*, 1993; Zak *et al.*, 1969). There are the same number of $\boldsymbol{k}$ points in reciprocal space as there are unit cells in the crystal and there is a correspondence between $\boldsymbol{k}$ points and supercell patterns in the crystal. Experimental structures are supercells in their reference lattice. As nearly all experimental structures contain a relatively small number of clustered parent lattice unit cells as sublattices, symmetry-breaking occurs at $\boldsymbol{k}$ vectors corresponding to this small cluster. These are $\boldsymbol{k}$ vectors at high symmetry points in the reference lattice Brillouin zone. If the experimental structure has a unit cell which is large or flattened/elongated, however, $\boldsymbol{k}$ vectors corresponding to this larger or longer/flatter group of parent lattice unit cells will be on high symmetry lines and planes. Symmetry-breaking pathways for experimental lattices pertaining to bcc and fcc reference lattices in the appendix show that some experimental structures contain large (i.e. CTBROM (More *et al.*, 1977)) or non-clustered parent lattice unit cells (i.e. MTRETC10 (Harrison *et al.*, 1972)) and their $\boldsymbol{k}$ vectors therefore are from high-symmetry plane and line IRs. Although these cases are less common, such isotropy subgroups could be included in the candidate lattices in Sec. 2.2.3. Providing such additional structures in the library would place additional constraints on the RP for the observed structures and therefore would further localize the RP for each observed structure at the expense of a much larger library.

In Sec. 2.2.3 we chose to consider only isotropy subgroups with one occupied Wyckoff point. This is a commonly used simplification (Verwer & Leusen, 1998). We have also discarded coupled IR isotropy subgroups for the same reason. Our method could be extended to include library structures with multiple Wyckoff points and those from

coupled IRs, although the computational demands are much higher because of the larger set of candidate lattices and so we do not pursue it here. The effect would be to further localize the RPs of each phase, at the expense of a much larger library.

Although tetrahedral molecules are used in the current example any molecular point group could be used without a dramatic increase in the number of potential coefficients $\boldsymbol{\nu}$ in the first two non-trivial manifolds. This is because the absence of IP coefficients in lower manifolds for high-symmetry molecules is offset by the larger number of parameters in higher manifolds. Consider the number of basis functions in the first two non-trivial manifolds of I_h (the icosahedral group) and C_1 (the point group of no special symmetry). These are the highest and lowest molecular point group symmetries, respectively. The first three manifolds of I_h containing a totally symmetric molecular representation are the zeroth, sixth, and tenth manifolds. The number of potential coefficients is seven on the sixth manifold, $\{\nu^{1,1}_{6,0,6}, \nu^{1,1}_{6,2,6}, ...\nu^{1,1}_{6,12,6}\}$, eleven on the tenth manifold, $\{\nu^{1,1}_{10,0,10}, \nu^{1,1}_{10,2,10}, ...\nu^{1,1}_{10,20,10}\}$, and seven for the cross manifolds, $\{\nu^{1,1}_{6,4,10}, \nu^{1,1}_{6,6,10}, ...\nu^{1,1}_{6,16,10}\}$. There are also two crystal field coefficients, $\{\nu^{1,1}_{0,6,6}, \nu^{1,1}_{0,10,10}\}$, giving a total of 27 potential coefficients for I_h. For C_1 the first three manifolds are the zeroth, first, and second. The coefficients on the first manifold are $\{\nu^{n_\sigma,n_\mu}_{1,0,1}, \nu^{n_\sigma,n_\mu}_{1,2,1}\}$. Although there are three copies of the totally symmetric molecular representation on the first manifold and therefore the molecular frame indices σ and μ in Eq. (1) go over $\{1,2,3\}$, we are free to choose the standard orientation for the molecule corresponding to Euler angles $\{0,0,0\}$. If it is chosen with the IP major axis parallel to the laboratory z-axis then only one of these is nonzero. This leaves two coefficients $\{\nu^{1,1}_{1,0,1}$, $\nu^{1,1}_{1,2,1}\}$. If the molecular minor axis in the standard orientation is oriented parallel to the laboratory x-axis only four of the σ and μ are nonzero in the second manifold. This leaves 48 coefficients on the second manifold, $\{\nu^{n_\sigma,n_\mu}_{2,0,2}, \nu^{n_\sigma,n_\mu}_{2,2,2}, \nu^{n_\sigma,n_\mu}_{2,4,2}\}$. With five additional crystal field coefficients $\{\nu^{1,1}_{0,1,1}, \nu^{1,n_\mu}_{0,2,2}\}$ and eight cross manifold coefficients

$\{\nu_{1,1,2}^{1,n_\mu}, \nu_{1,3,2}^{1,n_\mu}\}$ the total number of coefficients on the first three manifolds of C_1 is 60, an increase of roughly twofold from highly symmetrical I_h. This shows that although this method has been applied to tetrahedral molecules, it is applicable to other molecular point group symmetries with only a modest change in the number of parameters.

Throughout we have implemented a simple direct cutoff truncation scheme in which a doubly infinite summation is cut off at a maximum manifold number $\ell_i^{\max}$. An alternative is the manifold-sum cutoff in which we truncate such that $\ell_i + \ell_j \leq \ell^{\max}$. This truncation would include smoother functions before more rapidly oscillating ones which is consistent with our expectations of lower-energy electronic contributions to the potential. Also the number of potential parameters increases at a slower rate with this truncation. This is shown in Table 4 for both truncation schemes where we compare the cumulative number of parameters for the C_1, T_d, and I_h point groups at different manifolds. The truncation method used in this study is a square truncation of the double sum while the alternative is a triangular truncation. The manifold sum truncation adds new parameters into the potential more slowly than single manifold truncation. Therefore, the 15 coefficients used here are sufficient but may not be necessary. Further investigation of global phase diagrams with ℓ-sum truncation is needed to test this hypothesis. This is important since lower dimensional GPDs would be easier to construct and to use.

From the foregoing discussion we have seen 15 coefficients are sufficient to locate our dataset of experimental structures in potential parameter space. We have not investigated if a linear combination would be better. Principle component analysis could be used to identify linear combinations of basis functions that better fit molecules. It is possible that fewer than 15 coefficients are necessary.

An analog in the physics community to molecular crystal GPDs are the various Ising-type GPDs, such as Axial Next Nearest Neighbor Ising (ANNNI) GPDs. Despite

their near-trivial potentials, these have been used for decades to explore a rich variety of commensurate and incommensurate phases (Gendiar & Nishino, 2005; Kasama *et al.*, 2006; Y. Muraoka *et al.*, 1999; Sato & Matsubara, 1999; Kim *et al.*, 1995) that can be applied to realistic materials such as silicon carbide, titanium aluminum, group-IV element polytypes, and many others (Wong-Leung *et al.*, 2005; Colinet & Pasturel, 2002; Raffy *et al.*, 2002). These differ from molecular crystal GPDs in that their basis functions are simple binary variables $S_i \in \{1, -1\}$ coupled by a constant J, although such couplings typically exist with next-nearest neighbors or larger. It will be interesting to compare the results of these simplified models with a full molecular crystal GPD in our upcoming work (Keith *et al.*, 2009).

