Faraday Discussions C8FD90031K

We have presented the graphical abstract image and text for your article below. This briefly summarises your work, and will be presented with your article online. It will not appear in the print edition.

DISCUSSIONS

1

Crystal structure evaluation: calculating relative stabilities and other criteria: general discussion

Matthew Addicoat, Claire Adjiman, Mihails Arhangelskis, Gregory Beran, David Bowskill, Gerit Brandenburg, Doris Braun, Virginia Burger, Jason Cole, Aurora Cruz-Cabeza, Graeme Day, Volker Deringer, Rui Guo, Alan Hare, Julian Helfferich, Johannes Hoja, Luca Iuzzolino, Samuel Jobbins, Noa Marom, David McKay, John Mitchell, Sharmarke Mohamed, Marcus Neumann, Sten Nilsson Lill, Jonas Nyman, Artem R. Oganov, Pablo Piaggi, Sarah Price, Susan Reutzel-Edens, Ivo Rietveld, Michael Ruggiero, Matthew Ryder, German Sastre, Christian Schön, Christopher Taylor, Alexandre Tkatchenko, Seiji Tsuzuki, Joost van den Ende, Scott Woodley, Grahame Woollam and Qiang Zhu

Please check this proof carefully. **Our staff will not read it in detail after you have returned it.**

Proof corrections must be returned as a single set of corrections, approved by all co-authors. No further corrections can be made after you have submitted your proof corrections as we will publish your article online as soon as possible after they are received.

Please ensure that:

- The spelling and format of all author names and affiliations are checked carefully. Names will be indexed and cited as shown on the proof, so these must be correct.
- Any funding bodies have been acknowledged appropriately.
- All of the editor's queries are answered.
- Any necessary attachments, such as updated images or ESI files, are provided.

Translation errors between word-processor files and typesetting systems can occur so the whole proof needs to be read. Please pay particular attention to: tables; equations; numerical data; figures and graphics; and references.

Please send your corrections preferably as a copy of the proof PDF with electronic notes attached or alternatively as a list of corrections – do not change the text within the PDF file or send a revised manuscript. Corrections at this stage should be minor and not involve extensive changes.

Please return your **final** corrections, where possible within **48 hours** of receipt, by e-mail to: faraday@rsc.org. If you require more time, please notify us by email.

Researcher information

If any authors have ORCID or ResearcherID details that are not listed below, please provide these with your proof corrections.

Please check that the ORCID and ResearcherID details listed below have been assigned to the correct author. Authors should have their own unique ORCID iD and should not use another researcher's, as errors will delay publication.

Please also update your account on our online manuscript submission system to add your ORCID details, which will then be automatically included in all future submissions. See here for step-by-step instructions and more information on author identifiers.

First (given) name(s)	**Last (family) name(s)**	**ResearcherID**	**ORCID**
Matthew	Addicoat		
Claire	Adjiman		
Mihails	Arhangelskis		
Gregory	Beran		
David	Bowskill		
Gerit	Brandenburg		
Doris	Braun		
Virginia	Burger		
Jason	Cole		
Aurora	Cruz-Cabeza		
Graeme	Day		
Volker	Deringer		
Rui	Guo		
Alan	Hare		
Julian	Helfferich		
Johannes	Hoja		
Luca	Iuzzolino		
Samuel	Jobbins		
Noa	Marom		
David	McKay		
John	Mitchell		
Sharmarke	Mohamed		
Marcus	Neumann		
Sten	Nilsson Lill		
Jonas	Nyman		
Artem R.	Oganov		
Pablo	Piaggi		

Sarah	Price		
Susan	Reutzel-Edens		
Ivo	Rietveld		
Michael	Ruggiero		
Matthew	Ryder		
German	Sastre		
Christian	Schön		
Christopher	Taylor		
Alexandre	Tkatchenko		
Seiji	Tsuzuki		
Joost	van den Ende		
Scott	Woodley		
Grahame	Woollam		
Qiang	Zhu		

Queries for the attention of the authors

Journal: **Faraday Discussions**

Paper: **C8FD90031K**

Title: **Crystal structure evaluation: calculating relative stabilities and other criteria: general discussion**

For your information: You can cite this article before you receive notification of the page numbers by using the following format: (authors), Faraday Discuss., (year), DOI: 10.1039/C8FD90031K.

Editor's queries are marked on your proof like this ❶, ❷, etc. and for your convenience line numbers are indicated like this 5, 10, 15,

Please ensure that all queries are answered when returning your proof corrections so that publication of your article is not delayed.

Query Reference	Query	Remarks
1	Please confirm that the spelling and format of your name is correct. Names will be indexed and cited as shown on the proof, so these must be correct. No late corrections can be made. FOR YOUR INFORMATION: You can cite this paper before the page numbers are assigned with: (authors), Faraday Discuss., DOI: 10.1039/c8fd90031k.	

Faraday Discussions

Cite this: DOI: 10.1039/C8FD90031K

211

DISCUSSIONS

Crystal structure evaluation: calculating relative stabilities and other criteria: general discussion

Matthew Addicoat, Claire Adjiman, Mihails Arhangelskis, Gregory Beran, David Bowskill, Gerit Brandenburg, Doris Braun, Virginia Burger, Jason Cole, Aurora Cruz-Cabeza, Graeme Day, Volker Deringer, Rui Guo, Alan Hare, Julian Helfferich, Johannes Hoja, Luca Iuzzolino, Samuel Jobbins, Noa Marom, David McKay, John Mitchell, Sharmarke Mohamed, Marcus Neumann, Sten Nilsson Lill, Jonas Nyman, Artem R. Oganov, Pablo Piaggi, Sarah Price, Susan Reutzel-Edens, Ivo Rietveld, Michael Ruggiero, Matthew Ryder, German Sastre, Christian Schön, Christopher Taylor, Alexandre Tkatchenko, Seiji Tsuzuki, Joost van den Ende, Scott Woodley, Grahame Woollam and Qiang Zhu

DOI: 10.1039/C8FD90031K

(200:[200]200) **Christian Schön** opened the discussion of the paper by Gregory J. O. Beran: Concerning the multi-tier approach to improving the quality of the computations, where you did a careful analysis of different types of approximations, it is still very disconcerting that there appears to be no inherently systematic approach to gain an estimate of the errors involved in the calculations – often it seems that one does not really know where the error is hiding (*e.g.* fortuitous error cancellations). This contrasts with the analytical convergence criteria and error bounds one gets when using *e.g.* perturbation theory, and is quite bothersome (especially since many people do not even try to provide error estimates in the discussion sections of their papers). Do you see any progress in this area?

Gregory Beran replied: Quantification of uncertainty is a major challenge in electronic structure theory (and many other areas of chemical simulation). Our preferred, albeit imperfect, approach is to investigate the sensitivity of the prediction to the modeling choices. For example, one can explore how the structures, thermochemical properties, relative polymorph stabilities, or other properties of interest converge with respect to the basis set and level of theory. A few earlier studies from our group exemplify this. For aspirin forms I and II, the relative energy rankings proved quite insensitive to the electronic structure method and basis set, leading to relatively high confidence in the relative energetics.[1] The close similarities in crystal packing between the two forms allow for substantial error cancellation in the relative energies. In contrast, the energy rankings for the five polymorphs of oxalyl dihydrazide are very sensitive to the electronic structure treatment, making the quantitative stability ranking less certain.[2] In that case, the basis set superposition error (BSSE) associated with the intramolecular hydrogen bonding in the alpha polymorph is harder to correct than the BSSE for the purely intermolecular hydrogen bonding found in the other polymorphs. Only in the limit of the large basis sets does one achieve a balanced description of both types of interactions.

Perhaps the best example of sensitivity analysis/uncertainty quantification from our work occurs in our recent study, which predicted the methanol phase diagram.[3] In that case, by examining how the predicted phase boundaries change upon altering the relative free energies between two forms, we showed that our calculations reproduced the experimental phase diagram to within ∼0.5 kJ mol^{-1}. Furthermore, the sensitivity analysis was used to investigate which features of the calculations were most important and where one should focus the computational effort in order to obtain accurate predictions. The Electronic Supplementary Information for ref. 3 goes into considerable detail on these issues.

1 S. Wen and G. J. O. Beran, *Cryst. Growth Des.*, 2012, **12**, 2169–2172.
2 S. Wen and G. J. O. Beran, *J. Chem. Theory Comput.*, 2012, **8**, 2698–2705.
3 C. Cervinka and G. J. O. Beran, *Chem. Sci.*, 2018, **9**, 4622–4629.

(201:[201]201) **Volker Deringer** said: You have nicely discussed the different “tiers” of methodology, and how the acetaminophen example was only amenable to the computationally more affordable tiers. How do you expect this to change over time, when more and cheaper CPU time will become available? Do you expect “Tier 1” methods to be routinely in reach within a time frame of a couple of years – and what impact will that have on the field?

Gregory Beran responded: Future advances in computer power will certainly help, but as the timings indicate, the computational cost of a Tier 1 calculation is roughly 200 times that of a Tier 3 or Tier 4 one, and that ratio will continue growing with increasing molecular size. For many applications, the difference between Tier 1 and 3 results may not be worth the difference in computational cost. In the longer term, we think it will likely prove more beneficial to move beyond the quasi-harmonic approximation and find better ways to handle the dynamics and anharmonicity in the crystal, perhaps using machine learning or other approaches to develop accurate potentials that can be sampled more cheaply. High-quality energies/forces from fragment methods could potentially provide inputs for developing such potentials.

(202:[202]202) **Aurora Cruz-Cabeza** asked: Could you please give an estimate as to how your Tier 1–4 calculations compare in terms of computational expense?

Gregory Beran answered: Several factors play a role in the overall computational timings. Generally speaking, MP2 scales $O(N^5)$ with system size N, while large-scale DFT calculations typically scale $O(N^3)$ or better. However, things become more nuanced for the calculations performed here. The fragment-based hybrid many-body interaction (HMBI) energy approach decomposes the full crystal calculation to a series of quantum mechanical (QM) monomer and short-range dimer calculations, while longer-range dimers and many-body interactions are approximated at a lower level of theory (*e.g.* with the AMOEBA polarizable force field or in some cases here, periodic Hartree–Fock, HF). Evaluation of the QM dimers forms the computational bottleneck. At long range, the model switches to a force field or Hartree–Fock description of the dimer intermolecular interactions.

Furthermore, the HMBI approach exploits space group symmetry and requires evaluation only of the symmetrically unique dimers. Thus, the computational cost grows $O(N^5)$ with the size of the monomer, but it grows only linearly with the number of molecules in the asymmetric unit. The fragment-based calculations employ atom-centered Gaussian-orbital basis sets, and the convergence of the MP2 correlation energy with basis set size is slow—cubic, rather than the exponential convergence DFT would have in the same basis sets.

For fragment-based phonon calculations like those with HMBI, it turns out that the same QM dimer calculations are needed for either a Γ-point-only calculation or a supercell lattice dynamics calculation.[1] The extra work associated with the supercell occurs purely at the low level of theory (*e.g.* polarizable force field). In addition, the finite displacements are done on a smaller scale (in individual monomers or dimers, which often have fewer atoms than the whole unit cell), which further reduces the cost. So while MP2 frequencies are considerably more expensive than DFT ones, it might prove more efficient to use the fragment approach to capture phonon dispersion for sufficiently large cells.

For plane-wave DFT, the basis set grows with the size of the unit cell. This becomes especially problematic for large supercell phonon calculations. On the other hand, one can typically perform less *k*-point sampling in larger cells. DFT codes typically also exploit the symmetry of the cell to reduce the number of *k*-points.

Overall, one expects the cost of fragment-based MP2 calculations to grow more rapidly than plane-wave DFT as the molecules get larger, while the plane-wave DFT calculation cost grows more steeply with unit cell size. Acetic acid, imidazole and acetaminophen each have four molecules in the unit cell and one in the asymmetric unit. Carbon dioxide and ice are not discussed here, since their small molecular size and the extraordinarily high symmetry in carbon dioxide make them anomalously fast compared to larger organic molecules of interest to the community.

For acetic acid or imidazole, a single-point energy MP2/aug-cc-pVTZ + AMOEBA calculation is more expensive than the plane-wave B86bPBE-XDM (80 Ry plane-wave cut-off) one by a factor of $\sim$1.5.[2] The relative cost increases by a factor of $\sim$6–8× for a force or frequency calculation. The DFT basis set used in those calculations was probably larger than necessary, and for acetaminophen we reduced the plane-wave cut-off to 60 Ry. We also reduced the *k*-point mesh from 5×5×5 to 1×3×3. As a result, MP2 is $\sim$49×, $\sim$242×, and $\sim$153× more expensive than DFT for an energy, force, and frequency calculation. For comparison, if we kept the 80 Ry plane-wave cut-off, the ratios would be $\sim$3×, $\sim$16×, and $\sim$0.2×. With the very large DFT basis set, the frequency calculation is actually cheaper with fragment MP2/augcc-pVTZ, though again, that DFT basis set would probably be unnecessarily large.

Based on these test calculations, we estimated overall computational costs in CPU hours and relative costs for the Tier 1–4 calculations. This includes mapping out the energy–volume curves at 10 different geometries (assuming 100 optimization cycles per geometry optimization) and computing phonon frequencies at three volumes (for the reference frequencies and two more to compute the Grüneisen parameters). The results are summarized in Table 1. From this table, one sees that the Tier 1 calculations cost roughly 1–2 orders of magnitude more than the Tier 4 ones. On the other hand, for Tiers 3 and 4, the computational cost is dominated by the DFT geometry optimizations and phonon calculations. Single-point energy refinement at the MP2/aug-cc-pVTZ level in Tier 3 increases the computational cost only marginally—no more than $\sim$20% here. In contrast, switching to Tier 2 starts to increase the computational cost much more dramatically, especially for acetaminophen.

Even the single-point refinement calculations in the paper, which extrapolate to the complete basis set limit, correct the known deficiencies in MP2 with MP2C, and replace AMOEBA with periodic HF, do not increase the cost of Tier 3 too much. For the smaller two crystals, this refinement incurs only 10–20% higher cost compared to the pure DFT Tier 4 calculations. For acetaminophen, it roughly doubles the cost compared to the Tier 4 calculation. When combined with the quality of the predictions discussed in the paper, Tier 3 provides a valuable compromise between accuracy and computational cost.

1 S. Hirata, *J. Chem. Phys.*, 2008, **129**, 204104.
2 All timings were performed on a 16-core node with dual 2.4 GHz Intel Xeon E5-2630 v3 processors, 64 GB of RAM, and several terabytes of local disk storage.

(807:[203]203) **David McKay** communicated: The four tiers of computational methodology are set out in decreasing order of computational cost, with Tier 1 having HMBI energies, geometries and frequencies, then DFT coming in from right to left, *i.e.* frequencies, then geometries, then energies. Would the results of methods including DFT be improved (closer to the all-HMBI Tier 1) if, rather than considering the cost of the methods, you considered the competency of the methods at each stage, perhaps giving a Tier 2′ of HMBI, DFT and HMBI for energies, geometries and frequencies, respectively? Perhaps also a Tier 3′ of DFT, DFT, HMBI.

Table 1 Approximate computational costs for Tiers 1–4 using either HMBI (MP2/aug-cc-pVTZ + AMOEBA) or B86bPBE-XDM. The estimates assume 10 geometry optimization/energy calculations to map out the energy volume curves, 3 phonon calculations to obtain the reference phonons and Grüneisen parameters, and 100 geometry optimization cycles per geometry optimization

	CPU hours				Relative time			
Crystal	Tier 1	Tier 2	Tier 3	Tier 4	Tier 1	Tier 2	Tier 3	Tier 4
Acetic acid[a]	37 600	30 700	5330	5270	7.1	5.8	1.0	1
Imidazole[a]	92 800	74 100	13 100	13 000	7.1	5.7	1.0	1
Acetaminophen[b]	1 290 000	745 000	8000	6600	195	113	1.2	1

[a] DFT calculations employed an 80 Ry plane-wave cut-off and a 5×5×5 *k*-point mesh. [b] DFT calculations employed a 60 Ry plane-wave cut-off and a 1×3×3 *k*-point mesh.

Gregory Beran communicated in reply: Given the reality of needing to reduce computational cost, as evidenced by the timing data presented in response to the previous question, we did not investigate approaches of the sort you mention. What follows is therefore speculation: in the molecular quantum chemistry literature, there are usually not substantial differences between the quality of harmonic vibrational frequencies for intramolecular modes predicted with DFT *versus* MP2, assuming sufficiently large basis sets. The empirical factors people use to scale the frequencies to improve agreement with experiment are generally similar in magnitude (typically around 0.9–1.0, depending on the specific model chemistry). It is less obvious how well that insight translates to the soft lattice phonons, which contribute significantly to thermal expansion and thermal contributions to the enthalpy and entropy. On the other hand, the soft lattice modes are largely governed by the crystal packing and cell volume. So as long as the model gets those reasonably correctly, it should perform well. Our thermal expansion data and plot of the phonon frequencies for ice demonstrate generally similar performance for either HMBI or DFT on the low-frequency modes, so one might think there could be some freedom to mix and match the different models along the lines you suggest. However, in the end, the practical matters of computational effort drive us toward Tiers 3 and 4 discussed in the paper.

(205:[204]204) **Mihails Arhangelskis** enquired: My question concerns modelling unit cell expansion with quasi-harmonic calculations. How was anisotropic unit cell expansion accounted for? In the case of low symmetry (monoclinic and, especially, triclinic) unit cells I would imagine that a large number of structure distortions would have to be considered in order to map the temperature expansion in each crystallographic direction. Presumably, one would have to study expansion about each principal axis separately, while also considering the variations in unit cell angles. I would be curious to know whether such calculations are practical for low symmetry cells, or whether there this can be done in a more efficient way.

Gregory Beran responded: For the quasi-harmonic calculation, we perform the expansion/compression along a one-dimensional semi-anisotropic coordinate. Specifically, the energy *versus* volume curve is generated by optimizing the structure under various positive and negative hydrostatic pressures. This allows the volume to expand or contract by deforming the unit cell in whatever manner minimizes the energy change. A closely related approach would optimize the cell subject to a volume constraint for various different volumes. However, algorithms to perform such constant volume optimizations are not currently implemented in the software packages used in our work. Either way, the volume dependence of the phonons can subsequently be determined along the same one-dimensional coordinate. The semi-anisotropic approach used in our work is an approximation relative to a fully anisotropic expansion model which would allow all independent degrees of freedom to vary independently, but it should provide a more realistic description of how the unit cell varies than a simple isotropic scaling of the lattice parameters. For detailed analysis of the impact of including different degrees of anisotropy in quasi-harmonic calculations, see the recent work by Abraham and Shirts.[1] In the examples they looked at, the difference between an isotropic and a fully anisotropic model was up to ~1 kJ mol^{-1}. The differences involving quasi-anisotropic approaches like ours are considerably smaller. Regarding low-symmetry cells: the computational bottleneck for the hybrid many-body interaction fragment model is the number of dimer calculations that need to be evaluated. For careful calculations of the sort employed here, we typically employ conservative cut-offs that evaluate dimer interactions quantum mechanically to 9–10 Å, before smoothly switching to an MM treatment for longer-range interactions. See ref. 2 for more details. Exploiting space group symmetry typically reduces the number of dimers by a factor corresponding to the ratio between the total number of molecules in the unit cell (Z) and the number of molecules in the asymmetric unit (Z'). For example, in the orthorhombic polymorph of acetic acid ($Pna2_1$ space group, $Z = 4$, $Z' = 1$) with the 9–10 Å cut-off mentioned above, a typical single-point energy calculation on crystalline acetic acid might require over 212 dimers without space group symmetry, or 53 dimers with symmetry (savings factor = 4). See ref. 3 for more details. Four-fold symmetry is common to many of the most frequently occurring space groups, and such symmetry savings factors are common. Therefore, calculations on a unit cell with $P1$ symmetry will require several-fold more computational effort than calculations on a comparable cell with four-fold symmetry.

