Molecular Crystal Global Phase Diagrams III: Sufficient parameter space determination

J. Brandon Keith* and Richard B. McClurg

Department of Chemical Engineering and Materials Science,

University of Minnesota, Minneapolis, Minnesota

Abstract

In previous papers [1, 2] we developed a method for constructing global phase diagrams (GPDs) for molecular crystals and partially applied it to identifying reference lattices pertaining to all single component ordered crystal structures of tetrahedral molecules in the Cambridge Structural Database. Here we expand upon these results by outlining a method to derive a representative potential characteristic of all experimental structures in our data set. This is significant because our prior work [1] 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 15 parameters are sufficient to successfully plot the structures of our test data on molecular crystal GPDs.

*Electronic address: jbrkeith@gmail.com; Also at Department of Physics and Astronomy, Brigham Young University, Provo, Utah

I. INTRODUCTION

part 1

Crystal engineering is the design and synthesis of molecular solid-state structures with desired properties, based on an understanding and exploitation of intermolecular interactions. Two primary interactions holding together supramolecular synthons or secondary building units are hydrogen bonding and coordination complexation, though π-π, halogen-halogen, Au-Au, and ionic interactions have also been exploited. A simple example is the carboxilic acid dimer which is observed in approximately 30% of crystal structures in which it is theoretically possible. These synthons then can be assembled to one-dimensional tapes, two-dimensional sheets and three-dimensional 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 concommitantly a material's electrical, optical, thermal, and solubility properties[3–9]. However, when property and structural needs are not met by the same synthon or group of secondary units, it is less clear how to proceed.

The supramolecular chemistry inherent in molecular crystal synthesis deals with subtle interactions, and consequently control over the processes involved can require great precision. In particular, noncovalent bonds have low energies and often no activation energy for formation. This low bond energy results in difficult-to-control entropic effects, low melting points, and frequent polymorphsim. At low temperatures molecules may lose such entropy and seek structural changes[10]. Thus, thermodynamics is an important tool to design, control, and study molecular crystal chemistry. An example of the important effects of temperature on supramolecular systems is biological organisms, which rapidly cease to function outside a narrow temperature range.

The molecular environment around a supramolecular system is also of prime importance to the particular polymorph to which a set of molecules will crystallize. Many solvents have strong hydrogen bonding, electrostatic, and charge-transfer capabilities, and are therefore able to become involved in complex equilibria with the system, even breaking nucleation complexes completely. For this reason, the choice of solvent can be critical.

Thus, despite ongoing progress in understanding how molecular crystal structures form, there is still a lack of understanding beyond traditional synthons of the crystal forming

tendencies a molecular or collection of molecules will have as a function of temperature and solvent. Were a tool available to globally classify and understand molecular packing tendencies under complex conditions, crystal engineering could expand to its material base and exploit new conditions to form novel strutures.

part 2

Such a tool has been associated with binary fluid mixtures for many years. It began with the work of van Konynenburg and Scott which provided a general and systematic categorization of the different types of fluid-phase behavior in binary mixtures based upon the topology of their critical loci [11]. 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. Specifically, they showed the van der Waals equation of state (EOS), when applied to binary mixtures, can predict qualitatively almost all the known patterns of fluid-phase behavior, and they reveal the mechanisms of transition among the different types. Thus they describe five of the six main experimental classes of fluid behavior differentiated by their temperature-pressure projections.

Although additional fluid-phase behavior patterns have emerged and a revised nomenclature is now available [12], the original van Konynenburg system is still pervasive in the literature [13–15]. Many authors have repeated this type of study in which other EOSs based on simplified models, cubic EOSs, or molecular-based equations provide the GPDs [16–18].

In previous work we have proposed and demonstrated the construction of molecular crystal GPDs which differ from fluid-phase equilibria GPDs in fundamental ways. Molecular crystal GPDs are designed to elucidate the much more diverse and complex phase behavior of a set of arbitrarily-shaped objects arranged in a long-range ordered packing rather than in disordered fluid phases. They also do not use an EOS or mixing rules but rely on the fundamental postulates of equilibrium statistical mechanics, a quantum mechanical-like set of basis functions over rotational space $SO(3)$ and a statistically-derived set of translational packings over translational space T^3. Such basis functions are molecular orbitals interacting 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 potential coefficients indexed by three angular momenta. The mathematical details are explained in Ref. [1], but this leads to a small enough set of potential coefficients to enable a projective mapping to thee-dimensional space. Such a potential is shown to accurately model van der Waals or hydrogen-bonded forces which

are the most common in molecular crystal systems[1]. Crystal structures may thereby be identified in as volumes in three-dimensional projections of this small-dimensional space.