A further extension would be to include the intersection of each solid GPD with a fluid GPD. This would not only allow crystallographers to identify the first phase to precipitate from solution, but would also allow inclusion of specific crystallization conditions and kinetic effects. Thus the calculation of the exact location of the liquid/solid surface could be performed with as complex or simple a crystal growth model as one desires, using not only consideration of surface free energies, crystal habit, and surface defect structure, but also external conditions such as undercooling, seeding, etc. In theory even kinetic barriers to crystallization could be estimated between solution and GPD candidate phases, thus suggesting a more realistic progression of polymorph crystallization since GPDs find not only the thermodynamic ground state but also all metastable states.

3.3. Conclusions

We have shown that a molecular crystal global phase diagram (GPD) with a small number of reference lattices derived in (McClurg & Keith, 2009) can summarize the experimental data using a modest number of IP parameters. The data set is diverse

enough to test the GPD's ability to classify a wide range of space groups using a common intermolecular potential. Just as the van Konyenburg global phase diagram classification based on the simple van der Waals equation of state is nonetheless widely used to classify the phase behavior of real binary mixtures, molecular crystal global phase diagrams may be useful in elucidating phase behavior of a variety of real substances and, in turn, used to develop novel intermolecular potentials and materials. Further work to apply these results in constructing temperature-dependent GPDs is needed and is underway. Code to construct GPDs may be requested from the corresponding author.

This work received financial support from the American Chemical Society - Petroleum Research Fund (PRF #41774-AC10), DOE Grant No. DE-FG02-03ER46059, and NSF Grant DMR-0520547. Computational resources maintained by the University of Minnesota Supercomputer Institute were used for portions of this research.

References

Allen, F. H. & Motherwell, W. D. S. (2002). *Acta Cryst. B*, **58**, 407–422.

Aparicio-Martinez, S. & Hall, K. R. (2007*a*). *Ind. Eng. Chem. Res.* **46**, 273–284.

Aparicio-Martinez, S. & Hall, K. R. (2007*b*). *Ind. Eng. Chem. Res.* **46**, 285–290.

van der Avoird, A., Wormer, P. E. S. & Moszynski, R. (1994). *Chem. Rev.* **94**, 1931–1974.

van der Avoird, A., Wormer, P. E. S., Mulder, F. & Berns, R. M. (1980). *Topics in Current Chemistry*, **93**, 1–50.

Bolz, A., Deiters, U. K., Peters, C. J. & de Loos, T. W. (1998). *Pure Appl. Chem.* **70**, 2233.

Bradley, C. J. & Cracknell, A. P. (1972). *The Mathematical Theory of Symmetry in Solids*. New York: Oxford University Press.

Braga, D., Grepioni, F. & Orpen, A. G. (eds.) (1999). *Crystal Engineering: From Molecules and Crystals to Materials*. Boston: Kluwer Academic.

Briels, W. J. (1980). *J. Chem. Phys.* **73**, 1850–1861.

Cismondi, M. & Michelsen, M. L. (2007). *J. Supercrit. Fluid.* **39**, 287–295.

Colinet, C. & Pasturel, A. (2002). *Intermetallics*, **10**, 751.

Desiraju, G. R. (1989). *Crystal Engineering: The Design of Organic Solids*. New York: Elsevier.

Gavezzotti, A. (2007). *Molecular Aggregation: Structure Analysis and Molecular Simulation of Crystals and Liquids*. USA: Oxford University Press.

Gdanitz, R. J. (1997). In *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, edited by A. Gavezzotti, pp. 185–201. New York: Wiley.

Gendiar, A. & Nishino, T. (2005). *Phys. Rev. B*, **71**, 024404.

Grazulis, S., Chateigner, D., Downs, R. T., Yokochi, A. F. T., Quiros, M., Lutterotti, L., Manakova, E., Butkus, J., Moeck, P. & Le Bail, A. (2009). *J. Appl. Cryst.* **42**, 726–729.

Harrison, W., Marsh, W. & Trotter, J. (1972). *J. Chem. Soc., Dalton Trans.* p. 1009.

Hloucha, M., Lodge, J., Lenhoff, A. & Sandler, S. (2001). *Cryst.* **64**, 060104(R)–1–060104(R)–4.

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Cambridge: The MIT Press.

Holman, K. T., Pivovar, A. M., Swift, J. A. & Ward, M. D. (2001). *Acc. Chem. Res.* **34**, 107–118.

James, H. M. & Keenan, T. A. (1959). *J. Chem. Phys.* **31**, 12–41.

Kasama, T., Muraoka, Y. & Idogaki1, T. (2006). *Phys. Rev. B*, **73**, 214411.

Keith, J. B., McClurg, R. B. & Fultz, B. (2009). In preparation.

Keith, J. B., Mettes, J. A. & McClurg, R. B. (2004). *Crystal Growth and Design*, **4**, 1009–1012.

Kim, J.-J., Mori, S. & Harada, I. (1995). *Phys. Lett. A*, **202**, 68.

Kirkpatrick, S., Gelatt Jr, C. & Vecchi, M. (1983). *Science*, **220**, 671 – 680.

van Konyenburg, P. H. & Scott, R. L. (1980). *Philos. Trans. R. Soc. London, Ser. A*, **298**, 495–540.

Kovalev, O. V., Worthey, G. C., Stokes, H. T. & Hatch, D. M. (1993). *Representation of Crystallographic Space Groups: Irreducible Representations, Induced Representation and Corepresentations*. Berkshire, Great Britian: Gordon and Breach Science Publishers.

Lamoen, D. & Michel, K. H. (1999). *Phase Transitions*, **67**, 789–807.

Lang, D. T., (2003). Ggobi. www.ggobi.org.

Lin, J. Y. Y. & Keith, J. B., (2009). Virtual neutron facility. vnf.caltech.edu.

Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Mooij, W. T. M., Price, S. L., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2000). *Acta Crystallogr. B*, **56**, 697–714.

Lynden-Bell, R. M. & Michel, K. H. (1994). *Rev. Mod. Phys.* **66**, 721–762.

McClurg, R. B. & Keith, J. B. (2009). Molecular crystal global phase diagrams II: Reference lattices. Submitted.

Mettes, J. A., Keith, J. B. & McClurg, R. B. (2004). *Acta Cryst. A*, **60**, 621–636.

Michel, K. H., Copley, J. R. D. & Neumann, D. A. (1992). *Phys. Rev. Lett.* **68**, 2929–2932.

Missaghi, M. N., Mettes, J. A. & McClurg, R. B. (2009). In Preparation.

More, M., Baert, F. & Lefebvre, J. (1977). *Acta Cryst. B*, **33**, 3681–3684.

Moulton, B. & Zaworotko, M. J. (2001). *Chem. Rev.* **101**, 1629–1658.

Neumann, M. A., Press, W., Noldeke, C., Asmussen, B., Prager, M. & Ibberson, R. M. (2003). *J. Chem. Phys.* **119**, 1586–1589.