One nice feature of the fragment approach, however, is that the computational effort for the expensive quantum mechanical calculations grows linearly with the number of molecules in the unit cell. The quantum mechanical dimer interactions occur within a finite radius (*e.g.* the aforementioned 9–10 Å cut-off), regardless of whether those molecules occupy the central unit cell or a periodic image cell. In contrast, the computational cost of the periodic density functional grows cubically with the basis set size, and increasing the unit cell dimensions increases the number of plane-waves which fit in the unit cell for a given plane-wave cut-off. This behavior could make a fragment approach particularly attractive for modeling a defect in a larger supercell, for example.

1 N. S. Abraham and M. R. Shirts, 2017, arXiv:1712.00936 [cond-mat.mtrl-sci].
2 S. Wen, K. D. Nanda, Y. Huang and G. J. O. Beran, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7578–7590.
3 Y. N. Heit and G. J. O. Beran, *J. Comput. Chem.*, 2014, **35**, 2205–2214.

(203:[205]205) **Marcus Neumann** asked: In the calculation of the Grüneisen parameter you apply hydrostatic pressure in order to estimate the unit cell changes under thermal pressure. How well is this approximation validated? What is the expected free energy error of the approximation?

Gregory Beran answered: We have not investigated this issue in detail, though we provide some discussion regarding the isotropic *versus* the quasi-anisotropic hydrostatic approach in our answer to Mihails Arhangelskis' previous question. Abraham and Shirts recently investigated various ways of performing the quasi-harmonic approximation that include varying degrees of anisotropy.[1] They found that fully anisotropic expansion allowed the crystal to adopt structures up to 1 kJ mol^{-1} lower in energy than isotropic expansion. Free energy differences between polymorphs were even smaller, at a fraction of a kJ mol^{-1}. Our approximation, which lies intermediate between the isotropic and fully anisotropic models, seems like it should perform reasonably well.

1 N. S. Abraham and M. R. Shirts, 2017, arXiv:1712.00936 [cond-mat.mtrl-sci].

(204:[206]206) **Joost van den Ende** remarked: In Figure 4 and Figure S6 you show the dependence of the molar volume and electronic energy of CO_2 as a function of the different density functionals + dispersion correction used. Intuitively, I would think that using empirical dispersion corrections works better for larger molecules which fall “better” in the training set of the used dispersion correction. Therefore I would argue that CO_2 might not be the best test system to analyse the performance of empirical dispersion corrections. As a very experienced user of electronic structure calculations, do you think there is a size dependence of the studied molecules on the precision of empirical dispersion corrections?

Gregory Beran responded: Previous work by Gohr *et al.* suggests that the dispersion correction is quite important even for carbon dioxide.[1] For example, the lattice energy of phase I is 8.7, 22.7, and 23.5 using PBE, PBE-D2, and PBE-D3(BJ), respectively. Omitting the dispersion correction also leads to substantial overestimation of the molar volume and underestimation of the bulk modulus. Similar evidence for the reasonably good performance of various dispersion corrections for carbon dioxide, acetic acid and imidazole can be found in ref. 2.

More generally, modern dispersion corrections (Grimme’s D3, Distasio and Tkatchenko’s many-body dispersion, and Becke and Johnson’s XDM) are largely non-empirical and adapt well to a variety of systems. So I do not think this small molecules used for some of the benchmarks are a significant problem. Our paper did use the more empirical D2 for a few results, since D3 was not available in the Quantum Espresso software used, and it generally performed less well. We do not advocate general use of the D2 correction.

1 S. Gohr, S. Grimme, T. Soehnel, B. Paulus and P. Schwerdtfeger, *J. Chem. Phys.*, 2013, **139**, 174501.
2 A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Phys.*, 2012, **137**, 054103.

(206:[207]207) **Johannes Hoja** said: You have shown that the quasi-harmonic approximation at the Tier 3 level provides a good description of the thermal expansion and accurate sublimation enthalpies for small molecule crystals. Do you have an estimate of the level of molecular flexibility up to which one could obtain reliable sublimation enthalpies and entropies using the quasi-harmonic approximation?

Gregory Beran replied: This is a really important question and one we intend to investigate in the future. The quasi-harmonic calculations do appear to perform well for acetaminophen in the paper, which has slightly more flexibility than the completely rigid small molecules considered in earlier parts of the paper. Of course, many pharmaceuticals and other organic molecules of interest exhibit far more flexibility than any of the species considered in our paper. For sufficiently flexible species, it might make more sense to treat them with molecular dynamics (using a high-quality potential) or some other technique which can capture the anharmonicity in these species. On the other hand, our paper shows (*e.g.* Figure 2) that zero-point vibrational energy can contribute appreciably to the molar volume. Zero-point vibrational effects are not captured in classical molecular dynamics. One might also expect that the anharmonicity of the intramolecular vibrational modes will play an increasingly important role in larger molecules. The use of a one-dimensional Morse approximation for the anharmonicity described in your paper (DOI: 10.1039/C8FD00066B) might help address that problem. Future work will hopefully provide further clarity on the practical limits of the quasi-harmonic approximation for more flexible species.

(207:[208]208) **Michael Ruggiero** remarked: To follow up on the previous point, in Tier 1 the unit cell was constrained to an isotropic expansion/contraction, while in Tiers 2–4 the cell was allowed to optimize anisotropically. Does this have any influence on the results?

Gregory Beran responded: This is a good point. The Tier 1 results come from an earlier publication,[1] and our computational procedures have evolved since then. In that work, we optimized the geometries on-the-fly (instead of using the 1D energy–volume curves). While that allows the structures to relax in a fully anisotropic fashion, the Grüneisen parameters in that work were computed from volumes that were scaled isotropically. For ice and carbon dioxide, the experimental expansion is quite isotropic, so this is not a significant issue. However, acetic acid and imidazole expand somewhat more anisotropically, so there is probably some error introduced by computing the Grüneisen parameters under an isotropic approximation. Specifically, one would anticipate that the energy changes associated with those volume changes are too large, and that the frequency changes for the lowest-energy phonon modes will be overestimated. Hopefully the errors introduced by this approximation are small, and indeed the Tier 1 thermal expansion curves are fairly parallel to the experimental ones.

1 Y. N. Heit and G. J. O. Beran, *Acta Cryst. B*, 2016, **72**, 514–529.

(208:[209]209) **Michael Ruggiero** said: In the poster presented, the results using phonon dispersion are very compelling. In the cases where phonon dispersion isn't utilized and yields good results, in your opinion do you think that this is due to a cancellation of errors, a system dependent case where phonon dispersion isn't actually needed, or some combination of both?

Gregory Beran replied: For the Tier 3 hybrid approach, which we advocate due to its relatively low computational cost and good quality of results, the thermal expansion is pretty consistently underestimated across all the examples. As Jessica McKinley showed in her poster at the meeting (for another example, see the Electronic Supplementary Information for ref. 1), including phonon dispersion helps correct this. In contrast, the Tier 1 results, which do include phonon dispersion, track the experimental thermal expansion quite well. It is true that some of the calculations (most notably the Tier 4 DFT calculations, especially those with the D2 dispersion correction) expand the crystals more significantly with increasing temperature, even without including phonon dispersion. We suspect that this is due to error cancellation. As shown in Figures 2, 5 and S7 from the main paper (DOI: 10.1039/c8fd00048d), the expansion is suppressed somewhat once the MP2-level single-point energies are applied on top of those energies.

1 Y. N. Heit, K. D. Nanda and G. J. O. Beran, *Chem. Sci.*, 2016, 7, 246–255.

(209:[210]210) **Michael Ruggiero** asked: You've shown the importance of phonon dispersion for accurately determining thermodynamic properties, however such simulations are often prohibitively expensive. Is there any indication or marker for how much phonon dispersion to include to best marry computational cost and accuracy, or is it highly system dependent? To follow up, when thermodynamic values are computed accurately without phonon dispersion, what happens when dispersion is ultimately included? Are the non-dispersion simulations accurate due to a cancellation of errors, or do such methods simply not require dispersion?

Gregory Beran answered: We are pursuing additional strategies for capturing phonon dispersion effectively with reasonable computational cost and hope that it will be feasible to capture these effects in much larger crystals in the future. For example, Brandenburg *et al.* reported promising results in modeling the thermal expansion of carbamazepine with a density functional tight binding model.[1] Regarding the accuracy of the thermochemical predictions without phonon dispersion, one can obtain a reasonable estimate for the impact on the sublimation enthalpies or entropies by comparing the temperature dependent curvature of the Tier 1 *versus* Tier 3/4 results in Figures 6 and 7 of the main paper (DOI: 10.1039/c8fd00048d). Visual inspection suggests that the sublimation enthalpy errors at higher temperatures that are due to neglecting phonon dispersion are somewhere around 1–2 kJ mol^{-1} for carbon dioxide, acetic acid and imidazole, for example. The errors are considerably smaller for ice, which expands much less due to the strong three-dimensional hydrogen bond network. In other words, the magnitude of the errors will depend on the nature of the crystal packing. It is possible, of course, that fortuitous error cancellation will sometimes lead to reasonable-looking predictions of thermochemical properties despite modeling the thermal expansion poorly.

1 J. G. Brandenburg, J. Potticary, H. A. Sparkes, S. L. Price and S. R. Hall, *J. Phys. Chem. Lett.*, 2017, **8**, 4319.

(210:[211]211) **Sarah Price** said: Chick Wilson's paper[1] shows the thermal ellipsoids of a methyl group libration in form I paracetamol (acetaminophen) over the range of temperatures of the structures you have been modelling. The amplitude of the librations is huge, even as extrapolated to 0 K. Do you think that this may have affected your results for predicting the thermal expansion, as would the quasi-harmonic expansion be appropriate?

1 C. C. Wilson, Zero point motion of the librating methyl group in *p*-hydroxyacetanilide, *Chemical Physics Letters*, 1997, **280**, (5–6), 531–534.

Gregory Beran responded: This is an interesting question that merits future investigation. At the moment, we lack the data to answer this definitively. In other cases we have examined, neglecting phonon dispersion leads to an underestimation of the thermal expansion. It would be interesting to perform the phonon dispersion calculations and see how much of the thermal expansion the model captures. The magnitude of the residual disagreement between the predicted and experimental temperature dependence of the molar volumes could point toward the importance of effects such as the high-amplitude librations you point to.

(211:[212]212) **Sarah Price** remarked: Your work has been very important in showing that even zero-point lattice modes have a significant effect in expanding the lattice, implying that lattice energy minimisation with the perfect model would give errors for even the lowest-temperature crystal structures. The difference in methyl rotational amplitudes between polymorphs, depending on how tightly the methyl group is packed, could well affect the relative free energies. It would be unfeasible if we needed to calculate both quantum mechanical zero-point effects and had to do classical MD to get a full description of how the molecules are moving to get accurate thermodynamic and spectral properties.

Gregory Beran replied: This raises a good point. The quasi-harmonic approximation has clear limitations (*e.g.* high temperatures or highly flexible molecules). If one can obtain a good potential, switching to a dynamical treatment would be an appealing way to circumvent those limitations. However, classical molecular dynamics simulations do not capture the zero-point vibrational effects, which appear to contribute appreciably to molecular crystal volumes (*e.g.* Figure 2 in the main article (DOI: 10.1039/c8fd00048d)). There have been advances in path integral molecular dynamics techniques in recent years,[1] so perhaps it will become feasible at some point to use those types of approaches to capture the nuclear quantum effects more effectively.

1 T. E. Markland and M. Ceriotti, *Nature Rev. Chem.*, 2018, **2**, 0109.

(212:[213]213) **Seiji Tsuzuki** remarked: I have comments on the DFT calculations of the intermolecular interaction energy. You showed that the performance of the DFT calculations for evaluating the sublimation energy is not very good. The calculated sublimation energy depends on the calculated depth of the intermolecular interaction energy potential. The depth of the potential depends strongly on the choice of dispersion correction method and functionals used for the DFT calculations. I believe that the choice of PBE functionals is not so good. I believe that the performance of the calculations of the sublimation energies will be improved if you choose an appropriate dispersion correction method and functionals.

Gregory Beran responded: Our paper considers multiple different density functionals and van der Waals dispersion corrections. Our most important results use the B86bPBE-XDM model, which Erin Johnson and co-workers have demonstrated works well for molecular crystals and other systems where non-covalent interactions are important.[1,2] In addition, one of the important results in the paper is that when using hybrid approaches like Tier 3 that use wavefunction-based single-point energy refinement of DFT geometries/refinement, the final results are relatively insensitive to the choice of density functional/dispersion correction (see Figure 4 in the paper).

During the discussion I also shared two predicted values for the sublimation enthalpy of alpha methanol, which were taken from our recent paper.[3] These numbers compared the sublimation enthalpy computed with our fragment method at the CCSD(T)/CBS + periodic HF level of theory. The difference between the two values was whether we optimized the geometries and computed the phonons using MP2-level fragment calculations (*i.e.* Tier 1) or with plane-wave PBE-D3 (Tier 3). Because the single-point energies defining the energy–volume curve are at the CCSD(T) level, the results are not directly dependent on the PBE-D3 energy well depth. In that case, the Tier 1 sublimation enthalpy was 45.0 kJ mol^{-1}, which is in excellent agreement with the experimentally reported value of 45.7 ± 0.3 kJ mol^{-1}.[3] The Tier 3 value of 42.8 kJ mol^{-1} has an error that is several kJ mol^{-1} larger. This result is consistent with our other results presented in the *Faraday Discussions* paper (DOI: 10.1039/c8fd00048d), which demonstrate that while the hybrid wavefunction/DFT calculations perform well, they are not quite as good as those done entirely with higher-level wavefunction methods.

1 S. R. Whittleton, A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Theory Comput.*, 2017, **13**, 441–450.
2 S. R. Whittleton, A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Theory Comput.*, 2017, **13**, 5332–5342.
3 C. Cervinka and G. J. O. Beran, *Chem. Sci.*, 2018, **9**, 4622–4629.

(213:[214]214) **Seiji Tsuzuki** enquired: You used the HF method for the analysis of many-body interactions. The HF method overestimates the Coulombic interactions (electrostatic and induction interactions). The induction interaction is a major source of the many-body interactions in general. Therefore, the HF method probably overestimates the many-body interactions. On the other hand, DFT calculations can evaluate Coulombic interactions more accurately owing to the consideration of electron correlation. Why did you not use DFT calculations instead of HF calculations for the evaluation of the many-body interactions?

Gregory Beran answered: Our experience has been that when focusing on the long-range pairwise and many-body induction interactions, HF typically performs better than typical generalized gradient approximation (GGA) and can even be better than hybrid density functionals. There are examples of fragment methods that combine periodic DFT with monomer and dimer corrections computed with wavefunction methods,[1–3] However, those work best in cases where polarization effects are relatively weak. This can be understood in terms of the delocalization errors and problems in the many-body repulsion inherent in density functionals. Our earlier benchmark calculations on three-body interactions demonstrate this nicely.[4] Additional discussion of repulsion issues can be found in a paper by Gillan.[5] In the context of our fragment approaches, it is important to emphasize that the energies are dominated by the 1-body and shorter-range 2-body terms, which are computed with higher levels of theory. HF is only used for the non-pairwise additive contributions involving three or more molecules. In practice, such contributions are typically no more than 10–20%, and often they are 5–10%.

1 O. Bludsky, M. Rubes and P. Soldan, *Phys. Rev. B*, 2008, **77**, 092103.
2 S. Tsuzuki, H. Orita, K. Honda and M. Mikami, *J. Phys. Chem. B*, 2010, **114**, 6799–6805.
3 C. R. Taylor, P. J. Bygrave, J. N. Hart, N. L. Allan and F. R. Manby, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7739–7743.
4 J. Rezac, Y. Huang, P. Hobza and G. J. O. Beran, *J. Chem. Theory Comput.*, 2015, **11**, 3065–3079.
5 M. J. Gillan, *J. Chem. Phys.*, 2014, **141**, 224106.

(803:[215]215) **Gerit Brandenburg** communicated: The authors nicely analyzed how to combine different electronic structure methods to compute different components of the free energy (relaxed geometry, lattice energy, lattice dynamics). I wondering if their ranking (good to less good) is well justified in following the hierarchy of HMBI, DFT, FF. This will most likely depend on the type of many-body methodology that is used for the clusters and the type of embedding. As a first point, the many-body method employed is mostly MP2, which is certainly appropriate for CO_2 and ice I_h. However, for unsaturated π-stacked systems, MP2 is known to have deficiencies and DFT might be even more accurate. The literature data on the studied systems is summarized in Table 2. Could the authors comment on the accuracy of the employed HMBI variants for the lattice energy? The second issue is the force field describing the embedding. While I am confident that both geometries and energies from the HMBI approach will be only slightly perturbed by the embedding, this might not hold for the low-energy frequency modes. Have the authors studied this, for instance by enlarging the region that is treated with the many-body method?

1 A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Phys.*, 2012, **137**, 054103.
2 A. M. Reilly and A. Tkatchenko, *J. Chem. Phys.*, **2013**, 139, 024705.
3 A. Zen, J. G. Brandenburg, J. Klimeš, A. Tkatchenko, D. Alfè and A. Michaelides, *Proc. Natl. Acad. Sci. USA*, 2018, **115**, 1724.
4 S. Wen and G. J. O. Beran, *J. Chem. Theory Comput.*, 2011, **7**, 3733.
5 M. J. Gillan, D. Alfè, P. J. Bygrave, C. R. Taylor and F. R. Manby, *J. Chem. Phys.*, 2013, **139**, 114101.
6 M. Cutini, B. Civalleri, M. Corno, R. Orlando, J. G. Brandenburg, L. Maschio and P. Ugliengoa, *J. Chem. Theory Comput.*, 2016, **12**, 3340.

Gregory Beran communicated in reply: Lattice energies: We agree about the limitations of MP2 for describing non-covalent interactions. This issue in the context of molecular crystals is discussed in some detail in our earlier review.[1] There are multiple strategies for addressing these MP2 deficiencies, but one of our favored approaches and one we adopt in the current paper is MP2C.[2] This employs intermolecular perturbation theory to correct the treatment of the dispersion energy in MP2. Based on its excellent performance for non-covalent interactions and $O(N^5)$ scaling with system size N, David Sherrill and co-workers anointed MP2C the "bronze standard" for non-covalent interactions.[3] We have used MP2C to study the polymorphs of oxalyl dihydrazide,[4] and we have a new MP2D model[5] which recasts MP2C in a framework based on Grimme's D3

Table 2 Lattice energies of the studied crystals from MP2 and DFT compared to semi-experimental estimated and high-level results. Energies are given in kJ mol^{-1}

	Semi-exp.[a]	DMC[b]	MP2[c]	DFT[d]
CO_2	¬28.4	¬28.2	¬29.1	¬21.6
Ice I_h	¬58.8	¬59.2	¬58.7	¬72.8
Acetic acid	¬72.8	—	¬61.2	¬71.8
Imidazole	¬86.8	—	¬88.6	¬87.5

[a] Experimental sublimation enthalpies with estimations of zero-point and thermal effects from DFT computations as summarized in ref. 1–3. [b] Diffusion Monte Carlo data from ref. 3. [c] MP2 data from ref. 4–6, non converged basis set for acetic acid and imidazole. [d] B86bPBE-XDM for CO_2, acetic acid, imidazole, PBE-D3 for ice I_h.

dispersion correction. MP2D extends the range of applicability of MP2C to also include intramolecular interactions and facile geometry optimizations, which could prove useful in molecular crystals.

To compare against the lattice energies you cite, we extracted our best Tier 3 lattice energies at the MP2/aug-cc-pVTZ + AMOEBA and MP2C/CBS + periodic HF levels. Specifically, we evaluated the crystal energy at the minimum of the energy *versus* volume curves that result from single-point energy refinement on the B86bPBE-XDM geometries. As can be seen in Table 3, MP2/augcc-pVTZ + AMOEBA performs well for carbon dioxide, reasonably for ice, and poorly for acetic acid and imidazole. Imidazole in particular demonstrates the problem of MP2 over-binding the van der Waals dispersion interactions involving π systems.