As such molecular crystal GPDs are useful for exploring the complex phase and intermolecular potential possibilities common in modern crystal engineering. A similarity to fluid-phase GPDs is the use of free energy rather than a simple energy minimization as the criterion for stability. The free energy is calculated using a high-temperature reference state (called a reference lattice) in which molecules remain translationally ordered but are rotationally disordered–called a plastically crystalline state. Following a lowering in temperature, new phases arise through spontaneous symmetry breaking, leading to at least a partial rotational ordering of the molecules. This frequently involves a change in space group. A series of additional phase transitions also frequently arise as the temperature is lowered to 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 molecular symmetry. This basic 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 [**?**] and has since been applied to crystals such as solid C_{60} [19, 20].

Like van Konyenburg's 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) [21]. Also, by locating neighboring structures that may be more desirable than a given crystal, GPDs provide feedback for crystal engineering and molecular synthesis to achieve a desired crystal. Further, GPDs can provide rationalization and/or prediction of crystal polymorphs for a given molecule.

part 3

Previously [1] we demonstrated the construction of GPDs for tetrahedral molecules using an fcc reference lattice and outlined the first set of available phases as a function of intermolecular potential parameters. In a companion paper[2] we evolved a small set of tetrahedral molecular reference lattices based on a statistical analysis of the Cambridge Structural Database (CSD) [21]. That work found ...

part 4

This work discusses the issue of a minimal set of intermolecular potential coefficients needed to construct GPDs. A small number will disprove previous hypotheses that many

parameters are needed to describe intermolecular potentials [22] and will enhance the useability of GPDs in materials design. This will allow construction of manageably small-parameter diagrams with 3D-projection visualizations with which users may interact and explore cystal structure relationships and proximity in crystal phase space. Such a tool would be highly valuable to crystal engineers seeking to coerce an optically active molecule through synthetic perturbations to crystallize in a noncentrosymmetric space group, for example. Or it could prove useful in pharmaceutical studies when perturbing solvent, pH, or temperature conditions to enumerate (and control) all possible polymorphs a given drug may produce.

part 5

To explore the requisite parameter number of molecular crystals we begin in Sec. II by recalling the rotational potential used previously in Ref. [1]. We then outline the computational procedure to find potential parameters for each structure. This is done using global optimization techniques among parameter *and* configurational space. In Sec. III we discuss the reverse-engineered potentials and some aspects of our model. We end with a conclusion about our results.

II. 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 [2] we found the majority of the experimental structures in the CSD of tetrahedral molecules can be classified using only four reference lattices. Since there is little data for other reference lattices, we focus on the set {bcc, fcc, hcp, sc}. Each of these has at least four monomer crystal structures in the data set. In this section we seek to determine a *sufficient* parameter space such that experimental structures are each stable in a separate subspace. 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 shown in Fig. 1.

A. Mesoscopic Hamiltonian

Previously [1] we discussed the construction of a nearest neighbor rotational potential for van der Waals molecules. It is a level of abstraction above an atom-atom or site-site potential but retains a firm basis in quantum mechanics [23]. The potential 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 and coupling "matrix" $J^{\ell_i\ell_j}_{m_\tau n_\sigma m_\rho n_\mu}$,

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

The one half avoids overcounting, $\ell_i, \ell_j \in \mathbb{N}$, the natural numbers, and $|\ell_i - \ell_j| \le \ell \le \ell_i + \ell_j$.