Padmaja, N., Ramakumar, S. & Viswamitra, M. A. (1990). *Acta Cryst. A*, **46**, 725–730.

van Pelt, A., Peters, C. J., Swaan-Arons, J. & Dieters, U. K. (1995). *J. Chem. Phys.* **102**, 3361.

Polishuk, I., Wisniak, J. & Segura, H. (2002). *Phys. Chem. Chem. Phys.* **4**, 879.

Polishuk, I., Wisniak, J., Segura, H., Yelash, L. V. & Kraska, T. (2000). *Fluid Phase Equilib.* **172**, 1–26.

Raffy, C., Furthmuller, J. & Bechstedt, F. (2002). *Phys. Rev. B*, **66**, 075201.

Sato, A. & Matsubara, F. (1999). *Phys. Rev. B*, **60**, 10316.

Schmalzried, H. (2003). *Zeit. fur Physik. Chem.* **217**, 1281–1302.

Sherwood, J. N. (1979). *The Plastically Crystalline State*. New York: Wiley.

Simon, J. Bassoul, P. (2000). *Design of Molecular Materials: Supramolecular Engineering*. New York: John Wiley and Sons.

Stokes, H. T. & Hatch, D. M. (1988). *Isotropy Subgroups of the 230 Crystallographic Space Groups*. Teaneck, NJ: World Scientific.

Stokes, H. T. & Hatch, D. M. (2002). *Phys. Rev. B*, **65**, 144114–1–144114–12.

Stokes, H. T., Hatch, D. M. & Campbell, B. J., (2007). Isotropy. stokes.byu.edu/isotropy.html.

Stokes, H. T., Hatch, D. M. & Wells, J. D. (1991). *Phys. Rev. B.* **43**, 11010–11018.

Stone, A. J. & Tough, R. J. A. (1984). *Chem. Phys. Lett.* **110**, 123–129.

Thalladi, V. R., Goud, B. S., Hoy, V. J., Allen, F. H., Howard, J. A. K. & Desiraju, G. R. (1996). *Chem. Commun.* pp. 401–402.

Varshalovich, D. A., Moskalev, A. N. & Khersonskii, V. K. (1988). *Quantum Theory of Angular Momentum*. Teaneck, NJ: World Scientific.

Verwer, P. & Leusen, F. J. J. (1998). In *Reviews in Computational Chemistry*, edited by K. B. Lipkowitz & D. B. Boyd, vol. 12, pp. 327–366. New York: Wiley-VCH.

Wong-Leung, J., Linnarsson, M. K., B. G. Svensson, B. G. & Cockayne, D. J. H. (2005). *Phys. Rev. B*, **71**, 165210.

Y. Muraoka, Y., Oda, K. & Idogaki, T. (1999). *J. Magn. Magn. Mater.* **195**, 156.

Zak, J., Casher, H., Gluck, H. & Gur, Y. (1969). *The Irreducible Representations of Space Groups*. New York: Benjamin.

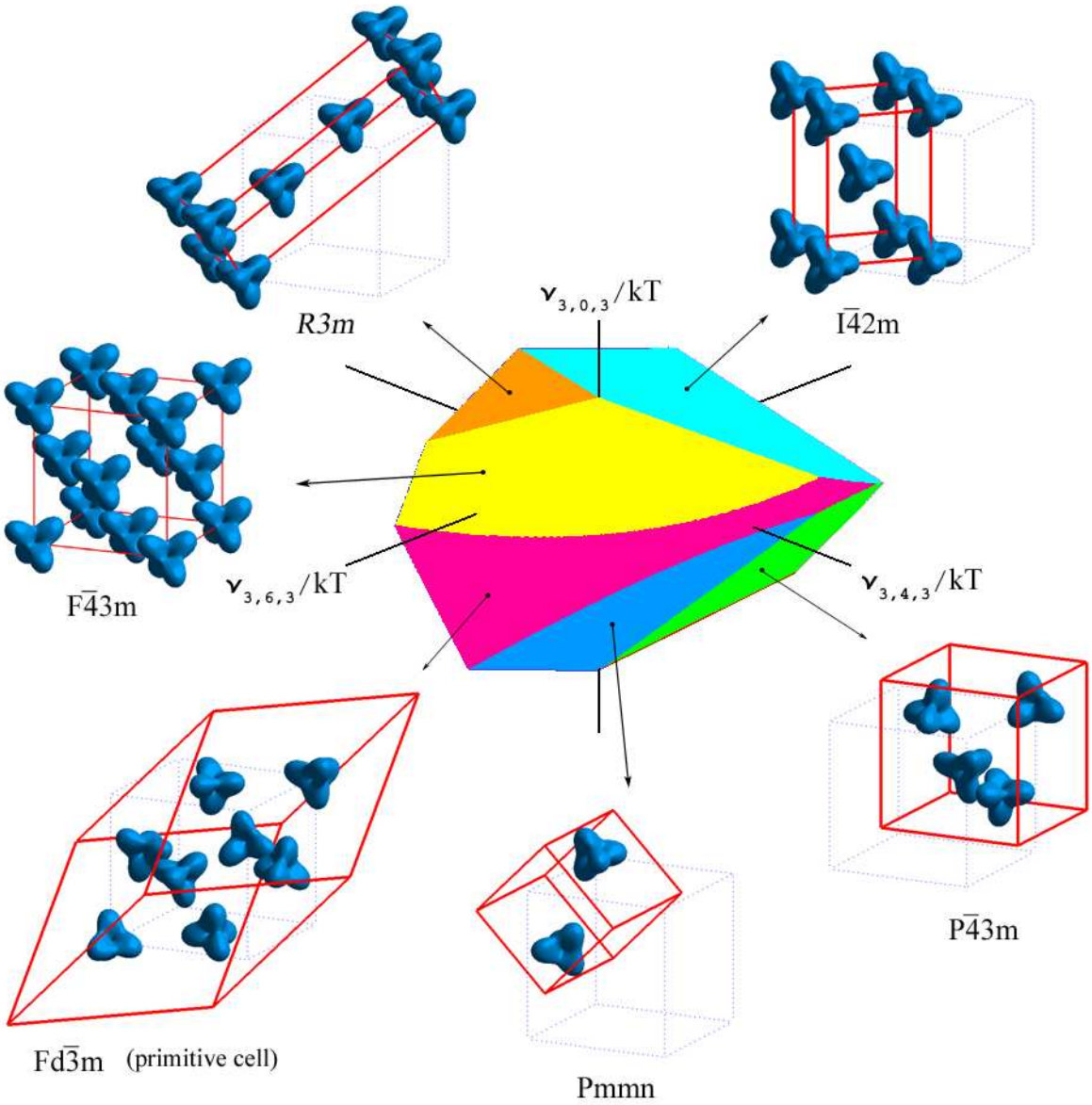

Fig. 1. [111] view of a Global Phase Diagram (GPD) for tetrahedral molecules in an fcc reference lattice. Axes are intermolecular potential coefficients $\boldsymbol{\nu}$ scaled by kT. The origin of the axes is the infinite-temperature reference state. Rays protruding from the surfaces pointing at distinct phases represent ratios of potential coefficients that form the indicated space groups. Colored surfaces are boundaries between daughter phases and the reference fcc phase. Reproduced with permission of the International Union of Crystallography from (Mettes *et al.*, 2004)