Refinement of the single-point energies by replacing MP2 with MP2C, extrapolating to the complete basis set limit, and improving the many-body treatment from the AMOEBA force field to periodic Hartree–Fock leads to considerable improvement. Three of the four MP2C/CBS + periodic HF lattice energies lie within 2 kJ mol^{-1} of the semi-experimental value. For acetic acid, the error is larger at 4 kcal mol^{-1}. More detailed analysis of basis set effects and post-MP2 correlation contributions are found in our earlier studies.[6–8] Aside from acetic acid, the MP2C lattice energies are appreciably better than the results we obtained with our own B86bPBE-XDM calculations, which are also reported in Table 3.

For additional comparison, consider the sublimation enthalpy data in our paper. For acetic acid at 50 K (the lowest temperature we have experimentally derived values for), the predicted and experimental sublimation enthalpies are 63.4 and 66.0 kJ mol^{-1}, or an error of 2.6 kJ mol^{-1}. Looking at all four crystals, the errors in the low-temperature sublimation enthalpies are in the range of ~1–3 kJ mol^{-1} with MP2C (Figure 6). The sublimation enthalpies computed with B86bPBE-XDM in our paper exhibit errors of ~3–8 kJ mol^{-1}. Overall, the accuracy of sublimation enthalpies and lattice energies appears fairly consistent, especially given that the experimental data may well have uncertainties of a couple of kJ mol^{-1} or more.

From a pragmatic point of view, dispersion-corrected DFT has advanced tremendously over the last decade, and it performs very well for many molecular crystal problems.[9–11] In many cases its accuracy is more than sufficient to make useful predictions. However, in cases where one is not confident in the DFT results (*e.g.* when particularly small energy differences are obtained or when one observes high sensitivity of the results to the specific functional or dispersion correction), then the fragment approaches provide an alternative that can allow

Table 3 Lattice energies for the four small molecule crystals (kJ mol^{-1}). Errors *vs.* the semi-experimental values are given in parentheses

	Semi-exp.	MP2/aug-cc-pVTZ + AMOEBA	MP2C/CBS + periodic HF	B86bPBE-XDM
CO_2	28.4	27.7 (¬0.7)	26.3 (¬2.1)	23.9 (¬4.5)
Ice	58.8	55.4 (¬3.4)	57.4 (¬1.4)	67.9 (9.1)
Acetic acid	72.8	64.5 (¬8.3)	68.7 (¬4.1)	73.5 (0.7)
Imidazole	86.8	101.2 (14.4)	88.8 (2.0)	90.8 (4.0)

one to push toward even higher accuracy. This philosophy fits nicely with the Tier 3 approach: if DFT modeling of a given system proves inconclusive, the results can be refined with moderate amounts of additional computational work.

Embedding: First let us clarify that we do not embed in the traditional sense (*e.g.* there is no electrostatic embedding). Rather, the quantum mechanical monomers and dimers are computed in isolation, and those terms are used to replace the corresponding polarizable MM contributions from the super-system calculation.[6] That said, one parameter in our model is the distance interval over which one should transition smoothly from the short-range dimers treated at the high level (*e.g.* MP2 or MP2C) to long-range dimers modeled at the low level (AMOEBA or periodic HF). We typically choose a conservative interval of 9–10 Å for this transition, which ensures that the important pairwise interactions are well described by the high-level method. We have examined these cut-offs in some detail previously.[12] For the many-body contribution, which is evaluated solely at the low level of theory, replacing AMOEBA with periodic HF provides an improved description and insight into the approximation made using AMOEBA. In addition to the data shown in the *Faraday Discussions* paper (DOI: 10.1039/C8FD00048D), the positive impact resulting from this model improvement is also demonstrated in our recent work on the methanol phase diagram.[13] Switching from AMOEBA to periodic HF for many-body treatment brings the predictions of the both the sublimation enthalpy and the molar volume of alpha methanol into excellent agreement with experimental results.

1 G. J. O. Beran, *Chem. Rev.*, 2016, **116**, 5567.
2 A. Hesselmann, *J. Chem. Phys.*, 2008, **128**, 144112.
3 L. A. Burns, M. S. Marshall and C. D. Sherrill, *J. Chem. Phys.*, 2014, **141**, 234111.
4 S. Wen and G. J. O. Beran, *J. Chem. Theory Comput.*, 2012, **8**, 2698.
5 J. Řezáč, C. Greenwell and G. J. O. Beran, *J. Chem. Theory Comput.*, 2018, submitted.
6 G. J. O. Beran and K. Nanda, *J. Phys. Chem. Lett.*, 2010, **1**, 3480.
7 S. Wen and G. J. O. Beran, *J. Chem. Theory Comput.*, 2011, **7**, 3733.
8 Y. Huang, Y. Shao and G. J. O. Beran, *J. Chem. Phys.*, 2013, **138**, 224112.
9 M. A. Neumann, F. J. J. Leusen and J. Kendrick, *Angew. Chem., Int. Ed.*, 2008, **47**, 2427.
10 S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, **116**, 5105.
11 J. Hoja, A. M. Reilly and A. Tkatchenko, *WIREs Comput. Mol. Sci.*, 2017, **7**, e1294.
12 S. Wen, K. Nanda, Y. Huang and G. J. O. Beran, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7578.
13 C. Červinka and G. J. O. Beran, *Chem. Sci.*, 2018, **9**, 4622.

(214:[216]216) **Ivo Rietveld** opened the discussion of the paper by Grahame R. Woollam: To what level do the results of the crystal structure prediction help explain the experimental results: the observed stability behaviour of the polymorphs and the conditions under which those polymorphs can be obtained?

Grahame Woollam responded: The experimentally observed stability relationship is only reproduced in the crystal structure landscape when the energy contribution of disorder is taken into account. The disordered model of Form I is found to be lower in energy than all of the other predicted structures and there is no indication of a missing thermodynamically more stable form, which represents the outcome of the experimental polymorphism screens. Experimentally Form II is metastable to Form I, converting during competitive ripening studies. The calculated internal energy difference for Forms II and I is 0.54 kcal mol^{-1}, which compares well to the experimentally measured (by DSC) exothermic crystallisation enthalpy of 0.33 kcal mol^{-1} (following the melt of Form II). Concerning the conditions to prepare Form II, it cannot be formed by a simple recrystallisation procedure; we need to do something extreme to prepare the metastable form. The toluene/TBME method gives it; in retrospect, we can access both conformers from solution. The kinetically driven process results in a structure that accommodates both conformers in the frozen-in disordered Form II structure. From the crystal structure landscape, we see that the Form II disorder model moves up in relative energy in between the single components of ranks 1 and 58. Conversely, the Form I disorder model moves down to a lower relative energy than either single component (ranks 11 and 45) by more than $kT \ln 2$.

(215:[217]217) **Graeme Day** enquired: You have shown in this paper that, by properly modelling disorder in Form I, the prediction results improve and disordered Form I gets resolved to below the global lattice energy minimum of the landscape of ordered crystal structures. To be confident in now having the correct energy ranking, wouldn't we have to look at every structure on the landscape and assess whether it has disordered configurations? Have you looked at modelling the disorder in some of the other predicted structures to see if they gain a similar amount of stabilisation? In terms of assessing disorder as part of a workflow, could you also comment on how automated the procedure could be made: identifying structures with disorder, creating symmetry adapted ensembles and assessing the resulting free energy? Identifying structures that could be disordered seems to be the trickiest step.

Grahame Woollam replied: The contrast of the *ab initio* crystal structure landscape featuring the single component of experimentally metastable Form II as rank 1 prompted a closer look into the structures containing two molecules in the asymmetric unit. If we accept that in the case of loratadine the ordered structures containing two molecules in the asymmetric unit provided insight into the relevance and importance of configurational variation, it is even more informative that only families 1 and 2 contained $Z' > 1$ (Forms II and I, respectively). As such, we have explored all families that may exist in other ordered arrangements with lower symmetry that were not generated by the CSP procedure.

This is a pertinent point; we are actively working in this area. A stepwise approach is required; the largely automated workflow follows the order identified by Graeme Day. For the moment, we are working pragmatically by benchmarking using systems with experimentally determined disorder. Indeed, identifying *a priori* that a molecule could feature thermodynamically favourable disorder is a challenge. However, going back to the CSP of loratadine, the $Z' = 2$ structures were informative; Form I conformers 1 and 2 were found 18 and 19 times, respectively. Of those 18 times, conformer 1 (AAAA) coincided with a second conformer 1 within a $Z' = 2$ structure just once (*i.e.* AAAA and AAAA). Whereas eight out of the 38 $Z' = 2$ structures contain both Form I conformers 1 and 2 (AAAA and AABA), suggesting that the low-energy structures containing both conformers 1 and 2 result in a thermodynamically stable product, owing to the number of times they were generated *ab initio* (thermodynamically favourable). Additionally, all of the AAAA and AABA structures are within 0.90 kcal mol^{-1} of the lowest-energy structure.

Conversely, Form II conformers 1 and 2 appear together in a $Z' = 2$ structure just once (ABAA and BBAA) in rank 12, highlighting that the structure is stable at only 0.61 kcal mol^{-1} higher energy than rank 1; however, not as significant as the Form I AAAA and AABA configuration. As such, a thorough investigation of the $Z' > 1$ structures may provide the necessary information to get started on the symmetry adapted ensembles to assess the resulting free energies.

(225:[218]218) **Jason Cole** said: Graeme Day made the point that it would be interesting to analyse the possibility of lowering the energies of all structures in the predicted landscapes by using the same substitution of conformations. I was wondering whether this had been attempted for any individual examples, other than the experimental cases, to get an understanding of what the overall distortion to the CSP landscape might be expected as a consequence: *i.e.* how common is the motif in the modelled structures that gets energetically improved by considering the disorder?

Grahame Woollam answered: There has been no disorder modelling of the predicted structures outside of Form I (family 2, energy rank structures 11 and 45) and Form II (family 1, energy rank structures 1 and 58). Families 3–5 and the individual structures have not been assessed; the response to Graeme Day's question on this paper applies here. For families 1 and 2 the motifs don't change following the introduction of configurational variation; they remain as rods and mosaics coming from the single component systems and matching the experimental structures of Forms II and I, respectively.

(234:[219]219) **Christian Schön** remarked: The discussion about disorder brings to my mind the issue of time scales and self-averaging. It appears to me that frozen-in disorder is just a matter of time scales, and similarly when we discuss dynamical disorder. It is really important to take the observational time scales into account when deciding whether one should add a configurational entropy term to the free energy. Now, in macroscopic (and often even in mesoscopic) systems, one can never fully equilibrate all the disordered states, since we cannot explore enough of the configuration space of the energy landscape for such systems. However, we can hope to use the self-averaging concept: here we would divide the macroscopic system into many subsystems that are just at the size limit of being a "macroscopic" system, in the sense that we would still recognize them as miniature representatives of the bulk system. If the equilibration time scales of these sub-systems are short enough that we can assume the sub-blocks are pretty close to equilibrium on the observational time scale, then we can approximate the full free energy by a sum over the free energies of each such sub-system, where each sub-system will have a slightly different free energy (as if they were members of a Boltzmann ensemble), but their average will be a good approximation to the total free energy of the macroscopic system (that is where the term "self-averaging" comes from, of course). Here, it is important to keep in mind that when trying to identify systems with "controlled disorder", such as solid solutions (the disorder only applies to the distribution of *e.g.* two cations on the positions of one cation in the aristotype of the solid solution structure, but not to the underlying cation–anion–lattice structure!) during a crystal structure prediction using global optimization techniques, we need to analyze all the minimum structures found and place them into structure families – often there are several structure families within one chemical system (*e.g.* a rocksalt family and a wurtzite family in mixed alkali–halide compounds[1]). This can lead to several solid solution phases present in the same diagram with different lifetimes,[2] *i.e.* on short observational time scales the system can be in one solid solution phase or the other (without knowing that another feasible – perhaps more stable – solid solution phase exists in the system)! Proceeding in this fashion for molecular crystals by identifying families of local minimum structures sounds very promising.

On another point, you also discussed the asymmetric unit in your study, but I am not quite sure that this makes sense – once you insert disorder, the symmetry of the structure is lost, and thus the asymmetric unit becomes the full set of atoms/molecules in the cell. So how can you proceed the way you do once this symmetry is lost?

1 J. C. Schön, I. V. Pentin and M. Jansen, *J. Phys. Chem. B*, 2007, **111**, 3943–3952.
2 M. Jansen, I. V. Pentin and J. C. Schön, *Angew. Chem., Int. Ed.*, 2012, **51**, 132–135.

Grahame Woollam responded: In our paper we used symmetry-adapted ensemble theory to compute the configurational free energy of Form I. The free energy of this form is associated with disorder that is a thermodynamic equilibrium property. The approach is indeed only valid if the system under consideration has the time to equilibrate. Since neither crystal structure prediction nor symmetry-adapted ensemble theory provide rate constants for configurational changes, we take a pragmatic approach to decide whether the method is applicable. We first use symmetry-adapted ensemble theory to compute occupancy factors for each molecular conformation. If the calculated occupancies match the experimental occupancies to within a few percent, we take this as a confirmation that the system is indeed close to equilibrium and that symmetry-adapted ensemble theory applies. If, on the other hand, the computed and experimental occupancies strongly disagree, we conclude that the disorder is not an equilibrium property. Experimental occupancies are determined for the atoms in the asymmetric unit. In symmetry-adapted ensemble theory, individual configurations indeed break the crystal symmetry, but not so the actual ensemble. By averaging over all configurations of the ensemble and their symmetry copies, a computed asymmetric unit with computed occupancies is obtained that can be compared to the experimental results. It is important to realise that molecular crystals can get into a state close to thermodynamic equilibrium even if the rate constants for configurational changes in the bulk are so slow that the disorder appears static. When crystals grow, there is a surface layer in which molecules are significantly more mobile than in the bulk and are subjected to intermolecular interactions that resemble the crystal field of the bulk. The surface layer can hence equilibrate towards a bulk-like equilibrium state. In such cases, the standard textbook formulas for the calculation of the configurational free energy apply at the temperature of the crystal growth process.

*P*1 structures are commonly used to construct or depict a supercell for simplicity. However, the supercells of loratadine do maintain symmetry elements; this is a key aspect of the SAE model, and the disordered components interacting with one another.

(216:[219]220) **Aurora Cruz-Cabeza** queried: I would like to know whether the disorder in loratadine Form I is static or dynamic.

Grahame Woollam answered: The ratio of Form I conformers 1 and 2 changed with increasing temperature; conformer 1 increased from 0.51 at 10 K through to 0.55 at 273 K, with the change marginally outside of 3 times e.s.d. At 353 K (just 54 K below the melting point of Form I) the increase in ratio was discontinuous.

(217:[220]221) **Aurora Cruz-Cabeza** asked: If disorder in Form I is dynamic, then rotation of the group would need to be considered through molecular dynamics. This may be a limitation for your static model of disorder and poses a challenge for the computation of entropies and free energies.

Grahame Woollam replied: Dynamic disorder is temperature dependent disorder; could we consider that Form I is dynamic based on the information above? Static disorder is considered random and free to orient; while the carbamate tail has a number of configurations available, the "forward" and "back" positions are favoured in the crystal systems with alternating frequency. The disorder models don't represent the model as random and free to orient, but instead thermodynamically favourable through alternating configurations in microdomains. Substitutional disorder, which is largely seen in inorganic chemistry, is where chemically inequivalent species replace the host molecules in a substitutional manner, as the name suggests. If this is considered to be like static disorder, perhaps Form I could be considered a chemically equivalent solid solution. Definitions are taken from Disorder in crystal structures: new approaches in finding the best model, Markus Neuburger, Basel, 2012.[1] The QM static calculation may be regarded as a microstate of what an MD simulation could provide (with improved accuracy). The static calculation may represent a frame from the MD trajectory(?). However, we do not capture the information over time scales and temperature using the static approach. The experimental occupancy measured in Form I changes with temperature, although marginally. As such, the difference is subtle, which may ultimately be within the error associated with the MD force field calculations.

[1] Disorder in crystal structures: new approaches in finding the best model, Markus Neuburger, Basel, 2012.

(218:[221]222) **Aurora Cruz-Cabeza** said: The disorder present in loratadine Form II is static. The molecules are "frozen" in the conformations available during crystallisation. I wonder if crystallisations of Form II have been carried out at different temperatures and whether or not a change in occupancies has been observed.

Grahame Woollam responded: Many crystallisations have been attempted to prepare Form II; the conditions are specific and prescribed, and as such the form is crystallised with an onset *ca.* 20 °C and isolated at 4 °C. No other preparations have yielded Form II; instead Form I is prepared from the same solvent and solute composition if isolated at higher temperatures. The best Form I crystals, for example (and those used in the vT-SCXRD experiments), were obtained from toluene/TBME isolated at room temperature. The Form II single crystal structure presented in the paper was solved at 10 K using Synchrotron radiation. A 100 K structure solution was made using equipment available in-house; however the quality of the data did not allow the same conclusions to be drawn for Form II as for Form I. This relates to the size and quality of the metastable crystals formed under severe kinetically driven conditions.

(219:[222]223) **Rui Guo** asked: If loratadine Form II is indeed frozen-in during the specific crystallization setup, can we assume that the global minimum found in the CSP is still out there, waiting to be found? Are there any temperature-varied experiments on Form II, similar to those for Form I in Table 3?

Grahame Woollam answered: If the question relates to the rank 1 structure, which is an ordered component of Form II, given our time frame and crystallisation conditions, it is hypothesised that Form II is an accommodating structure that is adopted due to its ability to feature the two configurations of the cycloheptane head group. The structures of family 1 are less numerous on the crystal landscape, and from a solution mediated process we would expect to convert to the more stable Form I as the crystal landscape is dominated by this family of structures; as such, many nucleation avenues lead to this dynamically favoured form. It may be postulated that the Form II structure could anneal in the solid state to a fully ordered single-component structure given sufficient time, the correct temperature and perhaps humidity to enable relaxation. However, the saddle point for conversion of the cycloheptane head group from the BBAA to the ABAA conformation in this accommodating structure is increased (as calculated from the Form II disorder model structure) to 2.5 kcal mol^{-1}, as compared to 2.0 kcal mol^{-1} in the Form I disorder model structure. Regarding vT-SCXRD, the response to the previous question applies.

(222:[223]224) **Rui Guo** queried: With regard to the supercells you used to calculate the energy gain from configuration disorder, how sensitive is this energy to the size and shape of the supercells?

Grahame Woollam replied: When disordered sites are close enough to interact strongly, the isolated site model is not applicable and in principle one needs to consider all combinations of disordered configurations in a cell or supercell. However, many of these combinations are symmetry equivalent, and by limiting lattice energy calculations to one representative per symmetry equivalent set and working out the correct multiplicities, the computational effort can be greatly reduced. This approach is called symmetry-adapted ensemble theory. Typically the number of molecules per cell or supercell would be the right choice to feature in the calculation of energy gain, though with Forms I and II of loratadine an alternative selection allowed for the conservation of computational time. In both cases it was clear from geometrical calculations that the disordered sites interact along one-dimensional rods, with little interaction between the rods. Therefore, the energies refer to the number of molecules per rod, with two rods per supercell. The evaluation of the energy differences was again based on the energies of the entire supercell. As such the energy calculations are not sensitive to the size and shape of the cell/supercell.

(221:[224]225) **Rui Guo** asked: Did you use the supercells of the same size and shape for both forms of loratadine? Isn't it reasonable to assume that a more balanced choice of supercells should be based on their specific ways of molecular packing, which as you have shown clearly for the two forms are indeed rather different?

Grahame Woollam answered: The supercells were constructed independently; both feature doubling of the shortest axis (incidentally the *b* axis in both systems, and as such 1×2×1 supercells resulted) in order to capture the interactions between neighbouring disordered molecules – translated and featuring the stacking of the tricyclic ring systems.

(223:[226]227) **Jason Cole** enquired: Have any efforts been made to explore incommensurate refinements or possible very large supercells with respect to the crystallography to resolve or more accurately model the underlying statistical disorder in the forms?