$U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ are functions of the orientation of the molecule through its Euler angles $\boldsymbol{\omega}_i$ using the passive convention [24]. They are projected from SO(3) irreducible representations (IR's), 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 [25]. Specifically, the left multiplication by $\boldsymbol{S}^{-1}$ in Eq. (2) gives basis functions transforming like Wyckoff point group IR's and the right multiplication by $\boldsymbol{S}$ gives basis functions of the molecular point group IR's. Subscript τ is a compound index referring to multiple copies of the Wyckoff point group IR subduced in ℓ_i and m_τ goes over 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 [25] and in Sec. II B. The symmetry adaption leaves relatively few basis functions 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 functions taking full advantage of 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,

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

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. [26] The full set of $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell,\ell_j}(r_{ij})$, denoted $\boldsymbol{\nu}$, are the axes of GPD's and the search space for reverse engineering experimental crystal structures.

B. 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 able to produce a symmetry change from the assigned reference lattice to the experimental phase. (3) Develop a library of structural types against which this energy 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.

1. *Figure of Merit*

In our previous work [1, 27] linear response theory was used to seek phase transitions as a model crystal is cooled from a disordered plastic crystalline 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,

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

Also, the free energy is equal to the potential in that limit,

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

Equations (4) and (5) are common approximations used in many crystal structure prediction codes [28] and are further justified by the exclusion of disordered structures in *CSDSymmetry* from which our data set is derived.

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 application. To determine an adequate truncation of the potential we make use of space group irreducible representations (IR's). Symmetry-breaking mechanisms are classified by space group IR's and an order parameter direction such as (a, b) [29]. 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 $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ [1]. Space group IR distortions of the crystal can be decomposed into point group IR distortions of the distinguishable molecules in the crystal [30]. To determine which IR's 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 [29] 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 most accessible in standard tables in the literature [25] and are shown for O_h and D_{3h} in Table I. 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 II. Point group IR's 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. For some pathways, such as that of carbon tetraiodide, 225a $\rightarrow$ 121a in Table II, there is a single space group IR Γ_5^- subducing a single point group IR T_{2u}. Table I 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. 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 I), 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 IR's. Three different coupled space group IR's decompose into three different point group IR's. One of these pathways uses IR's 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 IR's

appear by at least the third or fourth manifold in nearly all cases.

In real phase transitions, IR's on the first non-trivial manifold typically induce symmetry breaking. Second manifold induced symmetry breaking is uncommon and higher manifold induced phase transitions are not generally observed [31]. 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 [32]. 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.

3. Candidate Lattice Library

We seek a set of potential parameters $\boldsymbol{\nu}$ such that the energy is lower than other structures, whether observed or not. Therefore, a comparative list of structural types should consist of both the experimental structures and a large library of alternate structures. Each alternative structure serves to constrain the representative potential parameters $\boldsymbol{\nu}$ identified with each experimental structure. In our previous work [1], high symmetry point IR distortions [29, 33] 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 crystal structure prediction which generates thousands of multi-domain duplicate structures. Therefore they are convenient for constructing a library of candidate lattices. They are discussed further in Sec. III. 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 [34]. The library becomes inconveniently large if $Z' > 1$ is permitted. ISs whose *primitive* unit cell contains more than eight molecules are also discarded [35].

To minimize the energy of these candidate lattices, $U^{\ell_i}_{m_\tau n_\sigma}(\boldsymbol{\omega}_i)$ 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 for all other molecules in the Wyckoff orbit. This reduces the number of Euler angles needed to just one set for each candidate lattice. 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 basis functions 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 results in a generalization of and supergroup to high symmetry isotropy subgroups.

This procedure gives 75 candidate lattices for bcc, 67 for fcc, 19 for hcp, and 105 for sc. These augment the list of observed structures in our library. The unit cells of the candidate lattices are given in the supplementary material. The unit vectors are given in terms of their reference lattice vectors. Space group operations (of the reference lattice) that relate molecules on distinct lattice points are also given. 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.