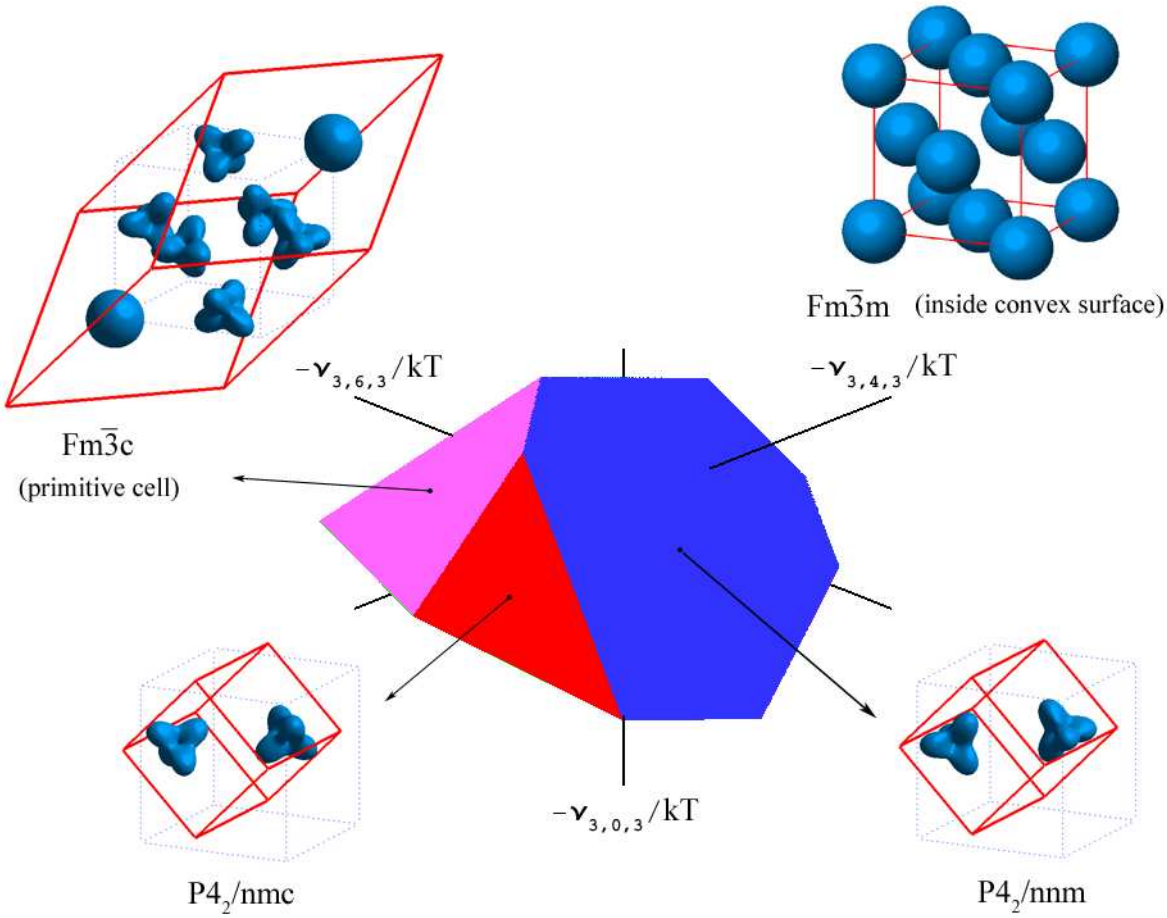

Fig. 2. $[\bar{1}\bar{1}\bar{1}]$ view of a GPD for tetrahedral molecules in an fcc reference lattice. Some molecules such as those in the $Fm\bar{3}m$ structure appear "spherical" because of rapid molecular reorientation. Reproduced with permission of the International Union of Crystallography from (Mettes *et al.*, 2004)

Table 1. *Presence of O_h and D_{3h} point group IRs for various manifolds of SO(3). O_h is the Wyckoff point group of the fcc, bcc, and sc reference lattices. D_{3h} is the Wyckoff point group of the hcp reference lattice. The first occurrence of a given IR is shown in bold.*

ℓ_i	O_h IRs	D_{3h} IRs
0	$\boldsymbol{A_{1g}}$	$\boldsymbol{A_1'}$
3	$\boldsymbol{A_{2u}} + \boldsymbol{T_{1u}} + \boldsymbol{T_{2u}}$	$A_1' + \boldsymbol{A_2'} + \boldsymbol{E'} + \boldsymbol{E''}$
4	$A_{1g} + \boldsymbol{E_g} + \boldsymbol{T_{1g}} + \boldsymbol{T_{2g}}$	$A_1' + \boldsymbol{A_1''} + \boldsymbol{A_2''} + 2E' + E''$
6	$A_{1g} + \boldsymbol{A_{2g}} + E_g + T_{1g} + 2T_{2g}$	$2A_1' + A_1'' + A_2' + A_2'' + 2E' + 2E''$
7	$A_{2u} + \boldsymbol{E_u} + 2T_{1u} + 2T_{2u}$	$A_1' + A_1'' + A_2' + 2A_2'' + 3E' + 2E''$ ‘
8	$A_{1g} + 2E_g + 2T_{1g} + 2T_{2g}$	
9	$\boldsymbol{A_{1u}} + A_{2u} + E_u + 3T_{1u} + 2T_{2u}$	

Table 2. *Group theoretical symmetry-breaking pathways of experimental lattices for the fcc reference lattice. Classified by space group IR and order parameter direction, each pathway shows the point group IR and minimal manifold of SO(3) in Eq. (1) required to achieve it. The order parameter directions are given in an abbreviated form in the notation of Stokes and Hatch (Stokes & Hatch, 2002). The pathways, space group IRs, order parameter directions, and point group IRs were computed using ISOTROPY. However, transitions belonging to a coupled IR between a high symmetry point and line are currently not a feature of ISOTROPY. These entries have been marked with an asterisk in the table.*