Grahame Woollam responded: No refinements were made to the experimental structures to resolve an incommensurate structure or very large supercell.

(224:[225]226) **Jason Cole** asked: Were satellite peaks observed in the structural data, as these would be indicative of a more ordered supercell?

Grahame Woollam replied: There were no interpretable additional peaks observed for any structures.

(226:[227]228) **Sten Nilsson Lill** enquired: In relation to the comment by Aurora Cruz-Cabeza, have you investigated the barriers and dynamics involved for interconversion of the conformers seen in the disordered parts of your crystal structures I and II?

Grahame Woollam answered: There have not been any MD calculations to investigate the flipping of the alternating configurations in the crystal. However, the saddle point for conversion of the cycloheptane head group from the BBAA to the ABAA conformer in the Form II disorder model structure was calculated to be 2.5 kcal mol^{-1}, as compared to 2.0 kcal mol^{-1} in the Form I disorder model structure. This encourages the idea that the alternative cycloheptane head configuration is frozen-in. In terms of varying the ethyl carbamate tail in Form I, Table 5 demonstrates the energy gain or penalty from having a favourable or unfavourable assembly, with the 1×2×1 supercell featuring the configuration of conformer 2, 1, 2, 1 in the abcd configuration, *i.e.* a 50 : 50 occupancy at 0 K (see Figure 14).

(227:[228]229) **Susan Reutzel-Edens** remarked: You have indicated that Form II, the polymorph with “frozen-in” disorder, was obtained under a very specific set of conditions, so perhaps it is unsurprising that only a limited range of disorder was ever observed for this polymorph. As it is not unusual to find limited conditions in which to nucleate a specific polymorph, I am curious as to whether you have considered seeding to possibly grow Form II crystals under diverse conditions to assess the accessible range of disorder experimentally. Can you please comment?

Grahame Woollam responded: This is a valid point; perhaps the limited range of crystallisation conditions provide a limited range of disorder observed experimentally, because the structure accommodates those configurations available from the solution to the interface/bulk of the crystal. On the flip side, it could be that the accommodating structure is only formed under such conditions, whereby the available solution conformers are provided at a specified rate during desupersaturation through both the cooling crystallisation and self-induced drown-out concomitantly to form such a metastable structure, which is fleeting in any other organic solvent systems, yet long-lasting in the solid state.

(229:[229]230) **Sharmarke Mohamed** said: This work has shown that the crystal form landscapes generated from crystal structure prediction methods could be used to rationalise the observed crystal disorder for conformationally flexible molecules. It is promising to see that GRACE is capable of locating the two ordered components of each disordered structure of loratadine. Could the authors comment on the capabilities of GRACE in modelling cases of dynamic proton disorder (*i.e.* in ionisable molecules leading to salts or cocrystals) where you have covalent bond breaking and formation as a result of proton disorder? This is not just a question that is of general academic interest (although the problem of proton disorder clearly continues to be a challenge for crystallographers), but as was evident in the discussions, there are clear regulatory requirements in formulating active pharmaceutical ingredients in solid forms with well-defined covalent bonding. The relative energies of the neutral or ionised forms of molecular complexes can be significant and so it is interesting to know whether GRACE can suggest the correct molecular complex (salt or cocrystal) as well as the energy barrier/stability for proton migration.

Grahame Woollam replied: The prediction of proton positions and proton disorder in molecular crystals of pharmaceutical interest is still a challenge. The problem is not specific to GRACE, but rather related to the underlying level of theory in the computation of the potential energy hypersurface. According to the experience of one of the authors (Marcus Neumann), calculations with the PBE functional and the Neumann-Perrin dispersion correction have a substantial failure rate in identifying the correct proton position in cases where the proton position is ambiguous.

(231:[231]231) **Doris Braun** remarked: In Table 4 of your manuscript an enthalpy of recrystallisation of 0.33 kcal mol^{-1} is given, which is also compared to the internal energy difference obtained for Forms II and I (0.54 kcal mol^{-1}). Does the value really correspond to the enthalpy of recrystallization? How was it measured/integrated?

Grahame Woollam responded: The value relates to the exothermic transition (crystallisation) measured by DSC following the melt of Form II and prior to the subsequent melt of Form I.

(232:[232]232) **Doris Braun** commented: The enthalpy of fusion rule[1] is often applied to derive the energy differences between polymorphs, which gives an energy dfifference (enthalpy of transformation) of 1.74 kcal mol^{-1} for the two loratadine polymorphs.

1 A. Burger and R. Ramberger, On the polymorphism of pharmaceuticals and other molecular crystals: I. Theory of thermodynamic rules, *Mikrochimica Acta*, 1979, **2**, (3–4), 259–271.

Grahame Woollam replied: For the 10 K min^{-1} heating rate and conditions applied, this value is correct and represents the difference in enthalpy of fusion between the polymorphic pair, assuming full and complete melting, without interference of the recrystallisation exotherm.

(233:[233]233) **Doris Braun** asked: By integrating the melting peak of Form II (5.09 kJ mol^{-1}) and the recrystallisation peak of Form I (¬6.83 kJ mol^{-1}, if complete recrystallisation occurred), the value should be exothermic and not endothermic.

Grahame Woollam answered: It is indeed exothermic.

(235:[235]235) **Virginia Burger** enquired: I have a question about how each of the partially occupied conformers is identified within a disordered polymorph. Is disorder identified within the electron density map and are the individual conformers then fitted to the map and assigned occupancies based on their electron densities? Are the actual partially occupied conformers visible within the electron density map? Or could, for example, polymorph Form I have been assigned three (or more) partially occupied conformers instead of just two?

Grahame Woollam responded: We see two distinct conformations in the electron density.

(236:[236]236) **Virginia Burger** asked: In the electron density for polymorph Form I, does the crystallographer see two distinct configurations (for conformers 1 and 2), or is the data just “blurry”?

Grahame Woollam answered: We see two distinct conformations in the electron density.

(237:[237]237) **Michael Ruggiero** opened the discussion of the paper by Samuel A Jobbins: What happens when the system requires some large-scale re-orientation, perhaps occurring along a non-intuitive pathway? What limitations are there to the *metashooting* technique? Do you limit yourself by inputting the transition state directly? How prone is the algorithm to “missing” the right mechanism?

Samuel Jobbins responded: Thanks for your questions. The transition mechanism is deduced by implementing “plain” transition path sampling, which is completed prior to the *metashooting* procedure being carried out. Thus, by the time the *metashooting* takes place, the most likely subset of the transition path ensemble has been obtained (which is what is sampled using the *metashooting*).

The shooting steps that take place in the *metashooting* part of the work are to generate the correct velocity distributions for new trajectories on the biased energy surface, in order to ensure control in the variation of the collective variable.

Large-scale reorientations may occur in the transition path sampling runs. When one generates an initial trajectory using the geometrical–topological procedure we have followed, it is quite easy to deliberately avoid unfavourable regions of the trajectory space. If any large-scale reorientation is to occur, the path sampling algorithm should find it – however, it may take more iterations of the algorithm before the converged subset of trajectories is found.

Inputting the transition state directly should not itself not be a limitation – in fact, we consider it a positive of the method. The transition state for both the transition path sampling and the *metashooting* must be dynamical, as it is mandatory that it links the two basins of attraction. The initial path sampling intermediate is generated from the geometric trajectory, and the initial *metashooting* intermediate may be any point along a (converged) trajectory that links the two basins. From here, shooting to the two basins and filling up from the "outside-in" allows for accurate weighting of the entire underlying free energy landscape.

Finally, the algorithm should not "miss" the ideal trajectory. Transition path sampling, by definition, should always find the most favourable subset of transition pathways in trajectory space. It is these that are biased upon in the *metashooting* technique, and the trajectories that we report in the paper. Further information can be found in papers on transition path sampling.[1–3]

1 C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.*, 1998, **108**, 1964–1977.
2 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.
3 C. Dellago, P. G. Bolhuis and P. L. Geissler, *Advances in Chemical Physics*, ed. I. Prigogine and S. A. Rice, John Wiley and Sons, 2002, pp. 1–78.

(238:[238]238) **Michael Ruggiero** asked: How do you determine which trajectories to follow? Since you start from a pre-determined intermediate, what if your initial transition state is inaccurate? Is it possible to seed the system with specific information, *e.g.* vibrational normal modes, to help "steer" the algorithm?

Samuel Jobbins replied: Thank you for your intriguing questions. The trajectories utilised in the *metashooting* component of the work are extracted from the prior transition path sampling scheme, which converged upon the most likely subset of reactive trajectories linking two basins of interest after more than two hundred iterations. For more information on the path sampling procedure and how it converges upon the most plausible subset of the transition path ensemble, I'd recommend reading the papers which illustrate the method in detail.[1–3]

In our work, we generated an initial trajectory using a geometrical–topological technique involving the transformation of periodic nodal surfaces, as used in previous work by our group.[4,5] Intermediates could then be generated by interpolation, and a structure which linked the two basins (by propagation with MD forwards and backwards in time) was selected to be the initial intermediate. One of the advantages of shooting-off from a generated intermediate in this way is that it does not matter at all whether or not the initial transition state is a plausible starting structure – the starting trajectory need only link the two basins of interest. In fact, it is highly likely that none of the intermediate states in our initial trajectory have any physical relevance at all. The TPS algorithm will gradually steer the system away from this initial unfavourable regime towards a more favourable one. Thus, the initial regime should not matter – however, generating an energetically reasonable first trajectory certainly decreases the number of TPS iterations required before initial trajectory decoherence, and minimises the computing expenditure required to move towards more relevant regions of trajectory space.

Another advantage of this geometrical–topological approach is that, if desired, certain characteristics can be hard-coded into the initial trajectory. The nonlinear mapping from the two limiting configurations proceeds according to a set of particular deformation modes that are defined within the model. As a result, specific features may be deliberated encoded (or avoided!) in the initial pathway to greatly reduce the number of path sampling iterations required. Additionally, a direct appreciation of the favourability of a particular model can be ascertained by monitoring the conservation or disappearance of initial transition motifs whilst the path sampling procedure is being performed.

1 C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.*, 1998, **108**, 1964–1977.
2 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.
3 C. Dellago, P. G. Bolhuis and P. L. Geissler, *Advances in Chemical Physics*, ed. I. Prigogine and S. A. Rice, John Wiley and Sons, 2002, pp. 1–78.
4 S. E. Boulfelfel and S. Leoni, *Phys. Rev. B.*, 2008, **78**, 125204.
5 H. G. von Schnering and R. Nesper, *Z. Phys. B: Condens. Matter*, 1991, **83**, 407–412.

(239:[239]239) **Marcus Neumann** said: You mentioned that it's difficult to pick the order parameter, but is that also true if the start point and end point of a transformation are known? What if the start point and end point are predicted structures in the crystal energy landscape of a flexible organic molecule? Are there any general rules for the construction of the order parameter?

Samuel Jobbins answered: Thanks for your questions. Selecting the collective variable(s)/order parameter(s) is often the most difficult part of the job! There are a number of general rules for constructing suitable collective variable(s):

1) The collective variable(s) must distinguish between the initial and final configurations of interest, and (if appropriate) all relevant intermediate and transition state regions, meaning no overlap between regions of importance;

2) The collective variable(s) must be able to describe all of the events of importance in a given transformation, particularly those that are "slow", and tolerate fluctuations in thermal energy;

3) The number of collective variables should not be too large – otherwise the calculation will be prohibitively expensive.

It is immediately apparent, however, that points 2) and 3) can be quite mutually exclusive, as characterising all the necessary reaction coordinates to describe a particular system can require quite a number of collective variables! In addition, omitting a relevant collective variable can lead to basin over-filling or failure to sample relevant regions of configuration space. (Indeed, this is why we have included the 3rd coordination sphere in the work as well as the 1st, as biasing only the 1st coordination sphere proved too coarse to distinguish between competing intermediates and produced an incorrect energy scheme.[1,2]) In my opinion, the best approach is to characterise the system using as few collective variables as possible, but by cleverly designing them such that they are truly "collective" and that each order parameter encapsulates as much information about the system and the transformation of interest as possible.

Knowing the start and end points of a simulation can obviously help to intuitively deduce a relevant set of collective variables. Sometimes it is quite evident – for example, when simulating a chemical reaction, a certain bond length might be the only variable of interest. However, in more complex systems with many degrees of freedom, ascertaining a fundamental set of collective variables can be a very non-trivial task. For example, in the case of flexible organic molecules, a reasonable start would be to address which particular bonds, angles, torsions, *etc.* play a vital role in the transformation under scrutiny. Some combination of these important reaction coordinates might then prove sufficient to correctly generate a free energy profile using metadynamics or the *metashooting* procedure.

One final point – in the scheme described in the paper, we used TPS initiated from a model first trajectory. As such, knowing the start and end points of the simulation are mandatory if one wishes to utilise transition path sampling and the *metashooting* technique in its current formulation. But, as I've said, whether or not this helps to determine the collective variables depends highly on the nature of the system and the event under scrutiny.

1 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 12562–12566.
2 A. Laio and F. L. Gervasio, *Rep. Prog. Phys.*, 2008, **71**, 126601.

(240:[240]240) **Julian Helfferich** remarked: We have so far discussed the collective variables, but I would like to address the methods. How far does your *metashooting* procedure depend on the methods you have chosen, metadynamics and the shooter method? Could you exchange them for other, potentially cheaper, biasing and transition path methods?

Samuel Jobbins responded: Thank you for your question. The *metashooting* procedure is by its very nature dependent on its two component techniques – metadynamics and transition path sampling. The TPS steps in the *metashooting* procedure control the variation in the collective variable, ensuring that it remains within the trajectory space corresponding to the reactive transition path ensemble. The metadynamics component deposits the bias within only this relevant region, allowing for the underlying energy surface corresponding to the transition to be explored efficiently and in precise detail. It is conceivable that the two core methods could be "switched out" for other biasing/transition path methods. However, we have illustrated that the union of TPS and metadynamics currently implemented can produce exquisite and accurate detail, particularly of the intermediate and transition state regions. Whether or not changing the methods for others would have an effect on the accuracy or the expense of the work, or whether unifying other techniques in this manner would be as robust or effective, remain open questions. For myself, rather than changing the methods utilised, I would advocate tuning and optimising the parameters involved in the two techniques of the procedure, in order to maximise the efficacy and speed of the *metashooting* analysis.

(241:[241]241) **Pablo Piaggi** said: I would like to comment on the performance of metadynamics. First of all, there is a common misconception that the metadynamics algorithm becomes slower as the simulation progresses. This is only true in the most rudimentary implementation, in which an increasing number of Gaussians have to be summed in order to calculate the bias potential. In modern implementations, a grid is used to construct the bias potential and therefore there is no need to sum all the Gaussians.[1] As a consequence, the cost of the algorithm is constant in time. Furthermore, other methods, such as adaptive biasing force, are also based on collective coordinates (CVs) and typically the calculation of the CVs is the bottleneck of these methods.

[1] https://plumed.github.io/doc-v2.4/user-doc/html/_metadyn.html

Samuel Jobbins replied: I wholly concur with this contribution. Thank you for bringing it to attention!

(242:[242]242) **Pablo Piaggi** remarked: It has been shown[1] that in finite size systems the barriers can be different to those in an infinite system. In the case under consideration the transformation will probably proceed by nucleation and growth, and I suspect the size of the simulation box is smaller than the critical cluster. Did you assess whether the size of the system affects the results?

[1] M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti and M. Parrinello, Molecular-dynamics simulations of urea nucleation from aqueous solution, *Proc. Natl. Acad. Sci. USA*, 2015, **112**, E6–E14.

Samuel Jobbins responded: Thanks for your question. We have not directly tested whether or not the size of the system would affect the results – however, we strongly believe that it would, as it is likely that the precise nature of the motifs seen in the transition path sampling runs is highly dependent on the simulation system size. Addressing this would be a useful validation of the method, and indeed would help to determine whether or not the system can be suitably scaled-down and studied by *ab initio* methods. However, based on previous work,[1,2] we believe that the system size utilised produced representative motifs and nucleation/growth events, whilst keeping the number of atoms in the system tractable for efficient MD/metadynamics simulations.

[1] S. E. Boulfelfel and S. Leoni, *Phys. Rev. B.*, 2008, **78**, 125204.
[2] D. Zahn, Y. Grin and S. Leoni, *Phys. Rev. B.*, 2005, **72**, 064110.

(243:[243]243) **Sarah Price** enquired: It is really spectacular that you have identified intermediate basins and transition states in this solid state transformation of ZnO. Your calculations were done with the simple exp-6 interatomic potentials that Robin Grimes derived in 1993, and were later validated for Zn0 in 2007 in your group, but this validation is unlikely to have included these local basins and transition state regions of the potential energy surface. Please can you comment on the extent to which your results could be an artefact of the potentials used?

Samuel Jobbins answered: Thank you for your question and for your high praise about the work! The structures represented by the extrema on the free energy surface that are described in this work correspond very closely to the previously reported structures. Notably, the five-coordinated tetragonal and hexagonal forms of ZnO have been discussed at great length in the literature. Indeed, there still remains considerable debate as to the roles of these two intermediates – for example, one *ab initio* study suggested that the tetragonal intermediate was vital to the WZ-RS transformation in ZnO,[1] however an experimental study utilising high resolution angular dispersive X-ray diffraction appeared to indicate that the hexagonal pathway was more favourable for zinc oxide transformations.[2]

Subsequent work has shown that the two intermediates are in direct competition with one another, and that a number of external factors are relevant, such as pressure, temperature and the arrangement of local coordination motifs.[3] In addition, the work of Zagorac *et al.* found a number of the metastable ZnO phases using *ab initio* techniques, some of which correlate directly with the configurations observed in this work.[4] Finally, in the group's previous work, the existence and energy rankings of intermediates found using this force field were validated using density functional theory calculations.[5]

Therefore, all of these works, both theoretical and experimental, report the same (or very similar) metastable configurations as those presented in this paper. As a result, I believe I can say with some certainty that the chances of the intermediate basins having little physical significance is very low – however, as always, further study (perhaps using a different potential model or a higher level of theory) would be a very suitable future endeavour to ensure this!

1 A. M. Saitta and F. Decremps, *Phys. Rev. B.*, 2004, **70**, 35214.
2 H. Liu, Y. Ding, M. Somayazulu, J. Qian, J. Shu, D. Häusermann and H. Mao, *Phys. Rev. B.*, 2005, **71**, 212103.
3 J. Cai and N. Chen, *J. Phys.: Condens. Matter*, 2007, **19**, 266207.
4 D. Zagorac, J. C. Schön, J. Zagorac and M. Jansen, *Phys. Rev. B.*, 2014, **89**, 075201.
5 S. E. Boulfelfel and S. Leoni, *Phys. Rev. B.*, 2008, **78**, 125204.

(245:[245]245) **David McKay** queried: The speaker responded to a previous question on the computational expense of metadynamics by saying that particular intermediate states are targeted to avoid spending time in unwanted minima. He also remarked that regarding accuracy, to verify these results he would like to be able to repeat the process using the DFT potential in metadynamics and transition path sampling. Rather than repeating the whole process with DFT, is there any scope for only using DFT as a correction to key structures along the potential energy surface?

Samuel Jobbins replied: Many thanks for your interesting question. It may be conceivable to try to treat the entire transition pathway (including the key structures) using density functional theory on top of the existing results. In particular, a few iterations of the *metashooting* procedure could be carried out using density functional theory, in order to see if there is significant deviation from the obtained results and how the energetics of the new trajectories compare with older ones.

However, both the transformation mechanism (as deduced by transition path sampling) and the free energy surface (produced by *metashooting*) have been ascertained by describing the atomic interactions using a classical Buckingham-style empirical pair potential.[1] One would hope that the scheme and (in particular) the intermediate regions produced by the potential would correlate well with any results that would be obtained using a higher level of theory. Previous work by our group has shown that the intermediates generated by the force field can be stabilised using density functional theory and have the correct energy rankings.[2] However, to be absolutely sure about this work, the entire process could be repeated again with a different potential model or higher level of theory.