4. Experimental Phase Energies

The potential V in Eq. (1) is a function of Euler angles ω, potential parameters ν_{ℓ_i,ℓ,ℓ_j}, and intermolecular angle Ω (measured with respect to the z-axis in the lab reference frame of molecule i). The intermolecular angles are determined by the reference lattice but the Euler angles are dependent upon the orientation of the molecules on the reference lattice. This is complicated by the fact that the experimental atomic positions in each molecule are

slightly distorted due to small strains resulting from embedding the experimental lattice in its reference lattice. We calculate the Euler angles of experimentally observed structures by rotating a rigid tetrahedral molecule in the reference lattice frame to minimize its atomic distances compared to the *fractional* coordinates of the experimental molecule. The reference lattice molecular positions are also shifted to coincide with the centers of mass of the experimentally determined fractional cell atoms. Thus the Euler angles obtained are a "best fit" to the molecular rotations were the experimental structures exactly proportional to their reference-lattice-derived ideal lattices. These Euler angles are presented in the supplementary material for our reference lattice set. The reference configuration 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. Typically only one atom lying along each of the three-fold axes was used when determining molecular orientations. Euler angles are not unique due to the 2π modulus in the definition of $\{\alpha, \beta, \gamma\}$ and the point group symmetry of the tetrahedral molecule. Once the Euler angles are determined, the rotator functions $U^{\ell_i}_{m_\tau n_\sigma}$ may be calculated directly using Eq. (2). These are listed in the supplementary material. $U^{\ell_i}_{m_\tau n_\sigma}$ is fully symmetric with respect to 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 $\boldsymbol{\nu}$ such as

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

for tetrakis(trimethystannyl)silane [CSD structure MEZDIE01], which crystallizes in space group 2.

5. Minimum Energy Gap

To find a representative vector $\boldsymbol{\nu}$ for which each experimental phase is the lowest energy structure (*i.e.* a phase with an energy such as Eq. (8)), we propose a minimum energy difference algorithm. First we project the $\boldsymbol{\nu}$ vector onto the unit hypersphere since only

relative $\boldsymbol{\nu}$ magnitudes affect the phase as $T \to 0$. Then taking the library of structures discussed previously in this section, we 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 we seek the vector ν_{RP} that minimizes

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

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 exhaustive library, ΔE is bounded below by 0 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 RP's converge.

Minimization methods using genetic algorithms (differential evolution) and simulated annealing [36] 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 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.

We anticipate that the lower manifold basis functions should be more important in the potential because the kinetic energy of molecular orbitals pertaining to these angular momenta is lower. However, the above scheme does not discriminate between ν_{ℓ_i,ℓ,ℓ_j} corresponding to

lower or higher manifolds. To favor lower manifolds the energy difference is minimized for $\ell_{\max} = 3$ 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.

III. DISCUSSION AND CONCLUSION

A. Representative Potentials

Calculating potential parameters according to the procedure outlined in Sec. II gives the results in Table III where the potential coefficients are given for experimental structures in the four most prevalent reference lattices. Calculations for other reference lattices such as Aa or A15 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 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 8 of the 35 structures, $\Delta E < 0$. This indicates the target structure is not in the library. Since $\Delta E = 0$ 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 5 of the 35 structures, $0 < \Delta E < 0.01$ and therefore a representative potential for these structures has likely 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 a slight mismatch in the Euler angles $\boldsymbol{\omega}_{\text{exp}}$ and $\boldsymbol{\omega}_{\text{min}}$ since the minimizer is free to alter the angle to attain the lowest possible energy. Thus $\boldsymbol{\nu}_{RP}$ is in the correct region but not necessarily centrally located and the target structure is in the library.

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 or three-body interactions, or missing quantum effects. Another reason may be that the reported experimental phase is in fact metastable in the limit $T \rightarrow 0$. Ostwald's step rule supports this, which is that during crystal nucleation and growth a metastable phase frequently crystallizes first which is closest to the liquid and which has the lowest free energy barrier. The crystallite then undergoes a series of metastable modifications until it reaches the thermodynamically stable structure [37]. The experimentally observed structure may have formed during an early stage of nucleation as one of these intermediate structures.

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 [22] that many manifolds are needed to accurately describe tetrahedral molecules such as methane and adamantane. However this supports current work [38] showing that fewer manifolds are necessary if the coefficients of the basis functions are evaluated to reflect primarily *low energy configurations* rather than the common evaluation method that equally weights all configurations. Current work is a proof by demonstration. Although visualization of high dimensional spaces is difficult, software such as GGOBI has made this possible in an interactive way. [39]. Any number of axes may be displayed and spatial relationships between the RP's analyzed.