Pathway	S. G. IR	OP Dir.	P. G. IR	$\ell^{\mathrm{req'd}}$
225a → 121a	Γ_5^-	P1	T_{2u}	3
225a → 142a	W_3	P2	T_{1g}, A_{2u}, E_u	3
225a → 114a	$X_2^- \oplus \Gamma_5^-$	P1 ⊕ P1	$A_{2u}, E_u \oplus T_{2u}$	3
	$X_3^+ \oplus \Gamma_5^-$	P1 ⊕ P1	$T_{1g} \oplus T_{2u}$	4
	$X_2^- \oplus X_3^+$	P1 ⊕ P1	$A_{2u}, E_u \oplus T_{1g}$	4
225a → 152b	Λ_3 (k=1/3)	P1	$E_g, T_{1g}, T_{2g}, E_u, T_{1u}, T_{2u}$	3
225a → 15e (REKYUB)	L_3^-	P7	E_u, T_{1u}, T_{2u}	3
225a → 12i	L_3^-	P2	E_u, T_{1u}, T_{2u}	3
225a → 64d,f	*			
225a → 14e (MECKUA)	$\Gamma_3^+ \oplus L_3^-$	P1 ⊕ C8	$E_g \oplus E_u, T_{1u}, T_{2u}$	4
	$\Gamma_3^+ \oplus X_5^-$	C1 ⊕ S7	$E_g \oplus T_{1u}, T_{2u}$	4
	$\Gamma_4^+ \oplus X_5^-$	P1 ⊕ C11 P1 ⊕ S7	$T_{1g} \oplus T_{1u}, T_{2u}$	4
	$\Gamma_5^+ \oplus L_3^-$	C2 ⊕ C8	$T_{2g} \oplus E_u, T_{1u}, T_{2u}$	4
	$\Gamma_5^+ \oplus X_5^-$	P1 ⊕ C11 P1 ⊕ S7	$T_{2g} \oplus T_{1u}, T_{2u}$	4
	$L_2^+ \oplus L_1^-$	P1 ⊕ P1	$A_{2g}, T_{1g} \oplus A_{1u}, T_{2u}$	4
	$L_2^+ \oplus L_3^-$	P1 ⊕ P7	$A_{2g}, T_{1g} \oplus E_u, T_{1u}, T_{2u}$	4
	$L_2^+ \oplus X_2^-$	P1 ⊕ P1	$A_{2g}, T_{1g} \oplus A_{2u}, E_u$	4
	$L_2^+ \oplus X_3^-$	P1 ⊕ P1	$A_{2g}, T_{1g} \oplus E_u, T_{1u}$	4
	$L_2^+ \oplus X_5^-$	P1 ⊕ P1	$A_{2g}, T_{1g} \oplus T_{1u}, T_{2u}$	4
	$L_3^+ \oplus L_1^-$	P7 ⊕ P1	$E_g, T_{1g}, T_{2g} \oplus A_{1u}, T_{2u}$	4
	$L_3^+ \oplus L_3^-$	P7 ⊕ P7	$E_g, T_{1g}, T_{2g} \oplus E_u, T_{1u}, T_{2u}$	4
	$L_3^+ \oplus X_2^-$	P7 ⊕ P1	$E_g, T_{1g}, T_{2g} \oplus A_{2u}, E_u$	4
	$L_3^+ \oplus X_3^-$	P7 ⊕ P1	$E_g, T_{1g}, T_{2g} \oplus E_u, T_{1u}$	4
	$L_3^+ \oplus X_5^-$	P7 ⊕ P1	$E_g, T_{1g}, T_{2g} \oplus T_{1u}, T_{2u}$	4
	$L_1^- \oplus L_3^-$	P1 ⊕ C8	$A_{1u}, T_{2u} \oplus E_u, T_{1u}, T_{2u}$	4
	$L_1^- \oplus X_2^-$	P1 ⊕ P1	$A_{1u}, T_{2u} \oplus A_{2u}, E_u$	4
	$L_1^- \oplus X_3^-$	P1 ⊕ P1	$A_{1u}, T_{2u} \oplus E_u, T_{1u}$	4
	$L_1^- \oplus X_5^-$	P1 ⊕ P1	$A_{1u}, T_{2u} \oplus T_{1u}, T_{2u}$	4
	$L_3^- \oplus X_2^-$	P7 ⊕ P1	$E_u, T_{1u}, T_{2u} \oplus A_{2u}, E_u$	4
	$L_3^- \oplus X_3^-$	P7 ⊕ P1	$E_u, T_{1u}, T_{2u} \oplus E_u, T_{1u}$	4
	$L_3^- \oplus X_5^-$	P7 ⊕ P1	$E_u, T_{1u}, T_{2u} \oplus T_{1u}, T_{2u}$	4
	$X_2^- \oplus X_5^-$	P1 ⊕ C11 P1 ⊕ S7	$A_{2u}, E_u \oplus T_{1u}, T_{2u}$	4
	$X_3^- \oplus X_5^-$	P1 ⊕ C11	$E_u, T_{1u} \oplus T_{1u}, T_{2u}$	4
225a → 14e (TOHSUE)	Δ_5 ($k = 1/4$)	C7	$T_{1g}, T_{2g}, T_{1u}, T_{2u}$	3
225a → 15f,f,f,f	C_2 ($k_1 = 1/4, k_2 = 3/4$)	C18,C19	$A_{2g}, E_g, T_{1g}, T_{2g}$, $A_{1u}, E_u, T_{1u}, T_{2u}$	3

Table 3. *RP components for crystals of tetrahedral molecules in the CSD. The "Identifier" is a representative structure for a given structural type. The "Multiplicity" is the number of such structures in our dataset. The ℓ_i truncation value and the presence, absence, or proximity of a global minimum are indicated. The parameters ν_{ℓ_i,ℓ,ℓ_j} are indexed according to Eq. (1) and have been mapped to the unit hypersphere.*

Identifier	Multi-plicity	ΔE	ν_{033}	ν_{044}	ν_{303}	ν_{323}	ν_{343}	ν_{363}	ν_{314}	ν_{334}	ν_{354}	ν_{374}	ν_{404}	ν_{424}	ν_{444}	ν_{464}	ν_{484}
bcc																	
217a HXMTAM07	11	0.000	-	0.000	0.970	0.000	0.230	-0.0790	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
161a TCYMET	1	0.007	-	0.355	0.303	0.380	-0.017	-0.130	-0.213	0.084	0.331	-0.116	0.409	0.496	0.075	-0.138	-0.089
2i MEZDIE01	1	0.001	-	-0.142	0.017	-0.215	0.042	0.222	0.221	-0.314	0.483	0.230	-0.287	-0.309	-0.019	-0.158	0.495
60c,d YIMWEW	1	-0.015	-	0.003	0.0	0.0	0.004	0.00	0.998	-0.007	-0.059	-0.008	0.002	0.0	-0.002	-0.010	0.009
2iii OHABEE	1	0.024	-	0.092	-0.003	-0.159	-0.003	0.016	-0.504	0.089	-0.348	0.087	-0.108	-0.102	-0.288	-0.239	0.643
fcc																	
121a ZZZKDW01	1	0.000	-	0.000	0.431	0.316	0.267	-0.802	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
142a KUJSIR	1	0.001	-	0.000	-0.935	-0.054	0.147	-0.319	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
114a ADAMAN08	2	0.018	-	-0.035	-0.007	-0.292	0.000	0.023	0.061	-0.046	0.191	0.053	-0.056	0.584	0.185	0.055	-0.698
152b MTRETC10	1	-0.080	-	-0.241	0.022	-0.082	0.106	-0.139	-0.909	0.002	0.109	-0.244	-0.001	-0.045	0.012	-0.067	-0.003
15e REKYUB	1	0.003	-	-0.210	-0.001	-0.359	-0.011	0.031	-0.157	0.135	-0.463	0.379	-0.209	-0.445	-0.340	0.157	0.208
12i MECKOU	1	0.000	-	-0.258	0.006	0.781	0.028	-0.058	-0.216	-0.056	-0.264	0.215	-0.194	0.286	0.166	0.029	0.074
64d,f METHANEIII	1	-0.238	-	0.112	-0.010	0.017	-0.020	0.006	0.001	-0.022	-0.002	-0.014	0.097	0.060	-0.772	-0.483	-0.379
14e MECKUA	1	0.005	-	-0.123	-0.070	-0.103	0.161	-0.744	-0.240	-0.084	-0.209	-0.345	0.008	-0.251	-0.056	-0.305	0.060
14e TOHSUE	1	-0.052	-	0.317	-0.011	-0.036	0.030	-0.137	0.615	0.221	-0.338	0.477	0.035	-0.163	-0.024	-0.003	0.284
15f,f,f,f CTBROM	2	-0.186	-	0.197	0.005	-0.044	-0.025	-0.151	-0.687	0.291	-0.284	0.486	0.037	-0.062	0.038	-0.077	0.222
hcp																	
165d DILWIE01	2	-0.034	-0.839	0.000	0.075	0.302	0.296	-0.334	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
147d ZIZHIZ	1	0.000	-0.793	0.000	-0.040	0.203	0.083	-0.568	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
176h CUCZUV	1	0.009	0.000	0.000	-0.251	-0.965	0.071	-0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
sc																	
215a FOHCUA	3	0.000	-	0.000	0.960	0.000	0.279	-0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
120c YEMRIR	1	0.016	-	-0.102	-0.111	-0.167	-0.439	0.077	0.403	-0.487	-0.078	0.281	0.169	0.056	-0.071	0.449	-0.167