1 D. J. Binks and R. W. Grimes, *J. Am. Ceram. Soc.*, 1993, **76**, 2370–2372.
2 S. E. Boulfelfel and S. Leoni, *Phys. Rev. B.*, 2008, **78**, 125204.

(246:[246]246) **David McKay** asked: Another consideration when going to more accurate methods is system size, which in this case is large for DFT. Does the study at force-field level provide any pointers as to how the system size could be reduced for DFT calculations, while retaining a reasonable model system?

Samuel Jobbins responded: It is of course very true that the mechanism deduced by transition path sampling is dependent on the size of the system. The system size of 1200 Zn–O pairs was chosen as this is small enough for computational tractability but large enough for critical nucleation and growth events to be visualised.[1,2] To scale down for a density functional theory-type study, it is conceivable that one could ascertain the size of these critical nuclei and the number of atoms required for their formation, and construct a new model system with a sufficient number of atoms to produce such features (with some additional neighbouring atomic sites, so that an interface is present when considering the periodic boundary conditions).

However, one must be very careful when scaling down the system. Making the system too small can completely suppress the formation of key intermediates, and the system may get “locked-in” to a regime of unrealistic trajectories, both in terms of the details which emerge and the time scales over which they occur. I would not be surprised if 1200 Zn–O pairs actually approached the minimum size required to generate such meaningful events, however it would be interesting to experiment with this to properly determine the minimum system size that could be studied with density functional theory.

1 S. E. Boulfelfel and S. Leoni, *Phys. Rev. B.*, 2008, **78**, 125204.
2 D. Zahn, Y. Grin and S. Leoni, *Phys. Rev. B.*, 2005, **72**, 064110.

(247:[247]247) **Christian Schön** said: We have seen similar intermediates (*e.g.* the 5-5 structure, the ionic analogue of the hexagonal BN structure) on the energy landscapes of related systems,[1–3] which were also stable minima when locally optimized on the *ab initio* energy landscape. Thus, I would believe they also exist in your system. On another issue, I would hesitate to make the system smaller – you might be in danger of locking the system into a limited number of (perhaps unrealistic) pathways between the regions you want to connect. If you recall the water-freezing simulations by Ohmine and co-workers *ca.* 15 years ago,[4] they found that even the large “crystalline” nuclei that had formed inside the liquid phase after about 100 ns vanished again (I think they needed about 200 ns before the whole system crystallized). So, a smaller system size might lead to unrealistically fast freezing rates. Do you ever see paths where seeds of the new phase grow and then vanish again?

You also mentioned that you could measure activation energies – can you also deduce activation entropies, or more generally measure the generalized barriers[5] *via* the probability flows on the landscape of the system between metastable configurations?[6,7] Perhaps *via* effective transition rates?

1 J. C. Schön and M. Jansen, *Comp. Mater. Sci.*, 1995, **4**, 43–58.
2 J. C. Schön, *Z. Anorg. Allg. Chem.*, 2004, **630**, 2354–2366.
3 D. Zagorac, J. C. Schön and M. Jansen, *J. Phys. Chem. C*, 2012, **116**, 16726–16739.
4 M. Matsumoto, S. Saito and I. Ohmine, *Nature*, 2002, **416**, 409–413.
5 J. C. Schön, M. A. C. Wevers and M. Jansen, *J. Phys. Cond. Matt.*, 2003, **15**, 5479–5486.
6 M. A. C. Wevers, J. C. Schön and M. Jansen, *J. Phys. Cond. Matt.*, 1999, **11**, 6487–6499.
7 S. Neelamraju, J. C. Schön, K. Doll and M Jansen, *Phys. Chem. Chem. Phys.*, 2012, **14**, 1223–1234.

Samuel Jobbins answered: Thanks for your correspondence and queries. I appreciate the vote of confidence with respect to the existence of the intermediates and transition states found in the work! I am very confident that they are meaningful metastable minima, as they have been seen in a great deal of previous work, including your own. I also am inclined to agree with your comment regarding system size – it would be interesting to see how small one could go whilst still observing plausible mechanistic details, however my intuition tells me that this limit would not be much smaller than the system size that we have used in this work! During the initial transition path sampling scheme, there are multiple examples of “failed” trajectories, in which a transformation begins, but then “fails” and reverts back to the existing structure. Such failed trajectories were often seen in the very early stages of transformation, particularly from rocksalt to zincblende, around the time of the formation of the cubic seed. Another frequent example of trajectory “failure” was when the transition became “stuck” in the intermediate 5-coordinate basin.

Pairing TPS and metadynamics allows us to globally map the free energy. In transition path sampling, the system still spends the most time around the main basins of attraction, while *metashooting* can weight even elusive intermediates accurately. TPS is a method that allows the generation of reactive trajectories without bias, as single trajectories follow their natural dynamics without external perturbation. Reaction rates are thus accessible by combining TPS with methods like Umbrella Sampling. A more precise outcome is provided by Transition Interface Sampling (TIS).[1] Additionally, the separation of enthalpic and entropic contributions can be addressed in metadynamics by introducing an instantaneous measure of entropy, based on pair correlation functions. All of these measures require a precise knowledge of the free energy, including possible intermediates, which is exactly what is provided by our *metashooting* methodology.

1 T. S. van Erp, D. Moroni and P. G. Bolhuis, *J. Chem. Phys.*, 2003, **118**, 7762–7774.

(248:[248]248) **Christian Schön** asked: Regarding your shooting algorithm: since there are many competing basins in your system, how can you sit “in the middle” and shoot, and still reach the two basins between which you try to establish a transformation route? Where is the “middle”, exactly?

Samuel Jobbins responded: The shooter algorithm here corresponds directly to the original algorithm reported in the transition path sampling literature. More information about the transition path sampling method can be found in ref. 1–3.

When referring to the “middle”, we simply mean a dynamical mid-point of the current trajectory that links the two basins of interest. In the beginning of the transition path sampling work, the mid-point is an interpolated intermediate from the geometric trajectory that successfully links the two basins of attraction. Subsequent mid-points are dynamical intermediates from the latest reactive trajectory. In the *metashooting* component, the transformation route has already been established by transition path sampling, and so shooting from the mid-point still reaches the two basins of interest along the converged subset of trajectories. By the time *metashooting* is carried out, the transition path ensemble has been sampled and the most likely subset of trajectories has been found – it is these trajectories which are biased and evaluated by the *metashooting* procedure.

1 C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.*, 1998, **108**, 1964–1977.
2 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.
3 C. Dellago, P. G. Bolhuis and P. L. Geissler, *Advances in Chemical Physics*, ed. I. Prigogine and S. A. Rice, John Wiley and Sons, 2002, pp. 1–78.

(249:[249]249) **Christian Schön** enquired: Does your procedure measure the probability flows directly? Also, we should not confuse the local “entropy”/“free energy” measured as a function of the order parameter during the metadynamics phase of the simulations with the “entropic” part of the generalized barrier that is related to the probability flows and thus contributes to the rate constants of the processes being analyzed. After all, a local free energy is only defined on observational time scales that are larger than the equilibration time and shorter than the escape time from the region, so you might be able to measure local densities of states – a static quantity – but it might not be permissible to compute the free energy, which is a cale dependent quantity!

Samuel Jobbins replied: I’m not sure I fully agree with this analysis. The Laio *et al.* paper on metadynamics describes in considerable detail how, at the infinite time limit, the deposited bias is approximately equal to the negative of the free energy, plus an irrelevant additive constant.[1] There are a number of ways that the error in this can be estimated, and also a number of methods which can be implemented to reduce this error. Utilising the well-tempered prescription of metadynamics also helps to overcome any potential issues here associated with over-filling (and hence over-estimating the free energy barriers). However – to reiterate – Laio *et al.* clearly show that the free energy (an equilibrium property) can be very well estimated using non-equilibrium dynamics in metadynamics. This is indeed one of the great selling points of the method, and as such both enthalpic and entropic contributions should be accounted for in the present scheme.

More generally, one can consider the situation in this way: large free energy barriers represent a major limitation for the exploration of reactive events – even more so if *ab initio* methods are used. In “plain” molecular dynamics, the escape time in the presence of energy barriers much larger than k_BT is, in principle, infinite. Metadynamics implements a coarse-grained representation of the true high-dimensional potential energy surface of interest, which will contain a large number of low-lying saddle points. In the lower dimensional coordinate space of the collective variables, the free energy surface is smooth. In this regime, the basin escape time is finite, allowing for a global mapping of the free energy, which converges within a finite time.

1 A. Laio and F. L. Gervasio, *Rep. Prog. Phys.*, 2008, **71**, 126601.

(250:[250]250) **John Mitchell** asked: Systems like chemical reactions, with few degrees of freedom, have tightly defined mechanisms corresponding to a specific trajectory or a narrow range of trajectories across the potential energy surface. Other energy landscapes, like those for protein folding, with very many degrees of freedom, have much looser families of trajectories connecting regions. Would you classify the phase transition you are looking at as a single mechanism, or a broad family of region-connecting trajectories? Do you think your methodology is better suited for tightly defined mechanisms or for broader families of trajectories?

Samuel Jobbins answered: Thanks for your intriguing question. The underlying energy landscape corresponding to a condensed matter phase transition can be very coarse, owing to the many degrees of freedom in the system. This is of course dependent on the number of atoms in the system, but for 1200 Zn–O pairs the complexity of the (complete) energy landscape is probably intermediate between the two examples you have cited here! In this particular transition, there are a great number of transition pathways possible, each of which has a particular probability (as determined by the energy of the transformation). The initial transition path sampling procedure samples reactive pathways in trajectory space and ultimately converges upon the most likely subset of reactive trajectories linking the two basins (prior to the *metashooting* technique actually being utilised). For more information on transition path sampling, I'd recommend taking a look at ref. 1–3.

These lowest energy reactive trajectories correspond to the transition mechanisms reported in the paper. The *metashooting* procedure is then only applied to this lowest energy subset of reactive trajectories in the coarse-grained, lower dimensional space of the collective variables, which is why the final free energy profile looks relatively smooth and well defined. I see no reason why the methodology should not work on more simple or complex cases. Systems with fewer degrees of freedom would be easier to characterise and define in terms of reaction coordinates, and would likely converge more quickly than their larger counterparts. More complicated systems may be more expensive to compute, however if an appropriate order parameter is chosen, transition path sampling should still be able to find the corresponding favourable reactive subset in trajectory space. These trajectories can then be biased using *metashooting* to generate the energy landscape corresponding to the "path of least resistance" for a given system.

1 C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.*, 1998, **108**, 1964–1977.

2 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.

3 C. Dellago, P. G. Bolhuis and P. L. Geissler, *Advances in Chemical Physics*, ed. I. Prigogine and S. A. Rice, John Wiley and Sons, 2002, pp. 1–78.

(251:[251]251) **Marcus Neumann** said: In a periodic supercell, many paths leading from a highly symmetric start point to a highly symmetric end point are related by symmetry. Does your method somehow exploit these energy relationships?

Samuel Jobbins replied: The initial trajectory is generated using a geometrical–topological approach, in which periodic nodal surfaces are wrapped around the limiting structures and transformed according to a pre-defined set of deformation modes. In this respect, symmetry can be incorporated into the initial trajectory. Such a scheme has been used in previous work by our group.[1,2] However, in its current implementation, the actual *metashooting* procedure does not take symmetry into account at all when sampling the trajectory space or applying the metadynamics scheme. However, this could certainly be worth looking into in the future, as it may well help to increase the efficiency of the technique.

1 S. E. Boulfelfel and S. Leoni, *Phys. Rev. B.*, 2008, **78**, 125204.
2 H. G. von Schnering and R. Nesper, *Z. Phys. B: Condens. Matter*, 1991, **83**, 407–412.

(819:[252]252) **Scott Woodley** communicated: What is the definition of your energy? You state in your paper that you use CP2K and mention the use of Buckingham potentials with parameters derived from Grimes *et al.*, but is this the Shell Model or rigid ions? If the latter, how would you expect your results to change if you added oxygen polarisation into your model?

Samuel Jobbins communicated in reply: In the molecular dynamics and transition path sampling runs, the total energy (as computed by CP2K) was used as the energy function. In the *metashooting*, the free energy is reconstructed as the negative of the deposited bias potential, as explained in the original paper on metadynamics. We indeed use the Buckingham potential of Binks and Grimes,[1] however in this work the oxygen polarisation term was not included. Initial calculations showed that just using the core charges and van der Waals interactions was sufficient to describe the relevant bulk phases of ZnO, with reasonable values for the geometric parameters and transition pressures. I personally suspect that the oxygen shells would actually make little difference to the transition mechanism – however, there may be some subtle effects, and there would be considerable differences if one, for example, wanted to calculate phonon dispersion curves. Perhaps repeating the procedure with the shells (or with another potential model) might be a good future endeavour to help validate the methodology.

1 D. J. Binks and R. W. Grimes, *J. Am. Ceram. Soc.*, 1993, **76**, 2370–2372.

(824:[253]253) **Noa Marom** communicated: You showed an example of solid state transformation from one crystal structure to another. Can *metashooting* also be used for amorphous to crystalline transformations?

Samuel Jobbins communicated in response: Thanks for your question! In principle, any system that can be described with a suitable set of collective variables/order parameters for the metadynamics and TPS stages of *metashooting* can be studied using our technique. In fact, we are currently in the process of studying a phase transition between two configurations of a metal–organic framework, which is believe to proceed *via* an amorphous intermediate. Once the most likely subset of trajectories linking the two configurations is found, we intend to apply the *metashooting* method to quantify the underlying free energy surface. Of course, as usual, the difficult part of this is determining which reaction coordinates are relevant to the transition! My advice for treating an amorphous system would be to attempt to exploit any short-range order present in the system, and to use mean coordination sequences or Steinhardt parameters to attempt to globally characterise the system.

(825:[254]254) **Noa Marom** communicated: Do you envision using machine learning to identify the best collective variables for a system?

Samuel Jobbins communicated in reply: What a fascinating idea! In principle, the idea of using a machine learning algorithm to substantially reduce the dimensionality of a problem to a handful of tractable collective variables is a very enticing prospect indeed (and arguably is a textbook example of the merits of machine learning). In practice, however, there are often so many possible variables to utilise in each system that it is quite hard to envisage a machine learning algorithm being able to successfully capture the often subtle chemical/physical nuances in every possible system. Often, a great deal of chemical intuition (and sometimes luck!) is required to ensure that only appropriate variables are chosen, and that all of the important variables are considered. I, however, am only just beginning to enter the machine learning world and would love to know if this could be done. Do you have any suggestions about how this could be done in practice? There are so many different systems (and hence collective variables) to study – how would one train such an algorithm to ensure that it could make useful suggestions for such a disparate selection of systems?

(300:[300]300) **Marcus Neumann** opened the discussion of the paper by Alexandre Tkatchenko:† Fig. 9 and Fig. 10 seem to suggest that thermal expansion affects the lattice free energy only weakly if phonons are treated in the harmonic approximation, and much more strongly when the anharmonicity of phonon modes is taken into account. Is this interpretation correct?

Johannes Hoja responded: This seems to be the case for this particular system. However, for a general statement further investigations involving a variety of systems would be necessary.

(302:[301]301) **John Mitchell** enquired: Thinking about the things you added to your model, I'm not surprised that you added mode-by-mode analysis of the vibrations. I am looking for an accuracy of 3 kJ mol^{-1} in my applications, and would also expect to require that. However, I am surprised by how much difference the many-body dispersion makes, compared to using just pairwise dispersion. This might be disturbing, since omission of many-body effects seems to lead to significant increases in error. Is this widely true across drug-like molecules? Are

---

† Alexandre Tkatchenko's paper was presented by Johannes Hoja, *University of Luxembourg*.

there other non-obvious sources of error that we need to be aware of that are currently missing from most models?

Johannes Hoja replied: For the five systems discussed in our paper, the mean absolute deviation between the PBE+TS and PBE+MBD energies varies between 1.8 and 3.2 kJ mol^{-1}. Therefore, we observe significant effects for all systems but their magnitude seems to depend on the molecular size and flexibility. Figure 2 in our paper shows a correlation plot between PBE+TS and PBE+MBD. Given the spread of the data points, even fitting the best possible pairwise model to the MBD data would not provide sufficient correlation. Note that the differences between PBE+TS and PBE+MBD originate from many-body dispersion interactions and also from the included dielectric screening effects. Therefore, Figure 3(b) illustrates purely the effect of many-body interactions. It can be seen that contributions of up to about 5-atom interactions are necessary to converge the relative stabilities. A discussion about many-body dispersion effects for molecular dimers, supramolecular systems and molecular crystals can be found in ref. 1 and 2. Non-obvious sources of error for static energy calculations could be revealed by benchmarks on larger and more flexible crystals using Quantum Monte Carlo (QMC).

1 A. M. Reilly and A. Tkatchenko, *Chem. Sci.*, 2015, **6**, 3289–3301.
2 A. Ambrosetti, D. Alfè, R. A. DiStasio Jr. and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2014, **5**, 849–855.

(303:[302]302) **Graeme Day** asked: I have a question about the validation of crystal structure prediction methods: when evaluating methods used to rank structures, how do we know when we have the correct ranking? Target systems for benchmarking should be chosen carefully and should be experimentally well characterised.

Are the blind test molecules the best systems to revisit when assessing new methods? It does not seem to me that the blind test molecules are selected in a way that makes them good benchmarks. Apart from a few exceptions, such as the polymorphic molecule XXIII, most of the blind test molecules have not been experimentally studied very much at all. They are included in the blind tests simply because their crystal structures have not been published. So, once they are no longer blind targets, I don't think that they are valuable for validation. Should the community agree on some better sets of molecules to use for testing?

Johannes Hoja answered: I agree that the blind test systems are not ideal benchmarks (with the exception of system XXIII). It would be better to create a set of experimentally well-studied systems, for which extensive polymorph screening has been done and for which reliable experimental relative stabilities are available (maybe at different thermodynamic conditions). Ideally, this set should cover a variety of different types of molecular crystals, *i.e.* crystals with different molecular size and flexibility, salts, co-crystals, solvates, and also disordered systems.

(306:[303]303) **Susan Reutzel-Edens** said: As a follow-up to Graeme's general question about the use of blind test molecules for benchmarking purposes, the evaluation of systems from the literature that are well characterized experimentally is likely to produce more meaningful results than for systems chosen simply for being blind test targets – unless, of course, experimental validation is forthcoming. In this work, what experimental validation data is available to assess the various energy models that were used for the CCDC blind test molecules?

Johannes Hoja responded: The available experimental data for the systems of the latest blind test[1] consists of the measured crystal structures (5 polymorphs for system XXIII and one form for the other systems). In addition, slurring experiments indicate that form A of system XXIII is the most stable polymorph at 257 K, while form D is the most stable polymorph at 293 K. Furthermore, all 5 polymorphs of XXIII co-exist under ambient conditions. Therefore, we can assume that these 5 polymorphs must have very similar stabilities.

1 A. M. Reilly *et al.*, *Acta Cryst. B*, 2016, **72**, 439–459.

(304:[304]304) **German Sastre** enquired: My question concerns large molecules (*e.g.* molecule XXVI in your Fig. 1) containing a large number of dihedral angles, where large variations imply very similar energies. What is the balance between intramolecular dispersion contributions and intermolecular dispersion contributions? There are not many comments in your paper regarding the effect of geometry (in particular dihedral angles) on energy. What degree of accuracy in the dispersion is necessary in order to obtain correct dihedrals?

Johannes Hoja replied: In our methods, we do not distinguish between intra- and intermolecular dispersion interactions. We evaluated the quality of our calculated structures by comparing the cell parameters and the root-mean-square deviation to those of the experimentally measured structures. In these metrics, PBE+TS, PBE+MBD and PBE0+MBD yield structures of quite similar quality (see Table 2). All the optimized structures are available in the Supporting Information. We also explored what impact the structure optimization method has on the resulting PBE0+MBD energies of the five XXIII polymorphs (see Figure 5). It can be seen that the optimization method used has only a small impact on the PBE0+MBD relative stabilities; the largest observed change in relative energy amounts to 0.8 kJ mol^{-1}.