Interpreting the potential coefficients in Table III is analogous to describing multipole interactions. These multipoles include contributions from all interaction modes and are not limited to electrostatic interactions. 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 interacting with the zeroth pole and 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 [23]. For instance, although dispersion and induction forces can be found in all components of $\boldsymbol{\nu}$, electrostatic forces contribute only to $\nu^{n_\sigma,n_\mu}_{\ell_i,\ell_i+\ell_j,\ell_j}$ [40] such as $\nu_{3,6,3}$, $\nu_{3,7,4}$, and $\nu_{4,8,4}$.

One drawback of using the algorithm in Sec. II B is that, by finding the maximum energy difference between a phase and all others, the RP's 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 RP's. However, different RP's are widely spread by our algorithm. Therefore RP's 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.

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 [2] 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 RP's 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.

B. 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 crystal structures is needed to localize the RP in $\boldsymbol{\nu}$ space. In our previous work [1] we considered just the high-symmetry point isotropy subgroups and in this work have followed suit since these are the most common [33]. Recall that space group IR's are indexed by reciprocal space vectors [41, 42]. 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 or non-clustered parent lattice unit cells such as 15f,f,f,f or 152b and their $\boldsymbol{k}$ vectors therefore are from high-symmetry plane and line IR's. Although these cases are less common, such isotropy subgroups could be included in the candidate lattices in Sec. II B 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. II B 3 we chose to consider only isotropy subgroups with one occupied Wyckoff point. This is a commonly used simplification. [28] 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, again at the expense of a much larger library.

Although tetrahedral molecules are used in the current example to reverse engineer the IP, 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 $\{\nu^{1,n_\mu}_{1,1,2}, \nu^{1,n_\mu}_{1,3,2}\}$ 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 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 IV 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 reverse engineer our dataset of experimental structures. 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 less than 15 coefficients are necessary.

C. Conclusions

We have shown that a molecular crystal global phase diagram (GPD) with a small number of reference lattices derived in [2] 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 Konynenburg 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.

Acknowledgments

This work received financial support from the American Chemical Society - Petroleum Research Fund (PRF #41774-AC10). Computational resources maintained by the University of Minnesota Supercomputer Institute were used for portions of this research.

[1] J. A. Mettes, J. B. Keith, and R. B. McClurg, Acta Cryst. A **60**, 621 (2004).

[2] R. B. McClurg and J. B. Keith (2008), submitted.

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

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

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

[6] J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, et al., Acta Crystallogr. B **56**, 697 (2000).

[7] K. T. Holman, A. M. Pivovar, J. A. Swift, and M. D. Ward, Acc. Chem. Res. **34**, 107 (2001).

[8] B. Moulton and M. J. Zaworotko, Chem. Rev. **101**, 1629 (2001).

[9] W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij, et al., Acta Cryst. B **58**, 647 (2002).

[10] M. A. Neumann, W. Press, C. Noldeke, B. Asmussen, M. Prager, and R. M. Ibberson, J. Chem. Phys. **119**, 1586 (2003).

[11] P. H. van Konynenburg and R. L. Scott, Philos. Trans. R. Soc. London, Ser. A **298**, 495 (1980).

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

[13] S. Aparicio-Martinez and K. R. Hall, Ind. Eng. Chem. Res. (????).

[14] S. Aparicio-Martinez and K. R. Hall, Ind. Eng. Chem. Res. (????).

[15] M. Cismondi and M. L. Michelsen, J. Supercrit. Fluid. (????).

[16] P. I., W. J., S. H., Y. L. V., and K. T., Fluid Phase Equilib. (????).

[17] I. Polishuk, J. Wisniak, and H. Segura, Phys. Chem. Chem. Phys. **4**, 879 (2002).

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

[19] K. H. Michel, J. R. D. Copley, and D. A. Neumann, Phys. Rev. Lett. **68**, 2929 (1992).

[20] D. Lamoen and K. H. Michel, Phase Transitions **67**, 789 (1999).

[21] F. H. Allen and W. D. S. Motherwell, Acta Cryst. B **58**, 407 (2002).

[22] W. J. Briels, J. Chem. Phys. **73**, 1850 (1980).