Table 4. *Comparison of the cumulative number of parameters in potential for a truncation at a given manifold for two truncation strategies and three molecular point groups.*

Direct Truncation				ℓ sum Truncation			
ℓ_i	C_1	T_d	I_h	$\ell_i + \ell_j$	C_1	T_d	I_h
0	*	*	*	0	*	*	*
1	3	0	0	1	1	0	0
2	60	0	0	2	6	0	0
3		5	0	3	15	1	0
4		10	0	4		1	0
6			8	6		6	1
10			19	10			8

* There is one isotropic basis function for $\ell_i = \ell_j = 0$ in each case, but it does not drive orientational ordering.

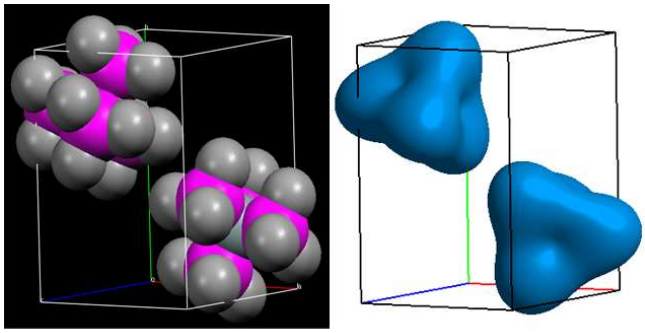

Fig. 3. Comparison of an experimental structure, tetrakis(trimethystannyl)silane [CSD structure MEZDIE01], which crystallizes in space group 2 at Wyckoff point i with arbitrarily-shaped tetrahedral figures whose orientation is determined by orientational energy minimization with molecular center of mass on an ideal bcc reference lattice.

Appendix A
Detailed derivation of Eq. 1

We choose to expand the potential using a complete set of two-molecule basis functions that span the rotational space $SO(3)$ parameterized by Euler angles $\boldsymbol{\omega}$ of molecule i, j and their intervening orientational space S^2 parameterized by solid angle $\boldsymbol{\Omega}_{ij} = (\theta_{ij}, \phi_{ij})$. Coupling two $SO(3)$ irreducible representations (IR's) $D^{\ell_i}_{m_i n_i}(\boldsymbol{\omega}_i)$ and $D^{\ell_j}_{m_j n_j}(\boldsymbol{\omega}_j)$ and a spherical harmonic $C^{\ell}_{m}(\boldsymbol{\Omega}_{ij})$ gives

$$W^{n_i n_j}_{\ell_i \ell \ell_j} = \sum_{m_i m m_j} \begin{pmatrix} \ell_i & \ell & \ell_j \\ m_i & m & m_j \end{pmatrix} D^{\ell_i}_{m_i n_i}(\boldsymbol{\omega}_i) C^{\ell}_{m}(\boldsymbol{\Omega}_{ij}) D^{\ell_j}_{m_j n_j}(\boldsymbol{\omega}_j) \tag{10}$$

where the explicit functional dependence of C^{ℓ}_{m} and $D^{\ell_j}_{m_j n_j}$ has been dropped. This is done in the following manner, where (Varshalovich *et al.*, 1988)

$$\begin{aligned} D^{\ell}_{mn}(\boldsymbol{\omega}_i) &= \exp(-im\alpha_i)\, d^{\ell}_{mn}(\beta_i)\, \exp(-in\gamma_i) && (11) \\ d^{\ell}_{mn}(\beta_i) &= \sum_{k=\max(0,m-n)}^{\min(\ell+m,\ell-n)} (-1)^k \\ &\quad \times \frac{[(\ell+m)!\,(\ell-m)!\,(\ell+n)!\,(\ell-n)!]^{1/2}}{k!\,(\ell+m-k)!\,(\ell-n-k)!\,(n-m+k)!} \\ &\quad \times [\cos(\beta_i/2)]^{2\ell+m-n-2k}\, [\sin(\beta_i/2)]^{2k-m+n} && (12) \\ C^{\ell}_{m}(\boldsymbol{\Omega}_{ij}) &= \sqrt{\frac{4\pi}{2\ell+1}}\, Y^{\ell}_{m}(\boldsymbol{\Omega}_{ij}), && (13) \end{aligned}$$

we first couple angular basis functions of molecules i and j

$$A^{\ell_{(ij)}}_{m_{(ij)}} = \sum_{m_i m_j} C^{\ell_{(ij)} m_{(ij)}}_{\ell_i m_i, \ell_j m_j} D^{\ell_i}_{m_i n_i} D^{\ell_j}_{m_j n_j} \tag{14}$$

where $C^{\ell_{(ij)} m_{(ij)}}_{\ell_i m_i, \ell_j m_j}$ is a Clebsch-Gordan coefficient (Varshalovich *et al.*, 1988). Coupling $A^{\ell_{(ij)}}_{m_{(ij)}}$ to the $C^{\ell}_{m}(\boldsymbol{\Omega}_{ij})$ and requiring the overall state to be a scalar W we obtain

$$W^{n_i n_j}_{\ell_i \ell \ell_j} = \sum_{m m_{(ij)}} C^{00}_{\ell m, \ell_{(ij)} m_{(ij)}} A^{\ell_{(ij)}}_{m_{(ij)}} C^{\ell}_{m} \tag{15}$$

By definition the Clebsch-Gordan coefficient $C^{\ell m}_{\ell_1 m_1, \ell_2 m_2}$ is zero unless $m_1 + m_2 = m$ so that $m_2 = -m_1$ if $m = 0$. It is also zero unless $|\ell_1 - \ell_2| \leq \ell \leq \ell_1 + \ell_2$, and since