The similarity of the structures is illustrated in Fig. 1 for system XXVI. Fig. 1 shows an overlay of the two molecules within the optimized unit cells obtained with PBE+TS (blue), PBE+MBD (red), and PBE0+MBD (green). To put this into perspective for dihedral angles, we compared the dihedral angles defined by atoms 1–4 in the plot. PBE+TS yields a dihedral angle of 51.3 degrees, while the PBE+MBD and PBE0+MBD dihedral angles amount to 52.2 degrees and 51.4 degrees, respectively. For this dihedral angle all three methods agree within 1 degree.

(305:[305]305) **Rui Guo** remarked: I noticed that in the paper, the constructed supercells are completely based on the unit cell parameters. I'd like to add that this practice largely mirrors that routinely used for metals or inorganic materials where the interacting species in the unit cell, metals or ions, have isotropic interactions. For organic molecular crystals, where intermolecular interactions or

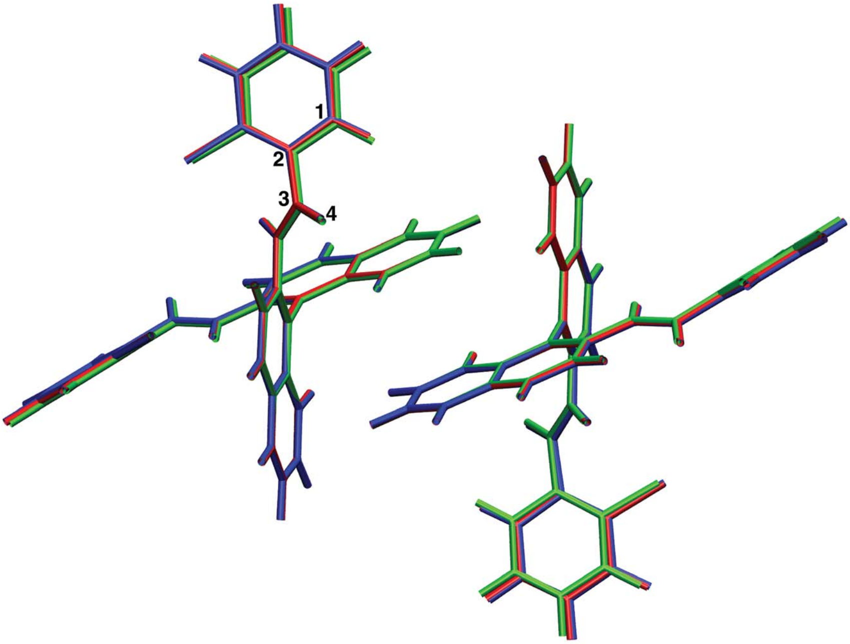

Fig. 1 System XXVI: overlay of the two molecules within the optimized unit cells obtained with PBE+TS (blue), PBE+MBD (red), and PBE0+MBD (green).

even the shapes of the molecules can be very anisotropic, a more informed approach, focusing more on molecular packing and intermolecular interactions, may lead to better computing efficiency. As an example, succinic acid form β has a molecular packing with hydrogen-bonded chains running along the [10–1] direction, and weak CH···O contacts between the chains. I recently calculated its phonon dispersion curves, and found that due to the specific molecular packing, extending the unit cell along the *c* axis for a supercell converges the integrated Helmholtz free energy much faster than extending along the *a* or *b* axes. It can be rationalized that extending the *c* axis preserves the network of weak CH–O contacts in the supercell, while extending the other two axes would have no such effect. A paper on this subject is currently being prepared with Prof. Keith Refson.

Johannes Hoja answered: Utilizing a more informed approach for the generation of supercells is a great idea for reducing the computational cost. In our paper, we employed supercells with minimal lengths of 10 Å in each direction. For a selected structure set we also evaluated the vibrational free energies using larger supercells with a minimal length of 14 Å. Increasing the supercell size changed the relative stabilities on average by 0.3 kJ mol^{-1}, and the maximum observed change amounts to 0.9 kJ mol^{-1}. For the structure showing the largest energy change upon increasing the supercell size, we plotted the phonon dispersion curves in the Supporting Information for the unit cell and the two calculated supercells. It can be seen that supercells of appropriate size are crucial for the correct description of the acoustic modes.

(335:[306]306) **Luca Iuzzolino** asked: With your methodology for optimising and ranking crystal structures, how many crystal structures do you think you could tackle with a computational expense feasible for a practically useful CSP study?

Johannes Hoja replied: I believe that the lattice relaxation with PBE+TS and a subsequent PBE0+MBD energy calculation would be feasible for up to around 100 structures. Vibrational free energies within the harmonic approximation are in my opinion reasonable for tens of structures, while the quasi-harmonic and Morse free energies should be done for a few selected structures.

(336:[307]307) **Luca Iuzzolino** said: The methods you use to calculate the relative energy differences between polymorphs are very accurate but also very computationally expensive. In this study they were applied to a few crystal structures; do you think they would be widely applicable to CSP studies where a large number of putative crystal structures need to be considered? Otherwise you will have to trust the energy ranking produced at the search stage to limit the number of crystal structures that are optimised to a feasible number, and this may lead to a loss of important forms if they are not well-ranked.

Johannes Hoja answered: Our presented approach is intended for the final stage of a crystal structure prediction procedure, assuming an already limited structure set. Therefore, one would need to trust the preceding methods not to miss important forms. However, our method could also function as reference data for the creation of tailor-made force fields or machine-learning models that can be used at the search stage.

(337:[308]308) **Artem R. Oganov** remarked: Last week I saw John Perdew, who was excited about his new SCAN functional. The recipe is to combine it with the rVV10 van der Waals functional when one needs to describe van der Waals systems. What is better – SCAN+rVV10 or PBE(PBE0)+MBD? Or perhaps SCAN+MBD (which I haven't seen used yet)?

Johannes Hoja responded: The performance of several density-functional approximations (including PBE, PBE0, SCAN) when paired with the MBD, VV10, and D3 van der Waals (vdW) models are discussed in ref. 1. SCAN+vdW can provide accurate interaction energies when the damping function used is optimized for the studied system. However, the effective range of the SCAN functional is much more system dependent than that of PBE or PBE0. Therefore, the PBE+vdW and PBE0+vdW methods provide a more universal description since the same range-separation parameter can be used for a variety of systems. Note that in contrast to MBD, VV10 is a pairwise method and provides therefore no description of many-body dispersion interactions.

1 J. Hermann and A. Tkatchenko, *J. Chem. Theory Comput.*, 2018, **14**, 1361–1369.

(338:[309]309) **Artem R. Oganov** enquired: In your work, you used PBE (or PBE0) as exchange–correlation functional, supplemented with many-body dispersion. Other people recommend SCAN exchange–correlation functional, supplemented with rVV10 functional for the description of van der Waals bonding. Which scheme is better? SCAN seems to be better than PBE; is this your impression too? What about the comparison between many-body dispersion and rVV10 functional?

Johannes Hoja replied: A comparison between PBE and SCAN with different van der Waals models (MBD, VV10, D3) can be found in ref. 1 (see the previous question).

1 J. Hermann and A. Tkatchenko, *J. Chem. Theory Comput.*, 2018, **14**, 1361–1369.

(341:[310]310) **Gregory Beran** said: Your results seem to show that as you refine the energies by increasing the sophistication of your computational models, the energy gaps between structures often decrease. We have seen similar behaviors in our own work. Pessimistically, this suggests that even though the field is increasingly able to predict the energetics accurately, that does not necessarily translate to an increased ability to rank/discriminate between structures. It also ties back to Sally Price's comment in the introductory lecture about how the field has transitioned away from the search for *the* unique crystal structure, whose energetic stability would ideally be well-separated from the other candidate structures/polymorphs, to a more nuanced understanding of how complicated crystal energy landscapes can be.

Johannes Hoja answered: I agree that we will often be faced with complicated crystal energy landscapes, where several structures are in a very narrow energy window. Such cases could lead to the identification of new crystal forms or highlight once again the necessity of also including kinetic effects. However, accurate first-principles calculations are a prerequisite for the understanding of such complicated energy landscapes.

(342:[311]311) **Virginia Burger** queried: You mentioned that someone is trying to crystallize the low energy polymorph (structure N70). Did you learn anything from your theoretical analysis that could guide the experimental search?

Johannes Hoja responded: Structure N70 has several similarities with the experimentally obtained Form A and both structures differ mainly by the stacking of the sheets. At the static lattice energy level structure N70 is about 1 kJ mol^{-1} more stable than Form A and is additionally significantly stabilized by vibrational free energies, *i.e.* temperature effects. Therefore, Form A could be dynamically favored over structure N70. We believe that, for example, extremely low melting beginning from Form A could potentially lead to the crystallization of structure N70.

(343:[312]312) **Virginia Burger** remarked: Regarding benchmarking: XtalPi has been testing our CSP work on axitinib, which has around five polymorph crystal structures, many more polymorphs and lots of experimental data, and we have found it to be a more useful benchmark case because of all the experimental data. Finally, you mentioned that simulations could provide more accurate free energy results. We have run such simulations for molecule XXIII from the blind test and have found that the more stable experimental structures have the lowest free energies at room temperature, which is exciting. We are putting together a paper about this now.

Johannes Hoja replied: Indeed, molecular dynamics simulations would provide a better treatment of anharmonic effects than our quasi-harmonic/Morse free energies. However, the accuracy of such an approach will also highly depend on the quality of the underlying energy evaluation method.

(805:[314]314) **Gerit Brandenburg** communicated: During the discussion of the author's excellent contribution, the importance of many-body dispersion effects was mentioned. I would like to clarify that the statements made crucially depend on the definition of the "body", *i.e.* a dipole excitation *vs.* an effective atomic excitation.

Alexandre Tkatchenko communicated in reply: Indeed, the absolute magnitude of the many-body dispersion energy will depend on at least two critical aspects: (1) the definition of the fragment and its polarizability (*i.e.* infinitesimal volume of electron density, localized orbital, or atomic density fragment), and (2) the definition of the coupling potential between such fragments. However, the point we make is rather independent of such considerations. We have demonstrated that MBD (where atomic fragments are employed) provides a quantitative description of the relative energies of molecular crystal polymorphs when combined with non-empirical hybrid exchange–correlation functionals (such as PBE0). Independently of which atom-pairwise model is used to describe vdW dispersion interactions, the correlation between MBD and such effective pairwise models leads to a spread that is larger than 4 kJ mol^{-1}. This is at least 4 times larger than the desired accuracy of 1 kJ mol^{-1}. We have also demonstrated that the convergence of the many-atom expansion within the MBD method is qualitatively different for various polymorphs. We expect these qualitative findings to hold independently of the method employed to describe many-body dispersion interactions.

(808:[315]315) **David McKay** communicated: Section 6 and Table 4 nicely describe the computational costs of the methods involved in the study. PBE0-MBD is the most successful method here, but is 17 times more expensive than PBE-TS. Could this cost increase be elaborated on? I have previously shied away from hybrid-DFT functionals in the solid state, as we tend to see these as prohibitively expensive. I assume there is a cut-off distance for the Fock exchange calculation – is this a parameter that must be converged, and is the cost kept low here with a short distance?

Johannes Hoja communicated in response: Hybrid density functional approximations like PBE0 are significantly more expensive than generalized gradient approximations like PBE because of the necessity to calculate the Hartree–Fock exchange. A detailed description of the periodic Hartree–Fock exchange implementation in FHI-aims utilizing numeric atom-centered basis functions is available in ref. 1. In our calculations we did not use any special settings or cut-off distances for the evaluation of the Hartree–Fock exchange. Note that the relative timings provided in our paper are for a particular molecular crystal of average size, which has 172 atoms in the unit cell. In general, the relative cost will depend on the calculated system and also on the HPC infrastructure used.

1 S. V. Levchenko, X. Ren, J. Wieferink, R. Johanni, P. Rinke, V. Blum and M. Scheffler, *Comput. Phys. Commun.*, 2015, **192**, 60–69.

(817:[316]316) **Volker Deringer** communicated: Towards the end of your paper you've taken the method to the limit of what can be done (in terms of computing cost), and in the conclusions you suggest that machine-learning methods might help with this in the future. There seem to be several possibilities indeed: one could fit potentials to MBD-corrected input data, or "learn" only the MBD correction, or possibly even thermal properties directly. Could you elaborate on this a bit – what do you think are the most promising directions there?

Alexandre Tkatchenko communicated in reply: Indeed, currently the cost of explicit PBE0+MBD calculations plus vibrational free energies is quite high. However, valuable and reliable reference data can already be produced with this methodology for 10–100 crystalline polymorphs of a given molecule. I see two possible approaches that combine machine learning with PBE0+MBD:

(1) Directly learn the forces/Hessians for different polymorphs and reconstruct a "molecule in a crystal" force field from a finite number of polymorphs.

(2) Learn the PBE0 energy (which should be rather local for closed-shell molecules) and learn the MBD Hamiltonian parameters and compute the MBD energy explicitly.

Machine learning ZPE and thermal properties should also be rather straightforward. We have demonstrated that for gas-phase molecules this can be done with quantitative accuracy.[1] Applying such models to crystals requires some non-trivial extension, but this should doable.

1 W. Pronobis, A. Tkatchenko and K.-R. Müller, *J. Chem. Theory Comput.*, 2018, **14**, 2991–3003.

(821:[317]317) **Seiji Tsuzuki** communicated: You showed that the contributions of the many-body dispersion interactions and vibrational energies to the lattice energies of crystals are not negligible. Do the contributions of many-body dispersion and vibrational energies have a significant impact on the relative stability of polymorphs, or are the contributions of these terms to the relative stability not significant due to cancellation?

Alexandre Tkatchenko communicated in response: As our paper shows, both the many-body dispersion and vibrational free energies are crucial for obtaining the correct description of free energy differences between polymorphs. The many-body energy decomposition shows that many-body effects vary significantly for different polymorphs. In the case of vibrational free energies, the low-frequency phonons depend heavily on the crystal structure and hence lead to substantially different entropy contributions for different polymorphs.

(823:[318]318) **David Bowskill** communicated: This approach determines the geometry of organic crystals using the PBE functional and then applies the MBD correction, followed by a proportion of exact exchange using the PBE0 functional. By applying these corrections, the static crystal structure and associated energy no longer reside at local minima. Do you have any evidence of how much this will affect the predicted crystal energies? For example, would full geometry optimisation using the PBE0 functional with MBD correction produce noticeably different results? In addition, how much does this affect the harmonic approximation, which assumes that the molecules are positioned at minima in the crystal energy landscape?

Johannes Hoja communicated in reply: In our paper we studied how the PBE0+MBD energies depend on the optimization method used (see Figure 5). It can be seen that the structure optimization method has only a small effect on the relative PBE0+MBD stabilities, with a maximum change of 0.8 kJ mol^{-1} in the relative stabilities. The structures used for the vibrational free energy calculations in the harmonic approximation were always optimized with exactly the same method as is used for the harmonic approximation to ensure that we are in a local minimum.

(307:[319]319) **Marcus Neumann** opened the discussion of the paper by Jan Gerit Brandenburg: What is the root mean square deviation between the lattice energy differences calculated with DFTB3-D3 and higher level methods for the compounds that you have studied?

Gerit Brandenburg answered: The root mean square (rms) deviation was calculated between the energies of the crystal structures after DFTB3-D3 and the final optimisations. The distributed multipoles and intramolecular energy were calculated with an improved charge distribution (PBE0 6-31G(d,p)). The FIT repulsion–dispersion potential was calculated relative to the global minimum structure in E_{latt} after the final optimisations (see Table 4). The rms values were calculated for all the CSP-generated crystal structures that were found to be within 50 kJ mol^{-1} after the final optimisations, as well as for a more limited set of crystal structures that matched the most important putative polymorphs found in the original CSP studies (see the Electronic Supplementary Information); this set included all the experimentally known forms with $Z' = 1$. The results in Table 4 confirm the conclusion of the paper that we cannot rely on DFTB3-D3 lattice energies.

Table 4 Root mean square (rms) deviations calculated between the energies of crystal structures

Molecule	$\sim\text{rms}_{\Delta E_{\text{latt}}}$ for all crystal structures/kJ mol^{-1}	$\sim\text{rms}_{\Delta E^{\text{latt}}}$ for crystal structures matching the most important PPMs/kJ mol^{-1}
XXVI	21	13
GSK269984B	34	21
XX	15	5
XXIII	21	10
Mebendazole	15	7
Overall	21	13

Related to this is a comparison with periodic DFT-D methods for selected crystal structures from the 6th crystal structure prediction blind test (POLY59 set[1]). The absolute lattice energies (X23 set) mean unsigned errors of DFTB3-D3 are competitive to PBE-TS, but for relative lattice energies, periodic dispersion corrected density functional (DFT-D) profits a lot from error compensation, which does not hold for the tight-binding Hamiltonian.

1 Brandenburg *et al.*, *Acta Cryst. B*, 2016, **72**, 502–513.

(308:[320]320) **Christian Schön** enquired: While looking like a nice compromise between *ab initio* calculations and empirical potentials, there is one big problem with DFTB: up to now, not enough pairwise tight-binding parameters are available. According to the practitioners, it takes $\frac{1}{2}$–1 year to generate a new set of parameters for your problem of interest, and you need such parameters for every pair of atom types (n atoms → $n(n+1)/2$ pairs). What is the story on this at the moment? We have found some parameters for organic molecules in the literature, but it would be incredibly helpful if there were also such parameters *e.g.* for interactions between metals and organic molecules.

Gerit Brandenburg replied: There are different groups working on DFTB parametrization. While we used the 3OB parameter set, a current list is available.[1] A different strategy from the group of Stefan Grimme is the development of a tight-binding Hamiltonian that relies on element parameters only and can thus be parameterized much more rapidly. Elements up to Rn are available.[2]

1 http://www.dftb.org/parameters/download
2 Grimme *et al.*, *J. Chem. Theory Comput.*, 2017, **13**, 1989.

(309:[321]321) **Matthew Addicoat** asked: Following up on parameters, you may have been referring to the QUASINANO parameters, which are single-element parameters up to element 90 (Th).[1,2] Did you try using the MIO parameters at all? I've had significant experience in both COFs and ionic liquids where we used the MIO parameters with Lennard-Jones dispersion – in one paper a reviewer did ask us to repeat our calculations with 3OB – we did so and got poorer results.[3]

In your paper, you did significant work on re-parameterising the dispersion component, but didn't add any extra hydrogen bonding or halogen corrections. Three of your molecules do have fluorine, chlorine or both. Did you see any problems or poor behaviour with those molecules?

1 M. Wahiduzzaman, A. F. Oliveira, P. Philipsen, L. Zhechkov, E. van Lenthe, H. A. Witek and T. Heine, *J. Chem. Theory Comput.*, 2013, **9**, 4006–4017.
2 A. F. Oliveira, P. Philipsen and T. Heine, *J. Chem. Theory Comput.*, 2015, **11**, 5209–18.
3 M. A. Addicoat, R. Stefanovic, G. B. Webber, R. Atkin and A. J. Page, *J. Chem. Theory Comput.*, 2014, **10**, 4633–4643.

Gerit Brandenburg answered: The QUASINANO set mentioned is part of the effort to generate a DFTB parametrization for the whole periodic table. However, it is still necessary to combine them with pair-parameters and those are only available up to Ca.[1] A more extensive full parametrization is provided by the GFN-xTB variant.[2]

A systematic comparison of QUASINANO, MIO, and 3OB parameter sets for DFTB3 combined with the D3 dispersion correction has not been published. All our initial tests on established benchmark sets for molecular dimers and crystals (S66, L7, S12L, X23 test sets) have shown that the use of 3OB parameters is better suited to describe the noncovalent interactions of mostly organic molecules. Our optimized damping parameter sets with corresponding statistical evaluation have been distributed in a private communication to ADF, where the QUASINANO set is used, and will be provided by us upon request. A closely related test has recently been done by combining DFTB3(3OB) and DFTB3(MIO) with the TS and MBD dispersion correction models. In the supporting information of ref. 3, we directly compare the intermolecular interactions and distances of molecular dimers (S66 set) obtained from different parametrizations (see Figure S7). In agreement with our previous experience, the 3OB parameters are recommended for this system class.