[23] A. van der Avoird, P. E. S. Wormer, and R. Moszynski, Chem. Rev. **94**, 1931 (1994).

[24] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Teaneck, NJ, 1988).

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

[26] A. van der Avoird, P. E. S. Wormer, F. Mulder, and R. M. Berns, Topics in Current Chemistry **93**, 1 (1980).

[27] J. B. Keith, J. A. Mettes, and R. B. McClurg, Crystal Growth and Design **4**, 1009 (2004).

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

[29] H. T. Stokes and D. M. Hatch, *Isotropy*, stokes.byu.edu/isotropy.html (2002).

[30] H. T. Stokes, D. M. Hatch, and J. D. Wells, Phys. Rev. B. **43**, 11010 (1991).

[31] R. M. Lynden-Bell and K. H. Michel, Rev. Mod. Phys. **66**, 721 (1994).

[32] M. Hloucha, J. Lodge, A. Lenhoff, and S. Sandler, Cryst. **64**, 060104(R) (2001).

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

[34] N. Padmaja, S. Ramakumar, and M. A. Viswamitra, Acta Cryst. A **46**, 725 (1990).

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

[36] S. Kirkpatrick, C. Gelatt Jr, and M. Vecchi, Science **220**, 671 (1983).

[37] H. Schmalzried, Zeit. fur Physik. Chem. **217**, 1281 (2003).

[38] M. N. Missaghi, J. A. Mettes, and R. B. McClurg (2005), in Preparation.

[39] D. T. Lang, *Ggobi*, www.ggobi.org (2003).

[40] A. J. Stone and R. J. A. Tough, Chem. Phys. Lett. **110**, 123 (1984).

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

[42] J. Zak, H. Casher, H. Gluck, and Y. Gur, *The Irreducible Representations of Space Groups*

(Benjamin, New York, 1969).

[43] H. T. Stokes and D. M. Hatch, Phys. Rev. B **65**, 144114 (2002).

TABLE I: Presence of O_h and D_{3h} point group IR's 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 IR's	D_{3h} IR's
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 II: 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 [43]. 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	$T_{1g} \oplus T_{1u}, T_{2u}$	4
		P1 ⊕ S7		
	$\Gamma_5^+ \oplus L_3^-$	C2 ⊕ C8	$T_{2g} \oplus E_u, T_{1u}, T_{2u}$	4
	$\Gamma_5^+ \oplus X_5^-$	P1 ⊕ C11	$T_{2g} \oplus T_{1u}, T_{2u}$	4
		P1 ⊕ S7		
	$L_2^+ \oplus L_1^-$	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	$A_{2u}, E_u \oplus T_{1u}, T_{2u}$	4
		P1 ⊕ S7		
	$X_3^- \oplus X_5^-$	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 III: 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 IV: 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.

TABLE V: 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 [43]. 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	$\ell^{\mathrm{req'd}}$
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 VI: Group theoretical symmetry-breaking pathways of experimental lattices for the hcp reference lattice.

Pathway	S. G. IR	OP Dir.	P. G. IR	$\ell^{\mathrm{req}'\mathrm{d}}$
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 VII: 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

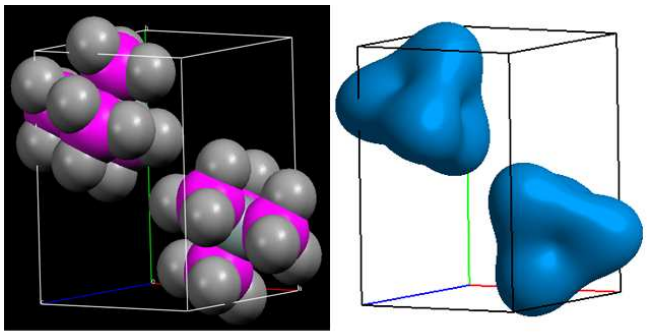

FIG. 1: Comparison of an experimental structure, tetrakis(trimethystannyl)silane [CSD structure MEZDIE01], which crystallizes in space group 2 at Wyckoff point i with arbitrarily-shaped tetrahedral figures whose orientation is determined by orientational energy minimization with molecular center of mass on an ideal bcc reference lattice.