$m_{1,2} \in \{\ell_{1,2} ... -\ell_{1,2}\}$ this implies that $\ell_1 = \ell_2$. Thus $C^{00}_{\ell m,\ell_{(ij)}m_{(ij)}}$ in Eq. (15) simplifies to $C^{00}_{\ell m,\ell\overline{m}}$. With the identity (Varshalovich *et al.*, 1988)

$$C^{00}_{\ell m,\ell\overline{m}} = \frac{(-1)^{\ell+m}}{\sqrt{2\ell+1}} \tag{16}$$

we have

$$W^{n_i n_j}_{\ell_i \ell \ell_j} = \sum_{m_i m m_j} C^{\ell\overline{m}}_{\ell_i m_i,\ell_j m_j} \frac{(-1)^{\ell+m}}{\sqrt{2\ell+1}} D^{\ell_i}_{m_i n_i} C^{\ell}_{m} D^{\ell_j}_{m_j n_j} \tag{17}$$

Likewise using the identity (Varshalovich *et al.*, 1988)

$$C^{\ell\overline{m}}_{\ell_i m_i,\ell_j m_j} = (-1)^{\ell_i-\ell_j-m}\sqrt{2\ell+1} \begin{pmatrix} \ell_i & \ell_j & \ell \\ m_i & m_j & m \end{pmatrix} \tag{18}$$

gives

$$W^{n_i n_j}_{\ell_i \ell \ell_j} = \sum_{m_i m m_j} (-1)^{\ell_i-\ell_j+\ell} \begin{pmatrix} \ell_i & \ell_j & \ell \\ m_i & m_j & m \end{pmatrix} D^{\ell_i}_{m_i n_i} C^{\ell}_{m} D^{\ell_j}_{m_j n_j}. \tag{19}$$

To remove the phase factor one may exploit the mirror symmetry of the 3jm symbol

$$\begin{pmatrix} \ell_i & \ell_j & \ell \\ m_i & m_j & m \end{pmatrix} = (-1)^{\ell_i+\ell_j+\ell} \begin{pmatrix} \ell_i & \ell & \ell_j \\ m_i & m & m_j \end{pmatrix} \tag{20}$$

leaving

$$W^{n_i n_j}_{\ell_i \ell \ell_j} = \sum_{m_i m m_j} \begin{pmatrix} \ell_i & \ell & \ell_j \\ m_i & m & m_j \end{pmatrix} D^{\ell_i}_{m_i n_i} C^{\ell}_{m} D^{\ell_j}_{m_j n_j}. \tag{21}$$

Equation (21) gives basis functions W from Eq. 10. without any symmetry adaptation. The $W^{n_i n_j}_{\ell_i \ell \ell_j}$ form a complete set of orthogonal IP basis functions over the eight-dimensional space $SO(3) \times S^2 \times SO(3)$. Similar but slightly different constructions are given in the literature(van der Avoird *et al.*, 1980; Briels, 1980; Stone & Tough, 1984; van der Avoird *et al.*, 1994).

While $W^{n_i n_j}_{\ell_i \ell \ell_j}$ are general, it is computationally advantageous to project out the point group symmetry of the molecule and that of the Wyckoff point. Appendix B reviews our use of projection operators which amount to matrix multiplication by a sparse unitary matrix $S^{\ell_i}_{n_i n_\sigma}$ where σ refers to a point group IR and n_σ is a particular

component of the IR. For example, using projection operators for the molecular point group yields symmetry-adapted matrix elements

$$T^{\ell_i}_{m_i n_\sigma}(\boldsymbol{\omega}_i) \;=\; \textstyle\sum_{n_i} D^{\ell_i}_{m_i n_i}(\boldsymbol{\omega}_i) S^{\ell_i}_{n_i n_\sigma} \tag{22}$$

which may be coupled to produce symmetry-adapted basis functions

$$F^{n_\sigma n_\mu}_{\ell_i \ell \ell_j} = \textstyle\sum_{m_i m m_j} \begin{pmatrix} \ell_i & \ell & \ell_j \\ m_i & m & m_j \end{pmatrix} T^{\ell_i}_{m_i n_\sigma}(\boldsymbol{\omega}_i) C^{\ell}_{m}(\boldsymbol{\Omega}_{ij}) T^{\ell_j}_{m_j n_\mu}(\boldsymbol{\omega}_j). \tag{23}$$

With these basis functions the potential is

$$V = \frac{1}{2} \textstyle\sum_{ij} \sum_{\ell_i \ell \ell_j n_\sigma n_\mu} \nu^{n_\sigma n_\mu}_{\ell_i \ell \ell_j}(r_{ij}) F^{n_\sigma n_\mu}_{\ell_i \ell \ell_j}(\boldsymbol{\omega}_i, \boldsymbol{\Omega}_{ij}, \boldsymbol{\omega}_j) \tag{24}$$

where the one half avoids overcounting, $\ell_i, \ell_j \in \mathbb{N}$, $|\ell_i - \ell_j| \leq \ell \leq \ell_i + \ell_j$, and $\nu^{n_\sigma, n_\mu}_{\ell_i, \ell, \ell_j}(r_{ij})$ are coefficients which are a function of the distance r_{ij} between molecular centers. The full set of these in the potential is termed $\boldsymbol{\nu}$. Subscripts σ, μ are a compound index referring to multiple copies of the molecular point group unit IR subduced in the ℓ_i, ℓ_j-th manifold of $SO(3)$ and n_σ, n_μ is its dimension which is always $n_\sigma, n_\mu = 1$. Point group IR subduction frequencies in spherical harmonics are discussed elsewhere (Bradley & Cracknell, 1972). All other point group IR's do not have the full molecular symmetry and so are zero to first order.

Projecting out Wyckoff point symmetry from the functions $T^{\ell_i}_{m_i n_\sigma}$ gives rotator functions

$$U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i) = \textstyle\sum_{m_i n_i} S^{\ell_i *}_{m_i m_\tau} D^{\ell_i}_{m_i n_i}(\boldsymbol{\omega}_i) S^{\ell_i}_{n_i n_\sigma}. \tag{25}$$

Expressing the potential using coupled rotator functions gives

$$V = \frac{1}{2} \textstyle\sum_{ij} \sum_{\ell_i \ell_j m_\tau m_\rho n_\sigma n_\mu} U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i) J^{\ell_i \ell_j}_{m_\tau n_\sigma m_\rho n_\mu}(\boldsymbol{\Omega}_{ij}) U^{\ell_j}_{m_\rho n_\mu}(\boldsymbol{\omega}_j) \tag{26}$$

where

$$\begin{aligned} J^{\ell_i \ell_j}_{m_\tau n_\sigma m_\rho n_\mu}(\boldsymbol{\Omega}_{ij}) \;&=\; \textstyle\sum_{\ell m_i m m_j} \nu^{n_\sigma, n_\mu}_{\ell_i, \ell, \ell_j}(r_{ij}) \begin{pmatrix} \ell_i & \ell & \ell_j \\ m_i & m & m_j \end{pmatrix} \\ &\quad \times S^{\ell_i}_{m_\tau m_i} C^{\ell}_{m} S^{\ell_j}_{m_\rho m_j} \end{aligned} \tag{27}$$

is a dimensionless coupling function. Subscripts τ, ρ are a compound index referring to multiple copies of the Wyckoff point group IR's subduced in the ℓ_i, ℓ_j-th manifold of $SO(3)$ and m_τ, m_ρ goes over the dimensions of each IR.