Hydrogen- and halogen-bonded molecules can be tougher to describe with a simple point-charge electrostatic model. Ref. 71-73 in our paper try to correct some of the systematic errors with additional pair-potentials. However, we found that these additional corrections do improve the interaction energies in the reference sets, but lead to only slight improvements in larger complexes.[4] Thus, we did not use these in our paper (DOI: 10.1039/C8FD00010G) and could not detect major problems with the target molecules including halogens. However we did not compare the energetic ranking in detail, as this was not sufficiently accurate for any of our targets, and we recommend a subsequent energy evaluation with a more accurate method.

1 Oliveira *et al.*, *J. Chem. Theory Comput.*, 2015, **11**, 5209.
2 Grimme *et al.*, *J. Chem. Theory Comput.*, 2017, **13**, 1989.
3 Mortazavi *et al.*, *J. Phys. Chem. Lett.*, 2018, **9**, 399.
4 Brandenburg *et al.*, *J. Phys. Chem. Lett.*, 2014, **5**, 4275.

(310:[322]322) **Christopher Taylor** enquired: You mentioned that you obtained significantly improved results when performing a single-point calculation and using an electrostatic multipole and empirical force field interaction model (*i.e.* using the DMACRYS program). Can you clarify this – do you mean that you ran only a single-point calculation in DMACRYS (*i.e.* the resulting structures are the same as those from your DF-TB calculations, and not an energy minimum at the multipole-and-force-field level of theory)? Or are you referring to the single-point *ab initio* calculation required to obtain said multipoles (*i.e.* the structures are fully minimised in DMACRYS after having been generated through DF-TB)?

Gerit Brandenburg replied: We did indeed follow the latter strategy, *i.e.* we used the rigid DFTB3-D3 molecular geometries to compute the distributed charge multipoles. Subsequently the DFTB3-D3 crystal geometry was relaxed with the DMACRYS code by combining the multipole electrostatic with the empirical FIT potential. This relaxation of the intermolecular degrees of freedom does not notably increase the computational cost and we thus recommend including this step. However, as a more general strategy, we have shown that even the intermolecular geometries of DFTB3-D3 (or DFTB3-MBD) can be used directly for computed density functional based single-point energies.[1]

1 Mortazavi *et al.*, *J. Phys. Chem. Lett.*, 2018, **9**, 399.

(311:[323]323) **Seiji Tsuzuki** said: I have two questions on the accuracy of DFT-B. The first question concerns the accuracy of the charge distributions obtained by the DFT-B method. The electrostatic interactions are important in determining the arrangement of molecules in crystals, since the electrostatic interaction is a highly directional interaction. It can be attractive or repulsive, depending on the orientation of the molecules. Therefore, accurate evaluation of the electrostatic interactions is important for crystal structure prediction. The accuracy of the calculated charge distributions can be judged by comparing the calculated dipole moment with the experimental one. *Ab initio* calculations show that a split valence basis set including polarization functions and electron correlation corrections are necessary for accurate calculation of the charge distributions. On the other hand, tight-binding calculations are semi-empirical calculations and use very small basis sets. I suspect that the DFT-B calculations cannot reproduce the experimental dipole moment well. Did you compare the dipole moments of molecules calculated by DFT-B with experimental values? I would like to know how large the errors of the calculated dipole moments are.

The second question concerns the accuracy of the calculated torsional potential by DFT-B. To reproduce the small energy difference between rotamers by *ab initio* calculations, a split valence basis set with polarization functions and electron correlation corrections are necessary. Is it possible to reproduce the experimental torsional potential by DFT-B? Have you compared the torsional potentials calculated by DFT-B with those obtained by high level *ab initio* calculations?

Gerit Brandenburg responded: The dipole moments are indeed a way to test the electrostatic interactions from DFTB3-D3. While similar studies have been conducted before, we have computed the dipole moment on a set of small semi-rigid molecules with experimental references; experimental values gathered in ref. 1. While DFT methods in converged basis set expansions (augmented triple-zeta or larger) indeed produce higher quality dipole moments, the results from DFTB3-D3 are reasonably good and actually better than those from a very recent tight-binding parametrization, GFN-xTB.[2] Correlations with the reference data (Pearson correlation and mean unsigned errors) are shown in Fig. 2. However, semi-empirical methods rely strongly on error compensations, *i.e.* the true test should be the total interaction energy that has been tested for molecular crystals in ref. 3.

It is correct that torsion potentials are crucial to describe the relative stabilities of molecular conformers. Individual torsions have sometimes been compared directly to high-level references. However, this is not trivial as any benchmark of full torsion potentials has to be based on multi-reference wavefunction methods. Thus, it is easier and more instructive to directly compare the relative stability of local equilibrium structures, *i.e.* molecular conformers. We have not done this for the target molecules in our paper (DOI: 10.1039/C8FD00010G), but in ref. 4 several established conformational benchmark sets were analyzed with different tight-binding approaches (including DFTB3-D3), showing reasonably good performance.

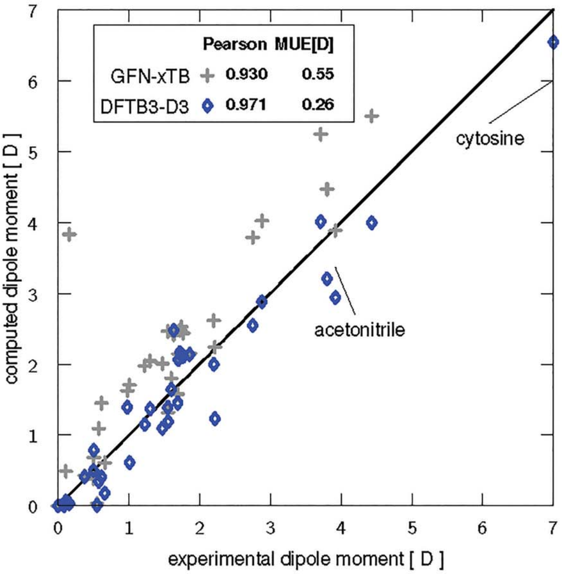

Fig. 2 Correlation with the reference data (Pearson correlation and mean unsigned errors).

1 Hickey *et al.*, *J. Phys. Chem. A*, 2014, **118**, 3678.
2 Grimme *et al.*, *J. Chem. Theory Comput.*, 2017, **13**, 1989.
3 Brandenburg *et al.*, *J. Phys. Chem. Lett.*, 2014, **5**, 1785.
4 S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989.

(312:[324]324) **Gregory Beran** asked: You noted that DFTB3-D3 provides useful predictions for the free energy contributions. Does it reproduce each individual phonon mode frequency well, or is it a case where good error cancellation occurs when summing over all the vibrational modes when computing free energies?

Gerit Brandenburg answered: In our paper (DOI: 10.1039/C8FD00010G), we computed free energy contributions based on the DFTB3-D3 model and contrasted it with rigid-body models. Some comparisons to higher-level DFT results and experimental measurements have been conducted.[1,2] However, our analysis is indeed a rather coarse-grained one, as only integrated quantities (like free energy contributions) are analyzed. In this regard, Prof. Beran is correct and we do rely on compensation of errors.

In earlier studies, the covalent modes from DFTB have been analyzed.[3] While the authors recommended a scaling factor, it is not clear how this would translate to soft non-covalent modes. Clearly, further detailed analysis for molecular crystals is needed in order to make faithful predictions. We would emphasize that methods need to be tested and validated for flexible pharmaceutical crystal structures to test the description of the coupling between the molecular and intermolecular modes, which seems likely have a significant effect on the relative free energies.

1 Brandenburg *et al.*, *J. Phys. Chem. Lett.*, 2017, **8**, 4319.

2 Buchholz *et al.*, *Cryst. Growth Des.*, 2017, **17**, 4676.
3 Witek *et al.*, *J. Comput. Chem.*, 2004, **25**, 1858.

(313:[325]325) **Marcus Neumann** remarked: Some years ago we introduced phonon free energies in GRACE based on our tailor-made force fields, and found out that those free energies actually had a negative impact on our energy rankings. Phonon free energies strongly depend on the vibrational frequencies of the soft modes, and small errors in the force calcuations result in large errors in the phonon free energies. The energy ranking performance of DFTB3-D3 does not seem to be better than the performance of tailor-made force fields, so I would expect the same for its use in phonon free energy calculations. What evidence do you have that phonon free energies calculated with DFTB3-D3 actually improve the energy ranking performance?

Gerit Brandenburg replied: DFTB3-D3 based lattice energies are not satisfactory for a faithful energy ranking and might not outperform the tailor-made force field within GRACE. On the other hand, the equilibrium geometries seem to be of good quality, as shown by the small deviations w.r.t. experimental crystal structures. How this translates into the properties of second derivatives is not clear *a priori*, especially as the low-energy modes are most important for the entropy estimates. From the ranking of our computed energy landscapes, we have some indication of an improved ranking, *e.g.* for target XXVI the experimental structure is rank 2 based on lattice energies and moves to rank 1 upon inclusion of the DFTB3-D3 based free energy contributions. A direct comparison of free energy contributions from DFTB3-D3 with DFT based references has been recently done on a carbamazepine polymorph.[1] This has to be extended to broader tests. The contributions of paper 11895 (DOI: 10.1039/C8FD00048D) and paper 11896 (DOI: 10.1039/C8FD00066B) might be ideally suited to conduct this analysis.

1 Brandenburg *et al.*, *J. Phys. Chem. Lett.*, 2017, **8**, 4319–4324.

(315:[326]326) **Johannes Hoja** asked: You mention in your paper that tighter convergence criteria had to be used in order to remove imaginary frequencies in the DFTB3-D3 phonon calculations. What were the tightest convergence settings necessary in order to avoid imaginary frequencies?

Gerit Brandenburg responded: The “tightened” optimization criteria correspond to the standard thresholds of CRYSTAL17, *i.e.* maximal gradient component of 0.00045 a.u., rms gradient of 0.0003, maximal displacement of 0.0018 a.u., and rms displacement of 0.0012 a.u. Additionally, the components of the dynamical matrix are computed by symmetric finite displacements (in contrast to the default of just one displacement).

(317:[327]327) **Ivo Rietveld** opened the discussion of the paper by Claire S. Adjiman: The potentials in the calculations of the crystal structures are adjusted using sublimation enthalpy data. How do you guarantee that the sublimation enthalpy data is accurate and not affected by impurities, and that the data is for the correct polymorph? How is the precision of the sublimation data dealt with, in particular when the vapor pressure is low? Sublimation data cannot be used for new polymorphs that are obtained purely by prediction. How is the available sublimation data generalised within the code to improve the stability prediction of unknown polymorphs?

Claire Adjiman answered: We have carefully selected data from the literature, paying attention to the experimental methods used, but ultimately we cannot guarantee the accuracy of all data points. The error in the data can be taken into account by adjusting the weight given to each data point (*e.g.* in the maximum likelihood approach, each data point is weighted based on the corresponding variance). We investigated the impact of such weights on the parameters for the hydrocarbons (*c.f.* Fig. 3 in the paper) and found that using equal weights for the structural data and for the sublimation enthalpies led to parameters which gave a good structure reproduction relative to what is achieved with a weight of 0 for sublimation enthalpies, and a good quality of reproduction of the sublimation enthalpies. We subsequently used equal weights. We do not use any information on predicted or unknown polymorphs for parameter estimation. In our data set, we have only one molecule with two polymorphs, tetracyanoethylene, and we ensured that the parameters yielded the correct energy ranking of these two polymorphs. In the future, we would like to include information on relative energies in the objective function.

(318:[328]328) **Jonas Nyman** asked: Have you benchmarked your new force field against the currently very popular X23 benchmark? Are some of the structures in the X23 set in your training and/or validation sets? Also, have the parametrizations been automated in such a way that you could fit other functional forms, such as a Lennard-Jones potential, or a damped Buckingham with additional higher-order C8/C10 dispersion terms?

Claire Adjiman replied: We have not benchmarked our force field against the X23 data set. The following X23 structures are included in our training set: 1,4-cyclohexanedione (CYHEXO), acetic acid (ACETAC07), adamantane (ADAMAN08), anthracene (ANTCEN16), benzene (BENZEN06), imidazole (IMAZOL06), naphthalene (NAPHTA31), pyrazine (PYRAZI01), succinic acid (SUCANH14), triazine (TRIZIN01), and trioxane (TROXAN). The parameter estimation code currently does not allow the direct use of alternative potentials. However, it can be modified relatively easily to work for the Lennard-Jones potential or a higher-order Buckingham potential.

(319:[329]329) **Sharmarke Mohamed** remarked: The work presented in your paper is a welcome contribution for practitioners of crystal structure prediction, who for a long time have used model intermolecular potentials that have been fitted to crystal structures that are often not representative of the systems under study. In work conducted within my group, I have found that the FIT force field actually performs well in modelling the intermolecular interactions in two-component molecular salt systems. However this force field is not appropriate for three-component molecular systems (*e.g.* cocrystal hydrates or ionic cocrystals). The *ab initio* model used to fit the potential parameters is an important factor in this and so it is very encouraging that the authors have developed and tested a method for deriving potential parameters that can be tailored to different *ab initio* models. I am curious, however, whether the authors feel that the lack of complex molecular systems (*e.g.* hydrates, cocrystals, salts) in their dataset of 106 crystal structures will present challenges in the use of these potential parameters for studies of multicomponent systems. Have the authors tested the transferability of these new parameter sets to multicomponent systems that display competing intermolecular interactions? Could the authors also comment on the typical cost in time (*i.e.* how many weeks?) that would need to be spent on using the method proposed in this paper to derive system-specific parameters (*i.e.* potential parameters for salt hydrates) using a specific *ab initio* model?

Claire Adjiman responded: Thank you for the positive comments. We have not yet tested the proposed parameters for such multicomponent systems. In our view, the dataset needs to be expanded to include hydrates and salts, as well as flexible molecules, to increase the transferability. We are currently working in this direction. The main cost in deriving parameters is to gather reliable data on crystal structures with sufficient diversity. The computations are highly parallelisable and their cost depends principally on the number of data points in the training set. Using one core, the regression of the parameters corresponding to one box in Fig. 2 of the paper requires approximately 150 days of wall time for a single core. This is equivalent to about 20 h per local minimisation, and all minimisations are independent of each other.

(320:[330]330) **Sten Nilsson Lill** asked: Your choice of functional (M06) is very sensitive to the numerical grid accuracy. Did you apply a tight grid in your calculations to ensure you had a stable final solution?

Claire Adjiman answered: Yes, we used a very tight grid for our calculations.

(321:[331]331) **Seiji Tsuzuki** enquired: I have a question about the exchange–repulsion potential in your force field. You showed that the performance of the calculation of sublimation energy is improved by the improvement of the force field, but the calculated sublimation energies still have large errors in some cases. It is well known that the exchange–replusion potentials of hydrogen and halogen atoms have strong anisotropy. The importance of the anisotropy of hydrogen atoms in reproducing the crystal packing of hydrocarbon molecules by force field calculations was suggested by Williams in the 1970s. I believe that it is necessary to consider the anisotropy of exchange–repulsion interactions to improve the performance of the calculation of sublimation energies. Why don't you use anisotropic exchange–repulsion potentials for the calculation of the sublimation energies?

Claire Adjiman responded: I agree that it is often important to include anisotropy in the calculations. We have designed our parameter estimation code so that the parameters can easily be re-estimated as one makes changes to the lattice energy model, either in the *ab initio* components of the model (intramolecular energy and electrostatics) or in the form of the repulsion–dispersion potential. In this paper, we focused on re-parameterising the most common potential, the isotropic Buckingham, for one specific level of theory for the DFT calculations (M06/6-31G(d,p)). In the future, we intend to re-use the methodology and code to develop anisotropic potentials.

(322:[332]332) **Luca Iuzzolino** said: You have shown that the new repulsion–dispersion parameters improve the absolute energies of the crystal structures that were considered in this study. Is this also true for the relative energy differences between polymorphic forms of the same molecules?

Claire Adjiman replied: The relative stability of TCNE polymorphs in the training set is correctly reproduced. We have not yet tested other polymorphic systems.

(324:[333]333) **Sarah Price** remarked: It is about time that something was done to improve the FIT potential. In 1996, David Coombes fitted a polar hydrogen potential to allow modelling of some hydrogen-bonded structures when used with in combination with Williams' old parameters from papers from the 1980s. Williams had fitted his parameters to few crystal structures and very few sublimation energies. The fluorine parameters have given some problems, probably because of the sublimation energy used in derivation, but generally it is surprising how many hundreds of CSP studies we have been able to do with the FIT potential, given how it was derived. I look forward to trying your new sets of parameters, with distributed multipoles for a good quality of charge density. However, I notice that you haven't got many cross-terms for two heteroatoms, such as nitrogen–oxygen, which we would need for amide hydrogen bonding. What are the difficulties? Would you advise trying combination rules for the missing parameters?

Claire Adjiman answered: Thank you for the positive comments. In the first instance, I would indeed advise using standard combining rules. The main limiting step is the creation of the training and validation sets, curating the data to ensure high reliability and adequate coverage of different types of structures. We are currently working to expand the set of parameters published here to include fluorine as well as heteroatom cross-interactions.

(800:[334]334) **Alan Hare** asked: Following up on the question of intramolecular smoothing: is the discontinuity in the potential true or apparent? While sometimes there can be an abrupt separation between two proximate, distinct functions, in others – like melting "squares of chocolate", as the poster reminded me! – a single, continuous function may really straddle the divide, or the "break between squares"… which can itself sometimes be transformed into a simpler function. On reflection, am I right in thinking that in the intramolecular case – with its multiplicity of rotations – the discontinuity is truly an abrupt one? If so, I imagine that smoothing across it may involve the use of a hyperbolic tangent (or possibly sigmoidal) function close to the "cut-over", so to speak. Or do you have another way of handling it altogether?

Claire Adjiman responded: This relates to the poster presented by Isaac Sugden. The functions are discontinuous because that there are two different values for the intramolecular energy at the boundary between two local approximate models (LAM points). This is because in the early versions of this approach each LAM has a well-defined region of use and we switch abruptly between LAMs as we cross boundaries. I have sketched this roughly to help visualise what happens (see Fig. 3). In an intuitive sense, how abrupt the discontinuity is depends on how different the two function values are. Regardless, the discontinuities are always expected to be present. These discontinuities are not physical – they are artefacts of the local models used. In this work, we corrected this by replacing the use of a single LAM by an average of all the LAMs. The average is weighted so that the LAM nearest to the current point of interest carries the most weight. We compute the weight of each LAM so that it decays exponentially as the distance between the point of interest and the LAM point increases.

(804:[335]335) **Gerit Brandenburg** communicated: The authors made a nice effort in parameterizing the FIT potential for use in organic molecular crystal structure prediction containing the elements H, C, N, O and Cl. While hydrogen has parameters that depend on its binding motif, other elements have a single set of parameters. Considering the London dispersion interaction, this might be a crude approximation, especially for carbon, which has quite a large variety in its dynamic polarizability. Fig. 4 shows computed (TD-DFT) dispersion coefficients for carbon in different environments, which are then used as references in the D3 dispersion correction scheme.[1] As the range of C_6^{CC} is from 20 to about 50 a.u., this might be relevant for the FIT type potential. The D3 coefficients are

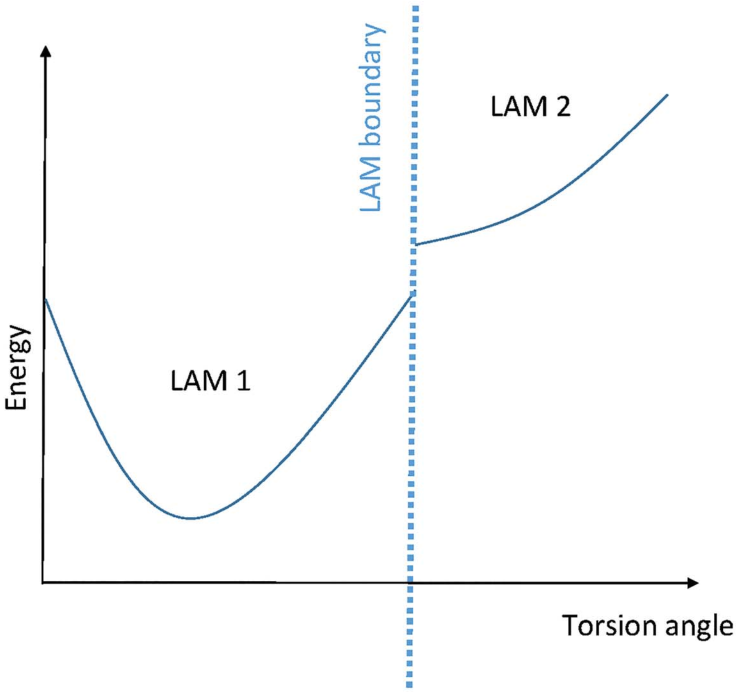

Fig. 3 Abrupt switching between LAMs at LAM boundary.

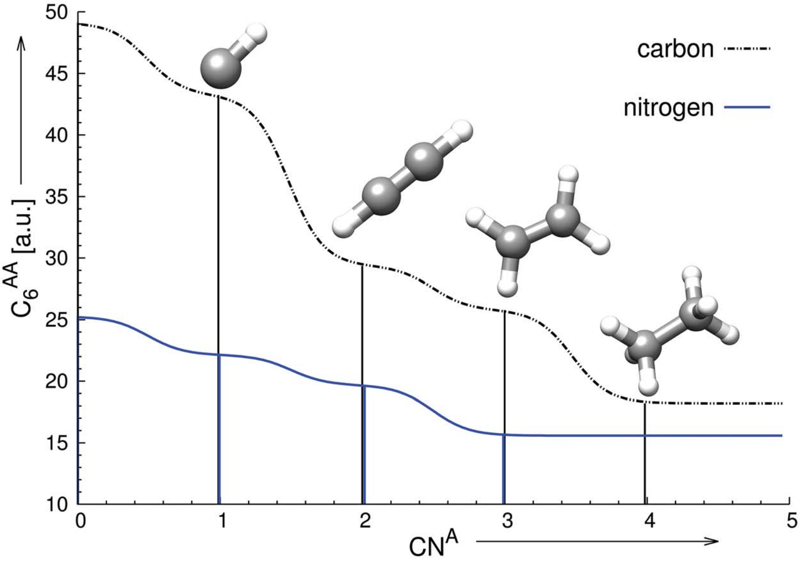

Fig. 4 CN dependent atom pairwise dispersion coefficient for homoatomic pairs C_6^{AA} as a function of the coordination number CN_A. The reference points $C_{6,ref}^{AA}$ are drawn as vertical sticks in the respective colors (black for carbon, blue for nitrogen). Figure reproduced with permission from ref. 3.

computationally easily accessible and yield high-quality inter-molecular dispersion coefficients (see Fig. 2 in ref. 2). Are the authors considering using environment dependent parameters for elements other than H?

1 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
2 E. Caldeweyher and J. G. Brandenburg, *J. Phys.: Condens. Matter*, 2018, **30**, 213001.
3 S. Grimme, A. Hansen, J. G. Brandenburg and C. Bannwarth, *Chem. Rev.*, 2016, **116**, 5105.

Claire Adjiman communicated in reply: It is indeed an approximation not to take into account the environment of other elements. We strive for a balance between the accuracy of the model and the statistical significance of the parameter values. As we increase the number of atom types and hence the number of parameters, we are likely to achieve greater accuracy, at the cost of decreased statistical significance as data scarcity is an issue. The set of atom types chosen has so far provided a good balance between accuracy and significance, with enough data available in most cases for training and validation of the models. We note that the parameters are effective values rather than a reflection of intrinsic repulsive–dispersive interactions: as such, we include multiple atom environments in the data set for training and obtain parameters that provide an average interaction. In addition, as discussed in the paper, the repulsion–dispersion term in our model of lattice energy includes all the aspects that are not taken into account in the model, such as entropic contributions or polarizability. This is very different in concept to the D3 coefficients, for which the objective is to obtain an intrinsic value of the dispersive interactions only. Thus, the proposed repulsion–dispersion potential contains an empirical correction of the potential. One could in principle modify our model to use D3 coefficients and introduce a different empirical correction.

(325:[336]336) **Marcus Neumann** opened a general discussion of the papers by Alexandre Tkatchenko, Jan Gerit Brandenburg and Claire S. Adjiman: Calculating the relative stabilities of crystal polymorphs is at the very heart of organic crystal structure prediction. Yet very little reliable information, especially for larger molecules, is available in the scientific literature. Experimental error bars are missing in most cases. If we really want to make progress in the energy ranking accuracy, a reliable experimental energy benchmark must be established. Ideally, 20–100 polymorphic enthalpy and free energy differences should be measured, each by different methods and different researchers.

Having talked to several experimentalists in academia about the issue, it seems that getting funding for such an endeavour is very difficult because accurate thermodynamic measurements by themselves are not considered innovative. As a community, we should try to convince a national or transnational funding agency to support the accurate measurement of polymorphic energy differences.

Claire Adjiman responded: I concur that it is important to generate such a benchmark data set. This could be done through collaborative projects where one student works across experimental and computational groups, measuring some data and assessing/improving the state-of-the-art in modelling. This would provide a coherent scientific problem and expose the students to different tools/learnings, and by working in tandem such projects may be more compelling for both researchers and funders. Of course one can imagine doing this across a larger consortium of collaborators, where we might explore a broader range of properties. ITN or COST would be good ways forward. There is also the example of the UNIFAC consortium, which has been obtaining data and developing models for fluid phase equilibria for many years, first in academia and now as an independent company.[1]

[1] http://unifac.ddbst.de/unifac_.html

Ivo Rietveld responded: As an experimentalist, I would like to add a few points. First of all, it is hard to give a PhD student such a subject, because he or she should at least have to deal with a scientific problem that will lead to a proper scientific thesis (and answer…?). Just data as such will not be enough, but in the end we also look for precise thermodynamic rankings, so it is a matter of matching the right substances with the right scientific questions. That will remain a sort of case-by-case issue.

Second of all, precision is difficult, also experimentally. The experimentalists may claim small error bars, and certainly with calorimetry and X-ray diffraction we are good at determining differences, as long as we do it on the same equipment(!), but again everything depends on the system, and whether the chemical compound can be measured with the precision that is required for computer calculations. Therefore, from an experimental point of view, the database is probably easier to define not in terms of substances, but in terms of transition values of which we are very confident. For ROY, for example (with which I am in fact not very familiar), there may be phases that are extremely well documented, whereas others are not because they are an experimental nightmare.

I think therefore that this question may be resolved by some sort of COST (or other!) network, where the uptake into the database of experimental systems that we are studying for other reasons anyway can be discussed. A decision to take up a substance or transition in the database within this COST network will initiate additional measurements by other experimentalists and calculations by computationists to come to consolidated transition values between very specific polymorphs. This latter part will likely need extra funding, but we may be able to leverage funding within such a COST network. This COST network should also function as quality control of the experimental values and computational values.

For the precise measurements, we may be able to apply for grants that emphasise the industrial development of scientific findings, particularly if all the large European pharmaceutical companies can be involved in the grant application. Some validation work may be done by a number of post-docs and papers may be published in the same vein as the polymorph prediction blind tests. Why not discuss the experimental and computational caveats and how we interpret, say, the experimental and computational results of 10 polymorphic transitions towards a rigorous database? More importantly: who is in?

Alexandre Tkatchenko responded: I strongly concur with Marcus and others as well. Perhaps the effort of encouraging our experimental colleagues would have a higher chance of success if all the participants (theorists, experimentalists and industry) show strong interest and push each other. For a start, it would be good if industry decided to disclose existing thermodynamic data for interesting polymorphic systems. I am sure this data is lying around in different companies. High-level quantum-chemical calculations are also very relevant and should be done. One current problem with non-canonical CCSD(T) and QMC is that accuracy is very hard to control to the required level of 1 kJ mol^{-1}, especially for larger molecules. So far, the accuracy of these methods in periodic calculations has been assessed by comparing to experiments, and never for energy differences between polymorphs. In this case, glycine, oxalic acid and coumarin (mentioned by Qiang Zhu) would be good systems to which to start applying QMC/CCSD(T). Regarding challenging systems, axitinib and ROY are good ones going forward. However, in the case of ROY, my understanding is that it is hard to ensure that experiments measure pristine and non-defect structures.

Christian Schön responded: I strongly concur with Marcus Neumann. I have tried for many years to convince my experimental solid chemistry colleagues to perform systematic analyses of their syntheses of well-defined systems, producing quantitative outcomes as function of synthesis parameters, in order to provide data to sharpen our theoretical tools for modeling chemical syntheses. But in the end, nobody was willing to give a student this endeavor – I guess it is not "sexy" enough (no new compounds or crystal structures being produced, *etc.*) and it might be too difficult to publish such a study. (I have heard similar complaints from people in the phase diagram community – they have a hard time getting funding agencies interested unless the chemical system has important applications.) I can see that a similar problem would appear in the accurate measurement of free energy differences, and I would strongly support such a measurement campaign.

Qiang Zhu responded: I agree. I think the first step is to create a good polymorph data set that can present the real challenges in energy ranking. Currently, the X23 set has been studied a lot,[1,2] but it seems to me that many methods produce nice results at about the same level. Do they really reflect the challenge in energy ranking? I have two systems which strongly oppose this conclusion – coumarin[3] and ROY (DOI: 10.1039/C8FD00039E). These might be very small molecules by Marcus Neumann's standards, but most methods still fail to reproduce the ranking nicely. As a community, we need to come up with a plan to create a list of challenges and then work it out. This requires a synergy between experiment and theory.

1 A. Otero-de-la-Roza and E. R. A. Johnson, *J. Chem. Phys.*, 2012, **137**, 054103.
2 A. M. Reilly and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2013, **4**, 1028.
3 A. G. Shtukenberg, Q. Zhu, D. J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk, A. Tkatchenko, A. R. Oganov, A. L. Rohl, M. E. Tuckermanejk and B. Kahr, *Chem. Sci.*, 2017, **8**, 4926.

(326:[337]337) **Gerit Brandenburg** remarked: The electronic structure community has made some substantial developments in the past years, with CCSD(T) being applicable to large and sometimes periodic systems (mostly relying on embedding or localization schemes). We have contributed to this endeavour by extending the applicability of Quantum Monte Carlo to organic crystals.[1] This approach can potentially be used to compute the static lattice energies of large molecular crystals and their polymorphs ($\sim$200 atoms per unit cell) in a reasonable time frame.

This could ideally be complemented by experimental relative stabilities to independently test the capabilities of CSP methods to predict relative lattice energies (*vs.* high-level theory) and relative free energy differences (*vs.* experiment). This is closely connected to Marcus Neumann's previous question and similarly requires funding to afford the substantial CPU time.

As an add-on concerning the CPU time: the computation of a lattice energy for benzene and anthracene crystals with sub-chemical accuracy (regarding the thresholds we can control) requires 6000 and 74 000 CPU hours. The time will clearly depend on the system size (roughly cubic scaling with the number of atoms), the target accuracy (*e.g.* reducing it from 0.5 kJ mol^{-1} to 0.1 kJ mol^{-1} increases the CPU time by an order of magnitude), and the employed time step (doubling it halves the CPU time), which varies a lot with the target molecule. Once a decision on target crystal structures has been made, more quantitative estimates are possible.

1 A. Zen, J. G. Brandenburg, J. Klimeš, A. Tkatchenko, D. Alfè and A. Michaelides, *Proc. Natl. Acad. Sci. USA*, 2018, **115**, 1724–1729.

Matthew Ryder replied: I agree that this approach could yield valuable insights into the capabilities of different CSP methods to predict lattice energies.

(328:[338]338) **Christian Schön** addressed Johannes Hoja and Gerit Brandenburg: It looks like we are getting good results on the energy evaluation front for molecules and molecular crystals. But what about the quality/computational expense of your methods for applications to molecules on surfaces? Can you handle this as "just another application" of your methods, or would one still need years of new developments/adaptations? What about the size of the (flexible) molecules – do you envision computing proteins on this level of accuracy?

Johannes Hoja answered: Our methods can be and have already been used to study the adsorption of molecules on surfaces (chemisorption, physisorption, and intermediate situations). Based on the Tkatchenko–Scheffler dispersion model, the vdW^{surf} method was developed to also incorporate screening effects from the substrate electrons.[1] Examples of applications of the DFT+vdW^{surf} and DFT+MBD methods can be found in ref. 2–5. These calculations provide accurate adsorption energies and geometries, and can readily be carried out on state-of-the-art computing facilities. Several DFT+MBD single-point energies could be obtained for isolated small proteins.

1 V. G. Ruiz, W. Liu, E. Zojer, M. Scheffler and A. Tkatchenko, *Phys. Rev. Lett.*, **108**, 146103.
2 V. G. Ruiz, W. Liu and A. Tkatchenko, *Phys. Rev. B*, **93**, 035118.
3 R. J. Maurer, V. G. Ruiz, J. Camarillo-Cisneros, W. Liu, N. Ferri, K. Reuter and A. Tkatchenko, *Progr. Surf. Sci.*, 2016, **91**, 72–100.
4 Y. Jiang, S. Yang, S. Li and W. Liu, *Sci. Rep.*, 2016, **6**, 39529.
5 S. Yang, Y. Jiang, S. Li and W. Liu, *Carbon*, 2017, **111**, 513–518.

Gerit Brandenburg answered: Molecules adsorbed on surfaces are indeed technically just another application. The computational expense grows (typically cubic scaling) with the size of the system and one has to converge the size of the unit cell to remove artificial interactions of the molecule with itself, *i.e.* to converge to the low-coverage adsorption, as well as the thickness of the surface layer. However, one has to carefully control the accuracy of the vdW corrected density functional as it is less established compared to organic crystals. This holds in particular for conducting surfaces, as the vdW corrections rely on local response models that might not be justified for conductors. Even the water adsorption on a wide gap hBN surface can be a challenge.[1] As the corresponding measurements are not trivial and reference quality calculations are rare, we need more experience for a better quality assessment.

VdW-inclusive DFT can be applied to small protein crystals like crambin within standard codes.[2] GPU accelerated program codes have been used for studying protein ligand binding affinities,[3] and linear scaling codes can be used to study the ion channel gramicidin A.[4]

1 Y. S. Al-Hamdani, M. Rossi, D. Alfè, T. Tsatsoulis, B. Ramberger, J. G. Brandenburg, A. Zen, G. Kresse, A. Grüneis, A. Tkatchenko and A. Michaelides, *J. Chem. Phys.*, 2017, **147**, 44710.
2 Piane *et al.*, *Chem. Sci.*, 2016, **7**, 1496; CRYSTAL17 code.
3 Ehrlich *et al.*, *ChemPhysChem*, 2017, **18**, 898; TerraChem code.
4 Todorović *et al.*, *J. R. Soc. Interface*, 2013, **10**, 20130547; CONQUEST code.

(329:[339]339) **Christian Schön** addressed Johannes Hoja and Gerit Brandenburg: What is a “large” molecule in the context of your energy calculations?

Johannes Hoja replied: The calculation time for molecular crystals does not directly depend on the size of the involved molecules, but rather on the number of atoms present in the unit cell of the studied molecular crystal. The largest unit cell in our study of the latest blind test systems consists of about 1000 atoms.

Gerit Brandenburg replied: From my experience this mainly depends on the type of application. The examples given in the answer to the previous question range from 1000 to 10 000 atoms per unit cell and would be considered large.

(330:[340]340) **Christian Schön** addressed Johannes Hoja and Gerit Brandenburg: Where would you place the current limits of your calculations, *i.e.* what kind of computational facilities (and what kind of time scales/levels of global optimization) are needed for which size of system?

Johannes Hoja responded: I believe that for production calculations the upper limit of PBE+TS or PBE+MBD lattice relaxations and subsequent PBE0+MBD energy calculations would be around 100 structures, and that vibrational free energies in the harmonic approximation could be calculated for tens of structures. The computation time actually required highly depends on the sizes of the sampled unit cells, and for the phonon calculations, also on the unit cell shape and symmetry. In general, a cluster with more than 500 cores would be advisable for these kinds of calculations. For example, a PBE+TS lattice optimization with subsequent PBE0+MBD energy evaluation and PBE+TS vibrational free energies (with settings as described in the paper) for Form A of system XXIII (172 atoms per unit cell) would require about 1100 CPU hours.

Gerit Brandenburg responded: Different codes scale differently and not all can fully use HPC facilities; for instance CRYSTAL17 has been shown to scale well up to 32 000 cores.[1] A DFT based re-ranking of crystal energy landscapes (or lists of putative polymorphs) has been conducted for the 6th CSP blind test;[2] the CPU time for local optimizations of 5000 structures are presented in Table S9 using 71 000 to 3 200 000 CPU hours. 100 000 CPU hours would amount to approximately 16 d elapsed time on a 256-core machine.

1 Erba *et al.*, *J. Chem. Theory Comput.*, 2017, **13**, 5019.
2 Reilly *et al.*, *Acta Cryst. B*, 2016, **72**, 439.

(331:[341]341) **Christian Schön** enquired: Experimental colleagues tell me that there is a net transfer of charge between the molecule and the metal surface. Does this complicate the calculations for you, or can you predict this charge transfer quantitatively?

Johannes Hoja answered: Van der Waals-inclusive DFT calculations capture charge transfer effects in principle, but I am not aware of any benchmark calculations.

Gerit Brandenburg answered: Technically this is not an issue, as the electron density is relaxed self-consistently. One can partition the self-consistent density to molecules, which would correspond to the charge transfer when compared to neutral molecules. However, some of the most frequently used DFT approaches of the generalized gradient type have a systematic error of too strong electron density delocalization, *i.e.* for a quantitative comparison to experiments one probably has to use a hybrid functional that reduces this error.

(340:[342]342) **Marcus Neumann** commented: Regarding a previous question about the CPU time requirements for PBE(0)+MBD+Fvib, I would like to mention that we recently implemented the approach in GRACE and are in the possession of some early CPU time benchmark results. The actual QM calculations were carried out with FHIaims through a driver implemented in GRACE. For blind test compound XXVI, we found that it is possible to process about 4 crystal structures on 384 modern Intel XEON cores in 1 day.

(345:[345]345) **Christian Schön** addressed Johannes Hoja and Gerit Brandenburg: What kind of codes support your methods – only FHI-AIMS?

Johannes Hoja replied: To my knowledge the Tkatchenko–Scheffler (TS) and the many-body dispersion (MBD) methods are implemented in FHI-aims, Quantum Espresso, VASP, CASTEP, ADF, Q-Chem, and DFTB+. Furthermore, a general many-body dispersion library is currently in development.[1]

[1] https://github.com/azag0/libmbd

Gerit Brandenburg replied: Both D3 as well as MBD dispersion corrections are implemented to be used in conjunction with DFTB3 or DFT in freely available codes.[1,2] Additionally, several codes support the corrections; a non-complete list: VASP (D3 and MBD), FHI-aim (D3 and MBD), ADF (D3 and MBD), DFTB+ (D3 and MBD), CRYSTAL17 (D3) CP2K (D3), Orca (D3), Turbomole (D3), Psi4 (D3), Gaussian (D3), QuantumEspresso (older D2 and TS), CASTEP (older D2 and TS), and Siesta (older D2).

[1] http://www.thch.uni-bonn.de/tc/dftd3
[2] http://www.fhi-berlin.mpg.de/~tkatchen/MBD

Conflicts of interest

Julian Helfferich and Marcus Neumann work for a company that develops software for organic crystal structure prediction. There are no other conflicts to declare.