Appendix B
Projection Operators

Symmetries of the molecule and Wyckoff point of the crystal exist within $D^{\ell_i}_{m_i,n_i}$ simultaneously and can be obtained by applying projection operators (Bradley & Cracknell, 1972)

$$P^{\tau}_{n_\tau n_\tau} = \sqrt{d_\tau/|G|}\,\textstyle\sum_{g\in G} D^{\tau *}_{n_\tau n_\tau}(g)\, g \tag{28}$$

where d_τ is the dimension of the IR τ belonging to the group G of order $|G|$, $\boldsymbol{D}^{\tau}$ is the matrix onto which the IR maps g, and subsequent orthonormalization is occasionally required. We have used a slightly different normalization which decreases the computation in this orthonormalization. Acting upon the elements D^{ℓ}_{mn} gives

$$P^{\tau}_{n_\tau n_\tau} \circ D^{\ell}_{mn} = \textstyle\sum_n D^{\ell}_{mn} S^{\ell}_{nn_\tau} \tag{29}$$

producing a linear combination with the symmetry of τ. The coefficients $S^{\ell}_{nn_\tau}$ form a sparse unitary matrix.

Supplementary Material

Table 1. *Group theoretical symmetry-breaking pathways of experimental lattices from the bcc reference lattice. Classified by space group IR and order parameter direction, each pathway shows the point group IR and minimal manifold of SO(3) in Eq. (1) required to achieve it. The order parameter directions are given in an abbreviated form in the notation of Stokes and Hatch (Stokes & Hatch, 2002).*

Pathway	S. G. IR	OP Dir.	P. G. IR	$\ell^{\mathrm{req'd}}$
229a → 217a	Γ_2^-	P1	A_{2u}	3
229a → 161a	$H_5^- \oplus \Gamma_2^-$	P3 ⊕ P1	$T_{2u} \oplus A_{2u}$	3
	$H_5^- \oplus \Gamma_4^-$	P3 ⊕ P3	$T_{2u} \oplus T_{1u}$	3
	$H_5^- \oplus H_4^+$	P3 ⊕ P3	$T_{2u} \oplus T_{1g}$	4
	$H_4^+ \oplus \Gamma_2^-$	P3 ⊕ P1	$T_{1g} \oplus A_{2u}$	4
	$H_4^+ \oplus \Gamma_4^-$	P3 ⊕ P3	$T_{1g} \oplus T_{1u}$	4
	$H_5^- \oplus H_2^+$	P3 ⊕ P1	$T_{2u} \oplus A_{2g}$	6
	$H_2^+ \oplus \Gamma_4^-$	P1 ⊕ P3	$A_{2g} \oplus T_{1u}$	6
	$H_1^- \oplus \Gamma_4^-$	P1 ⊕ P3	$A_{1u} \oplus T_{1u}$	9
	$H_1^- \oplus H_4^+$	P1 ⊕ P3	$A_{1u} \oplus T_{1g}$	9
229a → 2i (MEZDIE01)	$N_1^- \oplus \Gamma_4^+$	P1 ⊕ S1	$A_{1u}, E_u, T_{2u} \oplus T_{1g}$	4
	$N_1^- \oplus \Gamma_5^+$	P1 ⊕ S1	$A_{1u}, E_u, T_{2u} \oplus T_{2g}$	4
	$N_2^- \oplus \Gamma_4^+$	P1 ⊕ S1	$A_{2u}, E_u, T_{1u} \oplus T_{1g}$	4
	$N_2^- \oplus \Gamma_5^+$	P1 ⊕ S1	$A_{2u}, E_u, T_{1u} \oplus T_{2g}$	4
	$N_3^- \oplus \Gamma_4^+$	P1 ⊕ S1	$T_{1u}, T_{2u} \oplus T_{1g}$	4
	$N_3^- \oplus \Gamma_5^+$	P1 ⊕ S1	$T_{1u}, T_{2u} \oplus T_{2g}$	4
	$N_4^- \oplus \Gamma_4^+$	P1 ⊕ S1	$T_{1u}, T_{2u} \oplus T_{1g}$	4
	$N_4^- \oplus \Gamma_5^+$	P1 ⊕ S1	$T_{1u}, T_{2u} \oplus T_{2g}$	4
229a → 60c,d (YIMWEW)	*			
229a → 2iii	H_4^-	S1	T_{1u}	3
	H5-	S1	T_{2u}	3

Table 2. *Group theoretical symmetry-breaking pathways of experimental lattices for the hcp reference lattice.*

Pathway	S. G. IR	OP Dir.	P. G. IR	$\ell^{\mathrm{req'd}}$
194c → 165d	A_2	P3	A_2', A_1'	3
194c → 147d	$\Gamma_3^+ \oplus \Gamma_2^+$	P1⊕ P1	$A_2'' \oplus A_2'$	3
	$\Gamma_4^+ \oplus \Gamma_2^+$	P1⊕ P1	$A_1'' \oplus A_2'$	4
	$\Gamma_4^+ \oplus \Gamma_3^+$	P1⊕ P1	$A_1'' \oplus A_2''$	4
194c → 176h	K_4	P1	E'	3

Table 3. *Group theoretical symmetry-breaking pathways of experimental lattices for the sc reference lattice.*

Pathway	S. G. IR	OP Dir.	P. G. IR	$\ell^{\mathrm{req'd}}$
221a → 215a	Γ_2^-	P1	A_{2u}	3
221a → 120c	$R_5^- \oplus \Gamma_2^-$	P1⊕ P1	$T_{2u} \oplus A_{2u}$	3
	$R_4^+ \oplus \Gamma_2^-$	P1⊕ P1	$T_{1g} \oplus A_{2u}$	4
	$R_5^- \oplus R_4^+$	P1⊕ P1	$T_{2u} \oplus T_{1g}$	4
	$R_4^+ \oplus \Gamma_3^-$	P1⊕ P1	$T_{1g} \oplus E_u$	7
	$R_5^- \oplus \Gamma_3^-$	P1⊕ P1	$T_{2u} \oplus E_u$	7

Notes:

(1) As these tables are not meant to be exhaustive enumerations but only illustrative of the type of potentials necessary to find a given phase transition, we have truncated listings for 161a, 2i (MEZDIE01), and 14e (MECKUA) which have additional pathways similar to those shown.

(2) Inasmuch as a different method was used in this work to choose an embedding of the daughter lattice in the parent lattice than that used in (McClurg & Keith, 2009), the IRs inducing the phase transition from parent to daughter may be different from (McClurg & Keith, 2009). This, of course, does not affect our numerical results shown in Table 3.

(3) The pathways, space group IRs, order parameter directions, and point group IRs were computed using ISOTROPY. However, transitions belonging to a coupled IR between a high symmetry point and line are currently not a feature of ISOTROPY. These entries have been marked with an asterisk in the table.